



Agency conflicts and short- versus long-termism in corporate policies[☆]

Sebastian Gryglewicz^a, Simon Mayer^a, Erwan Morellec^{b,*}

^aErasmus University Rotterdam, the Netherlands

^bEPF Lausanne, Swiss Finance Institute, and CEPR, Switzerland

ARTICLE INFO

Article history:

Received 29 March 2019

Revised 18 June 2019

Accepted 28 June 2019

Available online 9 December 2019

JEL classification:

D86

D92

G30

Keywords:

Optimal short- and long-termism

Agency conflicts

Multitasking

ABSTRACT

We build a dynamic agency model in which the agent controls both current earnings via short-term investment and firm growth via long-term investment. Under the optimal contract, agency conflicts can induce short- and long-term investment levels beyond first best, leading to short- or long-termism in corporate policies. The paper analytically shows how firm characteristics shape the optimal contract and the horizon of corporate policies, thereby generating a number of novel empirical predictions on the optimality of short-versus long-termism. It also demonstrates that combining short- and long-term agency conflicts naturally leads to asymmetric pay-for-performance in managerial contracts.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Should firms target short-term objectives or long-term performance? The question of the optimal horizon of corporate policies has received considerable attention in recent years, with much of the discussion focusing on whether short-termism destroys value. The worry often

expressed in this literature is that short-termism—induced, for example, by stock market pressure—may lead firms to invest too little (see [Asker et al., 2015](#); [Bernstein, 2015](#); [Gutierrez and Philippon, 2017](#), for empirical evidence). Another line of argument recognizes, however, that while firms must invest in their future if they are to have one, they must also produce earnings today to pay for doing so. In line with this view, [Giannetti and Yu \(2018\)](#) find that firms with more short-term institutional investors suffer smaller drops in investment and have better long-term performance than similar firms following shocks that change an industry's economic environment.

While empirical evidence relating short- or long-termism to firm performance is accumulating at a fast pace, financial theory has made little headway in developing models that characterize the optimal horizon of corporate policies or the relation between firm characteristics and this horizon. In this paper, we attempt to provide an answer to these questions through the lens of agency theory. To do so, we develop a dynamic agency

[☆] We thank the editor (Toni Whited), the referee (Iván Marinovic), Rui Albuquerque, Peter DeMarzo, Pedro Gete, Mariassunta Giannetti, Zhiguo He, Julien Hugonnier, Dirk Jenter, Alejandro Rivera, Lukas Schmid, Alexander Szmayer, Felipe Varas (Cambridge discussant), Stéphane Villeneuve, Neng Wang, John Zhu (Cavalcade discussant), and seminar participants at Instituto de Empresa, the University of Hamburg, the Annual 3rd CEPR Symposium at Imperial College, the 2018 Cambridge Corporate Finance Theory Symposium, and the 2018 SFS Cavalcade at Yale University for useful comments. Erwan Morellec acknowledges financial support from the Swiss Finance Institute.

* Corresponding author.

E-mail addresses: gryglewicz@ese.eur.nl (S. Gryglewicz), mayer@ese.eur.nl (S. Mayer), erwan.morellec@epfl.ch (E. Morellec).

model in which the agent controls both current earnings and firm growth (i.e., future earnings) through unobservable investment. In this multitasking model, the principal optimally balances the costs and benefits of incentivizing the manager over the short or long term. As shown in the paper, this can lead to optimal short- or long-termism, depending on the severity of agency conflicts and firm characteristics. Additionally, we show that the same firm can find it optimal at times to be short-termist (i.e., favor current earnings) and at other times to be long-termist (i.e., favor growth). Our findings are generally consistent with the views expressed in *The Economist*¹ that “long-termism and short-termism both have their virtues and vices—and these depend on context.”

We start our analysis by formulating a dynamic agency model in which an investor (the principal) hires a manager (the agent) to operate a firm. In this model, agency problems arise because the manager can take hidden actions that affect both earnings and firm growth. As in [He \(2009\)](#) and [Bolton et al. \(2019\)](#), earnings are proportional to firm size, which is stochastic and governed by a (controlled) geometric Brownian motion (i.e., subject to permanent growth shocks). In contrast with these models, earnings are also subject to moral hazard and short-term shocks that do not necessarily affect (or correlate with) long-term prospects (i.e., shocks to firm size). The agent controls the drifts of the earnings and firm size processes through unobservable investment. Notably, the agent can stimulate current earnings via short-term investment and firm growth via long-term investment.

Investment is costly and the manager can divert part of the funds allocated to investment, which requires the compensation contract to provide sufficient incentives to the agent. Under the optimal contract, the manager is thus punished (rewarded) if either cash flow or firm growth is worse (better) than expected. Because the manager has limited liability, penalties accumulate until the termination of the contract, which occurs once the manager’s stake in the firm falls to zero. Since termination generates dead-weight costs, maintaining incentive compatibility is costly. Based on these trade-offs, the paper derives an incentive compatible contract that maximizes the value that the principal derives from owning the firm. It then analytically demonstrates that the optimal contract can generate short- or long-termism in corporate policies, defined as short- or long-term investment levels above first-best levels.

Our theory of short- and long-termism differs from existing contributions in two important respects. First, while most dynamic agency models focus either on short- or long-term agency conflicts, we consider a multitasking framework with both long- and short-term agency conflicts. We show that agency conflicts over different horizons interact, which can generate short- and long-termism in corporate policies. Second, unlike most models on short-termism, we do not assume that focusing either on the short or the long term is optimal. In our model, the optimal corporate horizon is determined endogenously

and reflects both agency conflicts and firm characteristics. These unique features allow us to generate a rich set of testable predictions about firms’ optimal investment rates and the horizon of corporate policies.

A first result of the paper is to show that short- or long-termism can only arise when the firm is exposed to a dual moral hazard problem. To understand why this condition is necessary, first consider long-termism. In our model, positive growth shocks lead to a permanent increase in earnings (and risk) and to a greater misalignment between shareholders’ interests and management’s incentives by diluting the manager’s stake in the firm. To offset these adverse dilution effects and reduce agency costs, the manager’s promised wealth must increase sufficiently in response to positive growth shocks. When the firm is exposed to both long- and short-term moral hazard, the contract optimally grants the manager a larger stake in the firm, which increases potential dilution effects. The principal then counteracts these dilution effects by tying the agent’s compensation more to long-term performance (i.e., long-term shocks), which leads to higher-powered long-term incentives.² The incentive compatibility condition with respect to long-term investment, which associates higher-powered incentives to higher levels of investments, in turn implies that the firm must also increase long-term investment, possibly beyond first-best levels. Our analysis demonstrates that long-termism is more likely to arise when cash flows are more volatile or when the investment technology is less efficient.

A second result of the paper is to show that short-termism can only arise if the firm is exposed to a dual moral hazard problem and there are direct externalities between short- and long-term investment. Notably, we show that a necessary condition for short-termism is that shocks to firm size and shocks to cash flows are correlated. When this correlation is negative—an assumption supported in the data (see, e.g., [Gryglewicz et al., 2019](#))—we additionally show that short-termism occurs when the agent’s stake in the firm is low and the risk of termination and agency costs are high. Indeed, in such instances, the benefits of long-term growth are limited. By contrast, stimulating short-term investment increases earnings and reduces the risk of termination and agency costs. Interestingly, a recent study by [Barton et al. \(2017\)](#) finds using a data set of 615 large- and mid-cap US publicly listed companies from 2001 to 2015 that “the long-term focused companies surpassed their short-term focused peers on several important financial measures.” While our model does indeed predict that firm performance should be positively related to the corporate horizon, it in fact suggests the reverse causality.³

² Throughout, we refer to the manager’s exposure to earnings shocks as short-term incentives and the exposure to growth as long-term incentives. However, we stress that long-term incentives are understood as incentives for a task that concerns the long term, rather than incentive pay contingent on outcomes that realize in the distant future.

³ Interestingly, this causality issue is already discussed in *The Economist*, Schumpeter’s article “Corporate short-termism is a frustratingly slippery idea,” who writes: “Do short-term firms become weak or do weak firms rationally adopt strategies that might be judged short term?” Similarly, [Barton et al. \(2017\)](#) write in their own study “one

¹ See “The tyranny of the long-term,” *The Economist*, November 22, 2014.

Third, we show that while Tobin's q unambiguously increases with financial slack (and in particular after positive cash flow realizations), long-term investment in general does not. This happens because long-term incentives are provided to implement optimal long-term investment and to insure the agent against dilution of her stake. Thus, long-term investment need not fully reflect the fundamental value of capital within the firm, as given by Tobin's q .

Incentives are provided in the optimal contract by making the agent's compensation contingent on firm performance via exposure to the firm's stock price and earnings. In previous dynamic contracting models, the optimal contract generates just enough incentives to the agent (i.e., incentive compatibility constraints are tight) because incentive provision comes with the threat of termination and is therefore costly to implement. A distinctive feature of our model is that the optimal contract introduces exposure to permanent shocks that is not needed to incentivize investment. In particular, the agent is provided minimal long-run incentives when the firm is close to financial distress and higher-powered long-run incentives after positive past performance when sufficient slack has been accumulated. In this region, incentives have option-like features and increase after positive performance.

To understand this result, note that when the manager's stake is large, and therefore subject to substantial dilution risk upon unexpected firm growth, it becomes optimal to mitigate these adverse dilution effect through high-powered incentive pay. This generates the distinct prediction that extra pay-for-performance is introduced when the manager's stake in the firm and dilution risk are large enough. We show indeed that in such instances the principal can eliminate dilution risk by fully exposing the manager's wealth to permanent shocks while maintaining incentive compatibility. When this is the case, long-run incentives are effectively costless and the manager is exposed to permanent, growth shocks beyond the level needed to incentivize long-term investment. In other words, positive permanent shocks lead to additional pay-for-performance, and negative permanent shocks eventually eliminate this extra sensitivity to performance implied by the optimal contract. Our model therefore provides a rationale for the asymmetry of pay-for-performance observed in the data (see, e.g., Garvey and Milbourn, 2006; Francis et al., 2013).

Our paper relates to the growing literature on short-termism. Influential contributions in this literature include Stein (1989), Bolton et al. (2006), and Aghion and Stein (2008) in which stock market pressure leads managers to boost short-term earnings at the expense of long-term value. In related work, Thakor (2018) builds a model in which short-termism is efficient, as it limits managerial rent extraction and leads to a better allocation of managers to projects. Narayanan (1985) develops a model in which short-term projects privately benefit managers by enhancing reputation and increasing wages. Von Thadden (1995) studies a dynamic model of financial contracting in

which the fear of early project termination by outsiders leads to short-term biases of investment. Marinovic and Varas (2019) and Varas (2017) develop dynamic contracting models in which the manager can undertake inefficient actions to boost short-run performance at the expense of the long run. Likewise, Zhu (2018) develops a model of persistent moral hazard in which the agent can choose between a short- and long-term action and characterizes the contract that implements the long-term action. In contrast with these models, we do not assume that focusing either on the short or the long term is optimal, and there is no intrinsic conflict between short- and long-termism in our setup. Hoffmann and Pfeil (2018) build a model in which the agent privately observe cash flows that he can divert and/or invest to increase the likelihood of adoption of future technologies.⁴ Their model does not address the issue of short- versus long-termism in corporate policies.

Our modeling of cash flows with permanent and transitory shocks is similar to that in Décamps et al. (2017) and Hackbarth et al. (2018). The model of Décamps et al. (2017) does not feature agency conflicts. The model of Hackbarth et al. (2018) shows that debt financing may render short-termism optimal for shareholders. Their dynamic agency model differs from ours in that it considers different managerial preferences and focuses on the agency-induced cost of debt (overhang). Consequently, the mechanism generating short-termism is distinct from ours. Notably, short-termism only arises because debt overhang reduces the benefits of long-term investment to shareholders that, in the presence of a resource constraint, leaves more resources for short-term investment. Unlike our model, their model does not feature long-termism or asymmetric pay-for-performance.

Our paper is more generally related to the growing literature on dynamic contracting. Most contributions in this literature study agency conflicts over the short run, using a stationary environment characterized by identically and independently distributed cash flow shocks; see, for example, DeMarzo and Sannikov (2006), Biais et al. (2007), Sannikov (2008), Zhu (2012), Miao and Rivera (2016), Malenko (2019), and Szydlowski (2019). Likewise, Biais et al. (2010) and DeMarzo et al. (2012) study dynamic contracting models with time-varying firm size in which cash flow shocks are short lived. In these models, the manager can affect current, but not directly future, firm performance. In contrast, He (2009) and He (2011) focus on agency conflicts over the long run by considering a framework in which the manager can affect firm growth.⁵ In these last two models, instantaneous earnings are not subject to short-term moral hazard. Our model combines both strands of the literature in a unified framework in which the optimal horizon of corporate policies arises

⁴ While Hoffmann and Pfeil (2018) find that overinvestment is more likely for firms with a superior investment technology, our model implies that overinvestment rather arises when the investment technology is inefficient. In Hoffmann and Pfeil (2018) firm size remains constant over time, thereby ruling out potential dilution of the managerial stake, so the mechanism leading to overinvestment differs from ours.

⁵ In a similar setting, Gryglewicz and Hartman-Glaser (2019) show that agency conflicts over the long run can lead to the early exercise of real options.

caveat: we've uncovered a correlation between managing for the long term and better financial performance; we haven't shown that such management caused that superior performance."

endogenously. Our framework is also related to [Ai and Li \(2015\)](#) and [Bolton et al. \(2019\)](#), who study optimal investment under limited commitment. These models do not feature moral hazard. [Ai and Li \(2015\)](#) demonstrate that shareholders' limited commitment can lead to overinvestment in a model in which firms are subject to permanent shocks. In contrast, we assume full commitment of shareholders (the principal) and identify agency frictions as a potential driver of overinvestment.

[Section 2](#) presents the model and its solution. [Section 3](#) analyzes the implications of the model for optimal investment. [Section 4](#) derives predictions on the horizon of corporate policies. [Section 5](#) shows how the optimal contract can be implemented by exposing the manager to the firm's stock price and earnings. [Section 6](#) focuses on asymmetric pay-for-performance. [Section 7](#) shows the robustness of our results to alternative model specifications. [Section 8](#) concludes.

2. The model

2.1. Assumptions

Throughout the paper, time is continuous and uncertainty is modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathcal{P})$ with the filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$, satisfying the usual conditions. We consider a principal-agent model in which the risk-neutral owner of a firm (the principal) hires a risk-neutral manager (the agent) to operate the firm's assets. In the model, firm performance depends on investment, which can be targeted toward the short- or long-run and entails a monetary cost. Agency problems arise because investment decisions are delegated to the manager, who can divert part of the resources allocated to investment.

The firm employs capital to produce output, whose price is normalized to one. At any time $t \geq 0$, earnings are proportional to the capital stock K_t (i.e., the firm employs an "AK" technology) and subject to permanent (long-term) and transitory (short-term) shocks. Permanent shocks change the long-term prospects of the firm and influence cash flows permanently by affecting firm size. Following [He \(2009\)](#), [DeMarzo et al. \(2012\)](#), and [Bolton et al. \(2011\)](#), we consider that the firm's capital stock (firm size) $\{K\} = \{K_t\}_{t \geq 0}$ evolves according to the controlled geometric Brownian motion process⁶:

$$dK_t = (\ell_t \mu - \delta) K_t dt + \sigma_K K_t dZ_t^K, \tag{1}$$

where $\mu > 0$ is a constant, $\delta > 0$ is the rate of depreciation, $\sigma_K > 0$ is a constant volatility parameter, $\{Z^K\} = \{Z_t^K\}_{t \geq 0}$ is a standard Brownian motion, and ℓ_t is the firm's long-term investment choice. For the problem to be well defined, we consider that $\ell_t \in [0, \ell_{\max}]$ with $\ell_{\max} < \frac{r+\delta}{\mu}$ where $r \geq 0$ is the constant discount rate of the firm owner. In addition to these permanent shocks, cash

flows are subject to short-term shocks that do not necessarily affect long-term prospects. Specifically, cash flows dX_t are proportional to K_t but uncertain and governed by

$$dX_t = K_t dA_t = K_t (s_t \alpha dt + \sigma_X dZ_t^X), \tag{2}$$

where α and σ_X are strictly positive constants, $s_t \in [0, s_{\max}]$ is the firm's short-term investment choice, and $\{Z^X\} = \{Z_t^X\}_{t \geq 0}$ is a standard Brownian motion. In the following, $\{Z^X\}$ is allowed to be correlated with $\{Z^K\}$ with correlation coefficient ρ in that⁷

$$\mathbb{E}[dZ_t^K dZ_t^X] = \rho dt, \text{ with } \rho \in (-1, 1). \tag{3}$$

Investment entails costs $\mathcal{I}(K_t, s_t, \ell_t)$. We assume that the investment cost is homogeneous of degree one in capital K_t , as in [DeMarzo et al. \(2012\)](#) and [Bolton et al. \(2011\)](#). That is, we have that $\mathcal{I}(K_t, s_t, \ell_t) \equiv K_t \mathcal{C}(s_t, \ell_t)$, where \mathcal{C} is increasing and convex in its arguments. Unless otherwise mentioned, we consider quadratic costs of investment

$$\mathcal{C}(s_t, \ell_t) = \frac{1}{2} (\lambda_s s_t^2 \alpha + \lambda_\ell \ell_t^2 \mu), \tag{4}$$

where s_{\max}, ℓ_{\max} are large enough to ensure that investment is interior at all times.⁸

The manager is protected by limited liability, does not accept negative payments from the principal dC_t , and cannot be asked to cover the investment cost $\mathcal{I}(K_t, s_t, \ell_t) = K_t \mathcal{C}(s_t, \ell_t)$ out of her own pocket. More specifically, the principal has to allocate funds to the manager before she can carry out the investment decisions s_t, ℓ_t . As a result, over $[t, t + dt]$ the agent is paid $dC_t + K_t \mathcal{C}(s_t, \ell_t) dt$ and wage payments net of investment cost dC_t must be positive (i.e., $dC_t \geq 0$). At any time t , the manager has full discretion over investment s_t, ℓ_t and can divert from the

⁷ In general, the correlation coefficient ρ between short-term and permanent cash flow shocks can be positive or negative. Considering, for example, the automobile industry, there is a general tendency for buyers of moving away from diesel cars toward electric cars. In the case of Volkswagen, this negative permanent demand shock on diesel cars has been compounded by the diesel gate, implying a positive correlation between short- and long-run cash flow shocks. In the case of Tesla Motors, the positive long-run demand shock on electric cars has been dampened by negative shocks on the supply chain (notably for Model 3), implying a negative correlation between short-run and long-run cash flow shocks. Additional examples of a negative correlation include decisions to invest in research and development or to sell assets. When the firm sells assets today, it experiences a positive cash flow shock. However, it also decreases permanently future cash flows. Examples of positive correlation include price changes due to the exhaustion of the existing supply of a commodity or improving technology for the production and discovery of a commodity. Using the same type of cash flow model as adopted in this paper, [Gryglewicz et al. \(2019\)](#) estimate that for Compustat firms between 1975 and 2014, the correlation between short-term and permanent cash flow shocks is, on average, negative.

⁸ The assumption of quadratic investment cost is made merely for analytical parsimony in that all our results in [Sections 2–5](#) hold true for any other cost function that is strictly convex in s, ℓ . This includes cost functions where short- and long-run investment are substitutes or complements, which occurs when $\frac{\partial^2 \mathcal{C}(s, \ell)}{\partial s \partial \ell} \neq 0$. We purposefully refrain from such a specification because interactions between short- and long-run investment arise endogenously in our model and we attribute these interactions entirely to the presence of moral hazard over different time horizons. The upper bounds on the investment levels can be related to the maximum time the manager can spend on the job. The upper bound on long-term investment (i.e., $\ell_{\max} < \frac{r+\delta}{\mu}$) also naturally arises in our model as a necessary condition to obtain finite firm values.

⁶ This specification for capital accumulation and revenue in which capital dynamics are governed by a controlled geometric Brownian motion has been used productively in asset pricing (e.g., [Cox et al., 1985](#); [Kogan, 2004](#)), corporate finance (e.g., [Abel and Eberly, 2011](#); [Bolton et al., 2019](#)), or macroeconomics (e.g., [Gertler and Kiyotaki, 2010](#); [Brunnermeier and Sannikov, 2014](#)).

funds $K_t C(s_t, \ell_t)$ she is handed over from the principal. In particular, the manager can change recommended short-run (respectively long-run) investment s_t (respectively ℓ_t) by any amount ε^s (respectively ε^ℓ) and keep the difference between actual investment cost and allocated funds, i.e.,

$$K_t [C(s_t, \ell_t) - C(s_t - \varepsilon^s, \ell_t - \varepsilon^\ell)]$$

for herself. Because $\{X\}$ and $\{K\}$ are subject to Brownian shocks—as long as $\sigma_X > 0$ and $\sigma_K > 0$ —there is moral hazard over short- and long-term investment decision. For simplicity, we assume that diversion does not entail efficiency losses.

In the baseline version of our model, we assume the agent has sufficient private funds so that she can, in principle, also boost firm investment (i.e., implement investment $\hat{s}_t > s_t$ or $\hat{\ell}_t > \ell_t$). While this assumption does not drive our main results, it offers several advantages. First, it considerably simplifies the analysis. Second, and most importantly, it allows us to connect more easily to the existing models of He, 2009 and DeMarzo et al. (2012) and to clearly demonstrate how the combination of short- and long-run moral hazard induces short- and long-termism. We analyze the case of limited private wealth in Section 5 and show that our results on short- and long-termism hold in this alternative setting.

As in DeMarzo and Sannikov (2006), Biais et al. (2007), and DeMarzo et al. (2012), the agent is more impatient than the principal and has a discount rate $\gamma > r$. As a result, the principal cannot indefinitely postpone payments to the agent. The agent possesses an outside option normalized to zero and maximizes the present value of her expected payoffs.⁹ Because the agent is protected by limited liability, her continuation value can never fall below her outside option, in which case she would profit from leaving the firm. Her employment starts at time $t = 0$ and is terminated at an endogenous stopping time τ , at which point the firm is liquidated. At the time of liquidation, the principal recovers a fraction $R > 0$ of assets, valued at RK_τ . Liquidation is inefficient and generates deadweight losses.¹⁰

Before proceeding, note that when $\sigma_K = 0$, we obtain the environment of the dynamic agency model of DeMarzo et al. (2012) or the financing frictions model of Bolton et al. (2011). Since there is no noise to hide the long-term investment choice, the long-term agency conflict is irrelevant in that case. By contrast, when $\sigma_X = 0$, we obtain the cash flow environment used in the dynamic capital structure (Leland, 1994; Strebulaev, 2007) and real options literature (Carlson et al., 2006; Morellec and Schürhoff, 2011) as well as in the dynamic agency models of He (2009, 2011). Since

⁹ As in Albuquerque and Hopenhayn (2004) and Rampini and Viswanathan (2013), we could assume that the manager can appropriate a fraction of firm value so that the manager has reservation value θK_t , where $\theta \geq 0$ is a constant parameter. The entire analysis can be conducted by replacing 0 with θ .

¹⁰ We could equally assume that the firm replaces the manager instead of liquidating when w falls to zero. The model results would remain unchanged as long as finding a new manager (i.e., replacement) is costly for the firm. For instance, one could assume some replacement cost kK_t , which could be microfounded by a costly labor market search.

there is no noise to hide the short-term investment choice, the short-term agency conflict is irrelevant in that case.

2.2. The contracting problem

To maximize firm value, the investor chooses short- and long-term investment $\{s\}$, $\{\ell\}$ and offers a full-commitment contract to the agent at time $t = 0$, which specifies wage payments $\{C\}$, recommended investment $\{s\}$, $\{\ell\}$, and a termination time τ . Because the agent cannot be paid any negative amount net of investment cost, the process $\{C\}$ is nondecreasing in that $dC_t \geq 0$ for all $t \geq 0$. Moreover, the contract cannot request the agent to finance investment so that she is handed over the investment cost $\mathcal{I}(K_t, s_t, \ell_t)$ at time t from the principal. We let $\Pi \equiv (\{C\}, \{s\}, \{\ell\}, \tau)$ represent the contract, where all elements are progressively measurable with respect to \mathbb{F} . With the agent's actual investment choice $\{\hat{s}\}$, $\{\hat{\ell}\}$, we call a contract incentive compatible if $s_t = \hat{s}_t$ and $\ell_t = \hat{\ell}_t$ for all $t \geq 0$ and focus throughout the paper on incentive compatible contracts, where we denote the set of these contracts by \mathbb{IC} . Since we only consider contracts of the set \mathbb{IC} , we will not formally distinguish between recommended and actual investment.

For an incentive compatible contract Π , let us define the agent's expected payoff at time $t \geq 0$ (i.e., her continuation value) as

$$W_t = W_t(\Pi) \equiv \mathbb{E}_t \left[\int_t^\tau e^{-\gamma(u-t)} dC_u \right].$$

$W_t = W_t(\Pi)$ equals the promised value the agent gets if she follows the recommended path from time $t \geq 0$ onwards. $W_0 = W_0(\Pi)$ is the agent's expected payoff at inception.

The principal receives the firm cash flows net of investment cost and pays the compensation to the manager. As a result, given the contract Π , the principal's expected payoff can be written as

$$\hat{P}(W, K) \equiv \mathbb{E} \left[\int_0^\tau e^{-rt} (dX_t - K_t C(s_t, \ell_t)) dt - dC_t \right] + e^{-r\tau} RK_\tau \Big|_{W_0 = W, K_0 = K}. \quad (5)$$

The objective of the principal is to maximize the present value of the firm cash flows plus termination value net of the agent's compensation, where we make the usual assumption that the principal possesses full bargaining power. Denote the set of incentive compatible contracts by \mathbb{IC} . The investor's optimization problem reads

$$P(W, K) \equiv \max_{\Pi \in \mathbb{IC}} \hat{P}(W, K) \text{ s.t. } W_t \geq 0 \text{ and } dC_t \geq 0 \text{ for all } t \geq 0. \quad (6)$$

With slight abuse of notation, we denote by $\Pi \equiv (\{C\}, \{s\}, \{\ell\}, \tau)$ the solution to this optimization problem.

2.3. First-best short- and long-term investment

We start by deriving the value of the firm and the optimal investment levels absent agency conflicts (i.e., when there is no noise to hide the agent's action in that

$\sigma_X = \sigma_K = 0$). Throughout the paper, we refer to this case as the first-best (FB) outcome.

Given the stationarity of the firm's optimization problem, the choice of s and ℓ is time invariant absent agency conflicts, and the first-best firm value reads

$$p^{FB}(K) = \max_{(s, \ell) \in [0, s_{\max}] \times [0, \ell_{\max}]} \frac{K}{r + \delta - \mu \ell} \times \left[\alpha s - \frac{1}{2} (\lambda_s \alpha s^2 + \lambda_\ell \mu \ell^2) \right] \equiv K p^{FB},$$

where the short- and long-term investment choice $\{s^{FB}, \ell^{FB}\}$ maximize firm value. We denote the scaled firm value absent moral hazard by p^{FB} . Simple algebraic derivations lead to the following result:

Proposition 1 (First-best firm value and investment choices). Assume the bounds i_{\max} for $i \in \{s, \ell\}$ are such that the first-best solution is interior. Then the following holds:

- (i) First-best short-term investment satisfies: $s^{FB} = \frac{1}{\lambda_s}$.
- (ii) First-best long-term investment satisfies: $\ell^{FB} = \frac{1}{\mu} \left[r + \delta - \sqrt{(r + \delta)^2 - \frac{\mu \alpha}{\lambda_s \lambda_\ell}} \right] = \frac{p^{FB}}{\lambda_\ell}$.

2.4. Model solution

We now solve the model with agency conflicts over the short and long term (i.e., assuming $\sigma_K > 0$ and $\sigma_X > 0$). Recall that the contract specifies the firm's investment policy $\{s\}$, $\{\ell\}$, payments to the agent C , and a termination date τ all as functions of the firm's profit history. Given an incentive compatible contract and the history up to time t , the discounted expected value of the agent's future compensation is given by W_t . As in DeMarzo and Sannikov (2006) or DeMarzo et al. (2012), we can use the martingale representation theorem to show that the continuation payoff of the agent solves

$$dW_t = \gamma W_t dt - dC_t + \beta_t^s (dX_t - \alpha s_t K_t dt) + \beta_t^\ell (dK_t - (\mu \ell_t - \delta) K_t dt). \tag{7}$$

This equation shows that the agent's continuation value must grow at rate γ to compensate for her time preference. In addition, compensation must be sufficiently sensitive to firm performance, as captured by the processes $\beta_t^s = dW_t/dX_t$ and $\beta_t^\ell = dW_t/dK_t$, to maintain incentive compatibility. To understand why such a compensation scheme may align incentives, suppose that the agent decides to deviate from the recommended choice and chooses investment $\hat{s}_t = s_t - \varepsilon$ during an instant $[t, t + dt]$. By doing so, she keeps the amount of investment cost saved

$$K_t (C(s_t, \ell_t) - C(s_t - \varepsilon, \hat{\ell}_t)) dt \simeq K_t C_s(s_t, \ell_t) \varepsilon dt = K_t \alpha \lambda_s s_t \varepsilon dt.$$

At the same time, however, she lowers mean cash flow by $K_t \alpha \varepsilon dt$ so that her overall compensation is reduced by $\alpha K_t \beta_t^s \varepsilon dt$. Therefore, the agent does not deviate from the prescribed short-run investment if $\beta_t^s = \lambda_s s_t$. Similarly, the agent does not deviate from the prescribed long-run investment if $\beta_t^\ell = \lambda_\ell \ell_t$. Both incentive compatibility constraints require that the agent has enough skin in

the game, as reflected by sufficient exposure to firm performance.

The investor's value function in an optimal contract, given by $P(W, K)$, is the highest expected payoff the investor may obtain given K and W . While there are two state variables in our model, the scale invariance of the firm's environment allows us to write $P(W, K) = K p(w)$ and reduce the problem to a single state variable: $w \equiv \frac{W}{K}$, the scaled promised payments to the agent as in He (2009) and DeMarzo et al. (2012).

To characterize the optimal compensation policy and its effects on the investor's (scaled) value function $p(w)$, note that it is always possible to compensate the agent with cash so that it costs at most \$1 to increase w by \$1 and $p'(w) \geq -1$. In addition, as shown by Eq. (7), deferring compensation increases the growth rate of W (and of w) and thus lowers the risk of liquidation but is costly due to the agent's impatience, $\gamma > r$. As a result, the optimal contract sets $dc \equiv \frac{dC}{K}$ to zero for low values of w and only stipulates payments to the manager once the firm has accumulated sufficient slack. That is, there exists a threshold \bar{w} with

$$p'(\bar{w}) = -1 \text{ and } dc = \max\{0, w - \bar{w}\}, \tag{8}$$

where the optimal payout boundary is determined by the super-contact condition:

$$p''(\bar{w}) = 0. \tag{9}$$

When w falls to zero, the contract is terminated and the firm is liquidated so that

$$p(0) = R. \tag{10}$$

When $w \in [0, \bar{w}]$, the agent's compensation is deferred and $dc = 0$. The Hamilton–Jacobi–Bellman equation for the principal's problem is then given by (see Appendix A.2):

$$(r + \delta) p(w) = \max_{s, \ell, \beta^s, \beta^\ell} \left\{ \alpha s - C(s, \ell) + p'(w) w (\gamma + \delta - \mu \ell) + \mu \ell p(w) + \frac{p''(w)}{2} [(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2] + 2 \rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) \right\}, \tag{11}$$

subject to the incentive compatibility constraints on β^s and β^ℓ .

Due to the scale invariance (i.e., $P(W, K_0) = p(w) K_0$) the investor's maximization problem at $t = 0$ can now be rewritten as

$$\max_{w_0 \in [0, \bar{w}]} p(w_0) K_0, \tag{12}$$

with unique solution $w_0 = w^*$ satisfying

$$p'(w^*) = 0.$$

As a consequence, the principal initially promises the agent utility $w^* K_0$ and expects a payoff $P(K_0 w^*, K_0) = p(w^*) K_0$. For convenience, we normalize K_0 to unity in the following and refer to $p(w^*)$ as expected payoff instead of scaled expected payoff. The following proposition summarizes

our results about the optimal contract. Its proof is deferred to [Appendix A.2](#).

Proposition 2 (Firm value and optimal compensation under agency). *Let $\Pi \equiv (\{C\}, \{s\}, \{\ell\}, \tau)$ denote the optimal contract solving problem (6). The following holds true:*

1. There exist \mathbb{F} -progressive processes $\{\beta^\ell\}$ and $\{\beta^s\}$ such that the agent's continuation utility W_t evolves according to (7). The optimal contract is incentive compatible in that $\beta^s = \lambda_s s$ and $\beta^\ell = \lambda_\ell \ell$, where $\{s\}, \{\ell\}$ are the firm's optimal investment decisions.
2. Firm value is proportional to firm size in that $P(W, K) = Kp(w)$. The scaled firm value $p(w)$ is the unique solution to Eq. (11) subject to (8), (9), and (10) on $[0, \bar{w}]$. For $w > \bar{w}$, the scaled value function satisfies $p(w) = p(\bar{w}) - (w - \bar{w})$. Scaled cash payments $dc = \frac{dc}{K}$ reflect w back to \bar{w} .
3. The function $p(w)$ is strictly concave on $[0, \bar{w}]$.

Before proceeding, note that w serves as a proxy for the firm's financial slack in our model so that states where w is close to zero—and the firm close to liquidation—correspond to financial distress. Since the firm has to undergo inefficient liquidation after a series of adverse shocks drive w down to zero, the principal becomes effectively risk averse with respect to the volatility of w so that the value function is strictly concave (i.e., $p''(w) < 0$ for $w < \bar{w}$). Put differently, the concavity of p implies that the principal would like to minimize the volatility of w while maintaining incentive compatibility.

Note also that overall value, $P(W, K) + W$, is split between the principal and the manager, where the manager obtains a fraction,

$$S(w) = \frac{W}{P(W, K) + W} = \frac{w}{p(w) + w},$$

of overall value. Because of $S'(w) > 0$ for all $w \in (0, \bar{w})$, the scaled continuation value w corresponds (monotonically) to the fraction of overall firm value that goes to the manager.¹¹ Therefore, we also refer to w as the agent's or manager's stake in the firm. When the manager's stake w falls down to zero, she has no more incentives to stay and accordingly leaves the firm. In this case, deadweight losses are incurred due to contract termination.

3. Short- versus long-run incentives

This section examines the implications of agency conflicts for long- and short-term investment choices. For clarity of exposition, we assume that the correlation between short- and long-run shocks ρ is zero and the parameters are such that investment levels s and ℓ are interior. [Section 4.2](#) analyzes the effects of nonzero correlation.

3.1. Short-term investment and incentives

Optimal short-term investment $s = s(w)$ is obtained by taking the first-order condition in [Eq. \(11\)](#) after using the

incentive compatibility condition $\beta^s = \lambda_s s$. This yields the following result:

Proposition 3 (Optimal short-term investment). *Optimal short-term investment is given by*

$$s(w) = \frac{\overbrace{\alpha}^{\text{Direct benefit of investment}}}{\underbrace{\lambda_s \alpha}_{\text{Direct cost of investment}} + \underbrace{-p''(w)(\lambda_s \sigma_X)^2}_{\text{Agency cost of investment}}}. \tag{13}$$

Short-term investment is strictly lower than under first-best except at the boundary in that $s(w) < s^{SB}$ for $w < \bar{w}$ and $s(\bar{w}) = s^{FB}$. If $\gamma - r$ and σ_K are sufficiently small, then $s(w)$ increases in w (i.e., $\frac{\partial s(w)}{\partial w} > 0$).

An important implication of [Proposition 3](#) is that agency conflicts lead to underinvestment for the short run (i.e., $s(w) < s^{FB}$ when $\rho = 0$). Upon increasing the investment rate s , the firm does not only incur direct, monetary cost of investment but also agency costs because higher s requires higher incentives β^s . Consequently, the agent's stake becomes more volatile, which raises the risk of costly liquidation and therefore leads to endogenous agency costs or incentive costs of investment. These agency costs decrease in the level of financial slack w and vanish at the payout boundary \bar{w} where $p''(\bar{w}) = 0$, at which point the firm's short-run investment reaches first best, $s(\bar{w}) = s^{FB}$.

3.2. Long-term incentives and investment

Next, we characterize the firm's optimal long-term investment ℓ and the agent's long-term incentives β^ℓ . Using the HJB [Eq. \(11\)](#) and the incentive compatibility condition $\beta^\ell = \lambda_\ell \ell$, we get the following result:

Proposition 4 (Optimal long-term investment). *Optimal long-term investment is given by*

$$\ell(w) = \frac{\overbrace{\mu(p(w) - p'(w)w)}^{\text{Direct benefit of investment}} + \overbrace{-p''(w)w\lambda_\ell\sigma_K^2}^{\text{Reducing dilution risk}}}{\underbrace{\lambda_\ell\mu}_{\text{Direct cost of investment}} + \underbrace{-p''(w)(\lambda_\ell\sigma_K)^2}_{\text{Agency cost of investment}}}. \tag{14}$$

The firm always underinvests for the long-term close to the boundary in that there exists $\varepsilon > 0$ such that $\ell(w) < \ell^{FB}$ for $w \in [\bar{w} - \varepsilon, \bar{w}]$.

To get some intuition for the results in [Proposition 4](#), let us consider the costs and benefits from marginally increasing long-term investment ℓ :

$$\frac{\partial p(w)}{\partial \ell} \propto \underbrace{\mu(p(w) - p'(w)w)}_{\text{Direct benefit}} - \underbrace{\lambda_\ell\mu\ell}_{\text{Direct cost}} + \underbrace{p''(w)\ell(\lambda_\ell\sigma_K)^2}_{\text{Agency cost}} - \underbrace{p''(w)w\lambda_\ell\sigma_K^2}_{\text{Reducing dilution risk}}. \tag{15}$$

Consider first the costs of raising long-term investment. The above expression shows that, in addition to the direct cost of investment, the firm incurs an agency cost. This

¹¹ In fact, $S'(w) \propto p(w) - p'(w)w$, which is always positive since $\frac{\partial}{\partial w}(p(w) - wp'(w)) = -wp''(w) > 0$ due to concavity of the value function. Since $S'(0) = R \geq 0$, it follows that $S'(w) > 0$ for all $w \in (0, \bar{w}]$.

agency cost arises because increasing long-run investment requires higher long-run incentives β^ℓ and therefore makes w more volatile. The agency cost of investment depends on the principal's effective risk aversion $-p''(w)$ and decreases optimal investment $\ell(w)$.

Consider next the benefits of raising long-term investment. The first difference between optimal short- and long-term investment is that the direct benefit of long-term investment is time varying and given by $p(w) - p'(w)w$. This value represents Tobin's marginal q in our model, as it equals the marginal value of capital $P_K(W, K)$. Note that long-term investment expenditures today lead to a higher average cash flow rate in the future. However, due to the possibility of firm liquidation owing to the moral hazard problem, the firm cannot perpetually enjoy this increase in the cash flow rate, so the benefit of long-term investment $p(w) - p'(w)w$ is strictly lower than p^{FB} . Ceteris paribus, this lowers the firm's investment rate $\ell(w)$. Remarkably, in contrast to the case of short-term investment, long-term investment is below the first-best level for high w close to \bar{w} . The reason is that while the agent becomes a residual claimant on cash flows at \bar{w} , she is not a residual claimant on the benefits of long-term growth at the first-best level because of the agency-induced firm liquidation in the future. It holds, however, that long-term investment is more profitable when the firm has more financial slack and the distance to liquidation is far (i.e., $p(w) - wp'(w)$ increases in w).

A second difference is that investment in $\ell(w)$ offers an additional benefit compared to investment in $s(w)$: it mitigates the dilution of the agent's stake w . Since $p''(w) \leq 0$, this effect unambiguously increases long-term investment. To understand the source of this effect, first note that by Ito's lemma, the dynamics of the agent's stake are given by¹²

$$dw = (\gamma + \delta - \mu\ell)w dt + \beta^s \sigma_X dZ^X + (\beta^\ell - w)\sigma_K dZ^K, \tag{16}$$

so the instantaneous variance of dw satisfies

$$\Sigma(w) \equiv \frac{\mathbb{V}(dw)}{dt} = (\beta^s \sigma_X)^2 + (\beta^\ell - w)^2 \sigma_K^2. \tag{17}$$

From Eq. (16), we see that a positive permanent shock $dZ^K > 0$ has two opposing effects on the manager's incentives. First, the agent is rewarded for strong performance via the sensitivity β^ℓ and is promised higher future payments W . This increases $w = \frac{W}{K}$ (via its numerator) by $\beta^\ell \sigma_K dZ^K$, which equals $\lambda_\ell \ell(w) \sigma_K dZ^K$. Second, firm size K grows more than expected, thereby reducing the agent's stake $w = \frac{W}{K}$ (via its denominator) by $-w\sigma_K dZ^K$. We refer to the reduction of the agent's stake upon a positive shock $dZ^K > 0$ as dilution and the volatility generated by this effect (i.e., $-w\sigma_K$), as dilution risk. Altogether, we have that $dw/dZ^K = (\beta^\ell - w)\sigma_K$. Because performance-based compensation and dilution move w in opposite directions,

long-run incentives β^ℓ mitigate the dilution effect that, ceteris paribus, lowers risk (see Eq. (17)) and is thus beneficial. This makes contracting for long-term investment cheaper and increases $\ell(w)$.

More generally, our model suggests that the manager's compensation should increase with firm size. Indeed, an increase in firm size (due to a positive permanent shock $dZ^K > 0$) raises both the firm's future cash flow rate and the magnitude of future cash flow shocks. As a result, the firm becomes not only more profitable but also more risky (in absolute terms). Both effects call for an increase in the manager's continuation value, which better aligns the manager's and the principal's interests and facilitates contracting for long-term investment.

It is illustrative to look at this effect from the perspective of agency costs. As long as $\beta^\ell < w$, raising β^ℓ lowers the volatility and instantaneous variance $\Sigma(w)$ of w , and therefore the risk of liquidation, so that the effective (marginal) agency cost of long-run investment is pinned down by the net change in risk, that is, by

$$\underbrace{-p''(w)\ell(\lambda_\ell \sigma_K)^2}_{\text{Agency cost } (>0)} + \underbrace{p''(w)w\lambda_\ell \sigma_K^2}_{\text{Reducing dilution risk } (<0)} = \underbrace{-p''(w)\sigma_K^2 \lambda_\ell (\lambda_\ell \ell - w)}_{\text{Effective agency cost } (\leq 0)}. \tag{18}$$

As is the case with the agency cost of investment, the benefits of mitigating dilution risk depend on how much volatility in w matters for the investor's value function (i.e., on principal's effective risk aversion $-p''(w)$). Therefore, it is most beneficial to alleviate dilution via long-run incentives β^ℓ when the concavity of the scaled value function is the largest. The effect disappears at $w = \bar{w}$ where $p''(\bar{w}) = 0$. When w is close to \bar{w} , and therefore $p''(w) \simeq 0$ and $p'(w) \simeq -1$, the firm always underinvests because direct benefits of investment $p(w) - wp'(w) \simeq p(w) + w < p^{FB}$ are reduced by the presence of moral hazard and agency-induced firm liquidation, which implies $\ell(w) = (p(w) + w)/\lambda_\ell < p^{FB}/\lambda_\ell = \ell^{FB}$.

4. Short- and long-termism in corporate policies

Because the manager's ability to divert funds decreases the benefits of investment, each moral hazard problem working in isolation leads to underinvestment relative to the first-best levels. The novel insight of our model is that a simultaneous moral hazard problem over both the short and long run can generate overinvestment. We call overinvestment for the long run (i.e., $\ell > \ell^{FB}$), long-termism and overinvestment for short-run (i.e., $s > s^{FB}$), short-termism. Below we analyze and contrast the circumstances that lead to long-termism and short-termism. We find that long-termism can arise irrespective of whether the different sources of cash flow risk are correlated, while short-termism requires $\rho \neq 0$.

4.1. Long-termism

Proposition 4 and Eq. (14) reveal that moral hazard decreases long-run investment via the direct benefit channel and the agency cost channel. The firm can potentially overinvest to reduce dilution risk. In the next

¹² As discussed in Appendix A.2, the dynamics in Eq. (16) are under an auxiliary measure \mathcal{P} rather than under the physical measure \mathcal{P} . The choice of the probability measure does not matter since w has the same volatility under both measures and volatility is the only quantity we study in in the following discussion.

proposition, we show that the last effect can dominate the former two effects and present sufficient conditions for overinvestment to arise.

Proposition 5 (Long-termism). *The following holds true:*

- (i) Long-termism (i.e., $\ell(w) > \ell^{FB}$) arises only if $\sigma_X > 0$ and $\sigma_K > 0$.
- (ii) Assume $\sigma_X > 0$ and $\sigma_K > 0$. Then, there exist w^L and w^H with $0 < w^L < w^H < \bar{w}$ such that $\ell(w) > \ell^{FB}$ for $w \in (w^L, w^H)$, provided that μ and $\gamma - r$ are sufficiently low.
The firm underinvests (i.e., $\ell(w) < \ell^{FB}$) when $w < w^L$ or $w > w^H$ (i.e., when w is close to zero or close to \bar{w}).
- (iii) Higher volatility $\sigma_X > 0$ or $\sigma_K > 0$ favors long-termism: if μ is sufficiently low and parameters are such that $\sup\{\ell(w) : 0 \leq w \leq \bar{w}\} = \ell^{FB}$, then there exists $\varepsilon > 0$ such that $\sup\{\ell(w) : 0 \leq w \leq \bar{w}\} > \ell^{FB}$ if σ_X or σ_K increases by ε .

The first part of Proposition 5 states that long-termism can only arise when firm cash flows are subject to both transitory and permanent shocks (i.e., when $\sigma_X > 0$ and $\sigma_K > 0$), and the firm is exposed to a simultaneous moral hazard problem over both the short and long run. When permanent cash flow shocks are removed from the model (i.e., $\sigma_K = 0$), long-term investment ℓ is observable and contractible. In addition, there is no risk of dilution of the agent's stake, as all shocks are purely transitory in nature. Under these circumstances, long-term investment satisfies

$$\ell(w) = \frac{p(w) - wp'(w)}{\lambda_\ell} < \frac{p^{FB}}{\lambda_\ell} = \ell^{FB}.$$

Because short-run agency lowers the direct benefits of long-run investment, the firm always underinvests for the long term.

To see why transitory shocks, or equivalently moral hazard over the short term, are essential for long-termism, we start with the following observation. Since the direct benefit of long-term investment under moral hazard is below the first-best level, it follows from Eq. (15) that a necessary condition for overinvestment in $\ell(w)$ is that the dilution effect exceeds the agency cost effect. Using Eq. (18), this is equivalent to requiring that the effective (marginal) agency cost is negative. Thus, overinvestment in ℓ or long-termism arises only if

$$-p''(w)(\lambda_\ell \ell - w)\lambda_\ell \sigma_K^2 < 0 \iff w > \lambda_\ell \ell = \beta^\ell,$$

that is, if the manager's stake is large relative to her long-term incentives. When $\sigma_X = 0$, the firm faces no transitory cash flow risk and therefore optimally grants the manager a relatively low stake, which puts a limit on potential dilution effects. More specifically, if it were that $w > \beta^\ell = \lambda_\ell \ell$, it follows from Eq. (17) that the firm would profit from decreasing w by making infinitesimal payouts $dc > 0$ and thus reducing the risk in w by

$$\Sigma(w) - \Sigma(w - dc) \simeq (w - \lambda_\ell \ell)dc > 0.$$

This strategy would reduce the risk the manager is exposed to and still provide sufficient incentives. Consequently, $\sigma_X = 0$ implies that $w \leq \lambda_\ell \ell$ for all w , the effective

agency cost of long-term investment is positive, and the firm underinvests in ℓ .¹³

When both σ_X and σ_K are strictly positive, the above argument does not work, as the firm also needs to account for short-run risk and incentives. To decrease termination risk, it can then be optimal for the firm to delay payments to the manager further, even if $w \simeq \beta^\ell$ and the manager's stake is barely exposed to permanent cash flow risk. This can lead to w exceeding β^ℓ , that is, to a negative effective agency cost and to overinvestment in ℓ . The mechanism is as follows. When the agent holds a large stake w , the risk of dilution identified above generates additional termination risk, which diminishes the risk reduction induced by postponing payouts. The principal can mitigate these adverse dilution effects by tying the agent's compensation more to long-term performance, which leads to higher long-run incentives β^ℓ . The incentive compatibility condition $\beta^\ell = \lambda_\ell \ell$ then implies that the firm must also increase long-term investment.

The second part of Proposition 5 shows that long-termism arises when the asset growth rate μ is low (i.e., when long-run investment is sufficiently inefficient). Proposition 5 therefore offers a potential explanation for the puzzling empirical evidence that in recent years capital is not allocated to the industries with the best growth opportunities (as recently shown by Lee et al., 2018). Additionally, long-termism arises when cash flow is sufficiently volatile in either time-horizon (i.e., $\sigma_X > 0$ and $\sigma_K > 0$ are large), and when the agent is sufficiently patient (i.e., $\gamma - r > 0$ is low).

The intuition for these findings is as follows. As explained above, long-termism requires the dilution effect to exceed the agency cost effect, which happens when the manager's stake is large relative to her long-term incentives, $w > \beta^\ell$. When this is the case, the effective agency cost is negative. Both higher cash flow risk (σ_X and σ_K) and lower cost of delaying payouts ($\gamma - r$) increase the value of deferred compensation so that \bar{w} rises, leading to an (average) increase in the manager's stake within the firm. On the other hand, low asset growth rate decreases contracted long-term investment ℓ and accordingly long-term incentives β^ℓ .

To generate long-termism, the agency-cost-based motives for overinvestment must also exceed the preference for underinvestment that arises because of the diminished direct benefit of investment. Recall that the marginal direct benefit of long-run investment under moral hazard equals $\mu(p(w) - p'(w)w)$ and is below its first-best counterpart, while the marginal direct cost $\lambda_\ell \mu$ is at the first-best level. Since both the direct benefit and cost are proportional to μ , this motive to underinvest is quantitatively low when μ is low and can then be overcome by the agency-cost-based preference for overinvestment.

¹³ In fact, the inequality is strict: $w < \beta^\ell = \lambda_\ell \ell(w)$. To get some intuition, note that in case $\beta^\ell = w$, the firm becomes riskless. The benefits of reducing w by an infinitesimal amount $dc > 0$ are proportional to $(\gamma - r)o(dc)$ and therefore of order $o(dc)$, while the cost—stemming from the additional risk of liquidation—are of order $o((dc)^2)$. Consequently, the firm would never set $\beta^\ell = w$ so that $\beta^\ell = \lambda_\ell \ell(w) > w$ for all $w \in [0, \bar{w}]$. This result is established in He (2009).

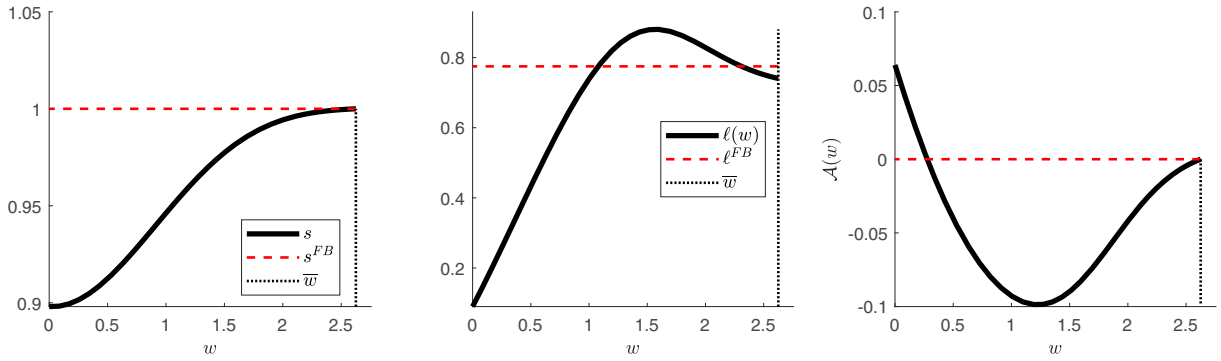


Fig. 1. Numerical example of long-termism. The first two panels depict optimal investment as functions of w . The third panel at the right displays effective agency cost $\mathcal{A}(w) = -p''(w)(\lambda_\ell \ell(w) - w)$. The parameters are $\alpha = 0.25$, $\sigma_K = 0.25$, $\sigma_X = 0.2$, $\rho = 0$, $\mu = 0.025$, $r = 0.046$, $\gamma = 0.048$, $\delta = 0.125$, $\lambda_s = \lambda_\ell = 1$, and $R = 0.2$.

Last, we note that while marginal q , $p(w) - wp'(w)$, unambiguously increases in w , long-term investment in general does not. This is because the marginal cost of capital—which includes the direct cost and the effective agency cost—is time varying. In particular, as explained above, the effective agency cost can increase in w . Thus long-term investment and marginal q can move in opposite directions. In general, the investment sensitivity with respect to marginal q possesses an ambiguous sign.¹⁴

Fig. 1 presents a quantitative example illustrating long-termism. The parameters satisfy the conditions set in Proposition 5 and are as follows. We set the discount rate parameters to $r = 4.6\%$ and $\gamma = 4.8\%$ and the depreciation rate to $\delta = 12.5\%$, similar to DeMarzo et al. (2012). The volatility parameter of the long-term shock is set to $\sigma_K = 20\%$, in line with Kogan (2004), while the volatility parameter of the short-term shock is set to $\sigma_X = 25\%$, in line with DeMarzo et al. (2012). The drift parameter for the profitability process is set to $\alpha = 25\%$. The left plot shows that the firm underinvests in the short run for all w . The middle plot shows that the firm overinvests in the long run for intermediate values of w . This is when the dilution effect, whose magnitude is proportional to $p''(w)w\sigma_K^2$, is the strongest. The right plot also shows that long-termism is related to a negative effective agency cost. Conversely, according to Proposition 5, long-termism never arises in financial distress (i.e., when w is close to 0), or when the firm is expected to make direct payments to the manager (i.e., when w is close to \bar{w}).

4.2. Correlated cash flow shocks and short-termism

As shown in Proposition 3, short-termism cannot occur in our baseline model with independent shocks (i.e., when $\rho = 0$). When permanent and transitory cash flow shocks are correlated, direct externalities between short- and long-term investment and incentives arise. These externalities can lead to corporate short-termism (i.e., to $s > s^{FB}$), as we demonstrate below.

To start with, note that when shocks are correlated, optimal short- and long-term investment are given by

$$s(w) = \frac{\alpha + \overbrace{p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell \ell(w) - w)}^{\text{Externality}}}{\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2} \quad (19)$$

and

$$\ell(w) = \frac{\mu(p(w) - p'(w)w) + \overbrace{p''(w)\rho\sigma_X\sigma_K\lambda_\ell\lambda_s s(w) - p''(w)w\lambda_\ell\sigma_K^2}^{\text{Externality}}}{\lambda_\ell\mu - p''(w)(\lambda_\ell\sigma_K)^2} \quad (20)$$

Compared to Eqs. (13) and (14), new terms appear that affect optimal investment levels and incentives. Since $s(w)$ depends on $\ell(w)$ and vice versa, there are direct externalities between investment levels and incentives. Intuitively, when the two sources of risk are positively correlated, exposing the manager's continuation payoff to both transitory and permanent shocks creates additional volatility and is therefore costly. Conversely, when the correlation is negative, exposure to both shocks partially cancels out, thereby reducing the volatility of the manager's continuation payoff w .

From Eq. (20), the externality of $s(w)$ on $\ell(w)$ is negative (positive) if $\rho > 0$ ($\rho < 0$). The magnitude of the externality scales with the curvature of the value function $p''(w)$ (i.e., the principal's effective risk aversion) and is therefore relatively weaker once w is sufficiently large and the risk of termination is sufficiently remote.

Likewise, Eq. (19) demonstrates that the choice of long-term investment $\ell(w)$ also feeds back into the choice of short-term investment $s(w)$. However, the externality effect in the numerator of $s(w)$ in Eq. (19) has two separate components:

$$p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell \ell(w) - w) = \underbrace{p''(w)\rho\sigma_X\sigma_K\lambda_s\lambda_\ell \ell(w)}_{\text{Agency cost}} - \underbrace{p''(w)\rho\sigma_X\sigma_K\lambda_s w}_{\text{Reducing dilution risk}} \quad (21)$$

This decomposition shows that when the correlation between shocks is nonzero, incentives for the short run are also used to counteract the dilution in the manager's stake arising upon positive permanent shocks $dZ^K > 0$. As

¹⁴ Bolton et al. (2011) also show that investment can be decreasing in marginal q when the marginal source of financing is credit line.

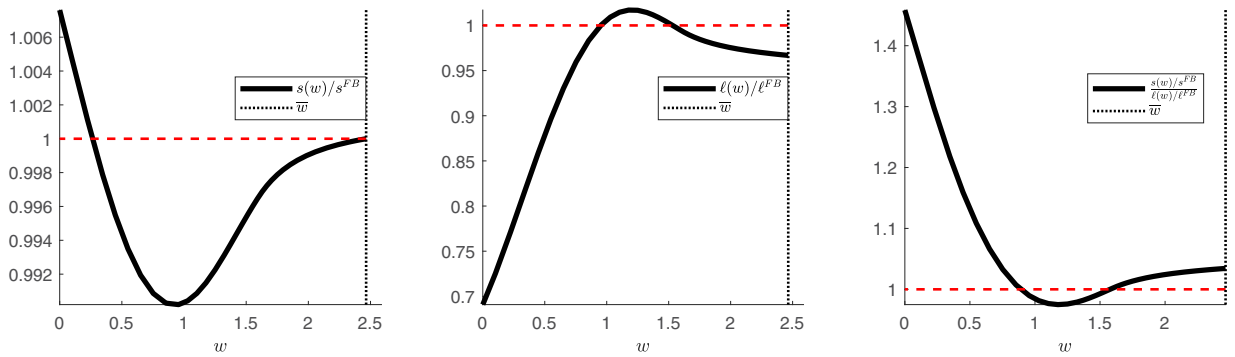


Fig. 2. Numerical example of short-termism. The parameters are $\alpha = 0.25$, $\sigma_X = 0.15$, $\sigma_K = 0.5$, $\rho = -0.75$, $\mu = 0.025$, $r = 0.046$, $\gamma = 0.047$, $\delta = 0.125$, $\lambda_S = 1.15$, $\lambda_\ell = 0.25$, and $R = 0.75$.

discussed in Section 4.1, with no correlation, the principal counteracts this dilution effect by tying the manager's compensation to permanent shocks and increasing long-term incentives. When the two sources of cash flow risk are positively (negatively) correlated, it is possible to reduce dilution risk also by means of higher (lower) short-term incentives.

Notably, when $\rho < 0$ and w is low, positive risk externalities of short- and long-term incentives emerge and may dominate dilution effects of short-term incentives. In this case, short-termism, $s(w) > s^{FB}$, can become optimal.

Proposition 6 (Short-termism under distress with $\rho < 0$) The following holds true:

- (i) Short-termism arises only if $\sigma_X > 0$, $\sigma_K > 0$, and $\rho \neq 0$. Conversely, if either $\sigma_X = 0$, $\sigma_K = 0$, or $\rho = 0$, short-termism cannot arise and $s(w) \leq s^{FB}$ for all w .
- (ii) Assume $\sigma_X > 0$, $\sigma_K > 0$, and $\rho < 0$. Then, there exist $w^L < w^H$ with $s(w) > s^{FB}$ for $w \in (w^L, w^H)$, provided σ_X is sufficiently small. When, in addition, λ_ℓ and $\gamma - r$ are sufficiently small, the set $\{w \in [0, \bar{w}] : s(w) > s^{FB}\}$ is convex and contains zero and $s(w)$ decreases on this set.

While long-termism occurs mainly for large values of the manager's stake w with the objective to alleviate the excessive dilution risk via long-run incentives β^ℓ , short-termism is more likely to occur for low values of w when the correlation between shocks is negative. When the agent's stake w is small, dilution risk is negligible and positive externalities between short- and long-term incentives induce more short-term investment. In addition, short-termism can arise when cash flow risk σ_X is small so that short-run agency cost is sufficiently low and does not dominate the externality effect.

Fig. 2 provides an example of short-termism when the correlation between long- and short-term shocks is negative. Consistently with Proposition 6, the firm overinvests in the short-run when in distress and w is close to zero. Fig. 2 further illustrates that both short- and long-termism may, but need not happen, within the same firm, depending on the level of financial slack as measured by w . In distress, the firm overinvests in generating (short-term) profits, while after a strong performance, the firm overin-

vests in (long-term) growth. While the effects of absolute short-termism appear to be quantitatively small, the effects of relative short-termism $\frac{s(w)/s^{FB}}{\ell(w)/\ell^{FB}}$, which determines whether investment is distorted toward the short-term compared to first-best, can be quantitatively large. Absent agency fictions, this ratio equals, by construction, one, and a value above (below) one indicates an investment distortion toward the (long) short run. The right-hand side plot of Fig. 2 presents the relative short-termism ratio, which, for our parameter values, is a nonmonotonic U-shaped function of w . The relative short-termism for large w close to \bar{w} arises for all parameters (compare Proposition 3 and 4). The ratio is below one for intermediate w whenever substantial absolute long-termism arises. Relative short-termism again dominates for low w , and this region exists due to negative ρ and relatively low cost of short-term investment (low σ_X and λ_ℓ ; cf. Proposition 6).

Our focus on the case of negative correlation is due to the findings in Chang et al. (2014) and Gryglewicz et al. (2019) that the correlation coefficient between permanent and short-term cash flow shocks ρ is, on average, negative. When this is the case, our model predicts that firms with a high risk of liquidation (i.e., firms that perform worse and have little financial slack) should find it optimal to focus on the short term (i.e., current earnings) while firms with a low risk of liquidation (i.e., cash-rich firms that perform well) should find it optimal to focus on the long term (i.e., asset growth). Interestingly, a recent study by Barton et al. (2017) find, using a data set of 615 large- and mid-cap US publicly listed companies from 2001 to 2015, "the long-term focused companies surpassed their short-term focused peers on several important financial measures." While our model does indeed predict that firm performance should be positively related to the corporate horizon, it in fact suggests the reverse causality.

For completeness, we also investigate optimal investment when the correlation between cash flow shocks is positive. In this case, the firm can overinvest in both short- and long-term investment at the same time. This happens when the agent's stake in the firm is large, thereby exposing the manager to a high risk of dilution. To reduce this dilution risk, the principal provides high-powered incentives to the manager. Importantly, when correlation

is positive, unexpected asset growth $dZ^K > 0$ triggers, on average, unexpected cash flow ρdZ^K , which leads a reward $(\beta^\ell + \rho\beta^S)dZ^K$ for the agent.¹⁵ Consequently, both short- and long-run incentives counteract the adverse dilution in the agent's stake so that the desire to mitigate dilution risk translates into high-powered incentives and, accordingly, to overinvestment for both time horizons. The next proposition characterizes this outcome.

Proposition 7 (Short-termism with $\rho > 0$) Assume $\sigma_X > 0$, $\sigma_K > 0$ and $\rho > 0$. Then, there exist $w^L < w^H$ with $s(w) > s^{FB}$ for $w \in (w^L, w^H)$, provided $\sigma_X > 0$ and $\gamma - r$ is sufficiently small. When in addition μ is sufficiently small, the set $\{w \in [0, \bar{w}] : s(w) > s^{FB}\}$ is convex with $\inf\{w \in [0, \bar{w}] : s(w) > s^{FB}\} > 0$ and $\sup\{w \in [0, \bar{w}] : s(w) > s^{FB}\} = \bar{w}$.

5. Incentive contracts contingent on stock prices

The optimal contract provides short- and long-run incentives, β_t^S and β_t^ℓ , by conditioning the agent's compensation on earnings and asset size. In practice, executive compensation is commonly linked to stock prices (via stock and option grants) and to accounting results (via performance-vesting provisions of these grants and via performance-based bonuses). The use of both stock prices and accounting measures in designing executive compensation has been increasing over time. Stock and option grants constitute a majority of CEO compensation (Edmans et al., 2017). A majority of equity grants have accounting-based performance-vesting provisions with earnings being the most common metric (Bettis et al., 2018). We now show that the optimal contract implied by our model is broadly consistent with these patterns and can be implemented by exposing the manager to stock prices and earnings.

We start with writing the dynamics of earnings and stock prices. The firm's (instantaneous) earnings net of investment cost are given by

$$dE_t = (\alpha s_t - C(s_t, \ell_t))K_t dt + K_t \sigma_X dZ_t^X,$$

while the stock price (which, with full equity financing and the total share supply normalized to one, is equivalent to firm value) evolves according to¹⁶

$$\frac{dP_t}{P_t} = \mu_t^P dt + \Sigma_t^X dZ_t^X + \Sigma_t^K dZ_t^K,$$

where the expressions for μ_t^P , Σ_t^X , and Σ_t^K are given in the Online Appendix. The principal provides the incentives to the manager by choosing the manager's exposures to earnings and stock price changes, respectively, defined by

$$\beta_t^E = \frac{dW_t}{dE_t} \text{ and } \beta_t^P = \frac{dW_t}{dP_t}.$$

¹⁵ To see this, one can decompose $dZ_t^X = \rho dZ_t^K + \sqrt{1 - \rho^2} dZ_t^I$, where $\{Z^I\}$ is a standard Brownian motion, independent of $\{Z^K\}$. Hence, $\mathbb{E}(Z_t^X | Z_t^K) = \rho Z_t^K$ or in differential form $\mathbb{E}(dZ_t^X | dZ_t^K) = \rho dZ_t^K$.

¹⁶ Note that P_t is the per-unit price of equity fully owned by the principal. The manager's value W_t , which could consist of restricted stock units or stock options, is held on an incentive account and is not traded in the market. As such, P_t reflects the market price of traded equity. Since P_t is the investors' firm valuation net of transfers dc to the manager, future dilution by stock vesting or managerial stock option exercise is already accounted for.

The exposures β_t^E and β_t^P are set so as to generate the required short- and long-run incentives under the optimal contract. The Online Appendix derives the following expressions for the exposures implied by the optimal contract

$$\beta_t^P = \lambda_\ell \ell_t \left(\frac{1}{p(w_t) + p'(w_t)(\lambda_\ell \ell_t - w_t)} \right)$$

and

$$\beta_t^E = \lambda_s s_t \left(\frac{p(w_t) - p'(w_t)w_t}{p(w_t) + p'(w_t)(\lambda_\ell \ell_t - w_t)} \right).$$

An appropriate exposure to the firm's stock price (which takes into account the nonlinear relation between stock price and asset size) provides the right amount of long-run incentives. It additionally provides some short-run incentives, as the stock price is also subject to short-run shocks. The exposure to earnings is set to provide the required residual exposure to short-run shocks. This characterization of the optimal contract highlights an important implication of our model: while stock prices account for both short- and long-run shocks to firm value, exposing the manager solely to the firm's stock price cannot in general provide a right mix of short- and long-run incentives. To achieve optimal incentives, the manager also needs to be exposed to short-run accounting performance metrics such as earnings.

Last, we analyze how the manager's exposure to the firm's stock price relative to her exposure to earnings changes over time. To do so, we analyze the ratio:

$$\frac{\beta^P}{\beta^E} = \frac{\lambda_\ell \ell}{\lambda_s s} \times \frac{1}{p(w) - wp'(w)}.$$

This expression shows that the ratio β^P/β^E largely commoves with the ratio of long-term over short-term investment, ℓ/s . The longer the time horizon of the firm's investment policies, the more contingent the manager's compensation is on the firm's stock price. Fig. 3 depicts a typical pattern of β^P/β^E and ℓ/s as functions of w . The manager's compensation depends the least on the firm's stock price after poor past performance and, in particular, under financial distress. In such instances, long-term investment is of little value so that managerial incentives primarily motivate short-term investment and rely on short-term accounting measures. Additionally, stock price volatility sharply increases under distress, whereas earnings volatility remains constant over time. This implies that relatively low β^P generates the required long-term incentives for w close to zero.

Remarkably, after sufficiently strong past performance, when $w = \bar{w}$, it holds that $\beta^P/\beta^E = 1$. As a result, the firm compensates the manager equally based on stock price and earnings, regardless of parameter values reflecting firm fundamentals. Thus, our model implies that time-varying agency conflicts—which are temporarily resolved at \bar{w} —account for the difference between stock-price-based and earnings-based compensation sensitivities.

6. Asymmetric pay in executive compensation

We now turn to analyze the dynamics of incentive provision and show that the optimal contract induces

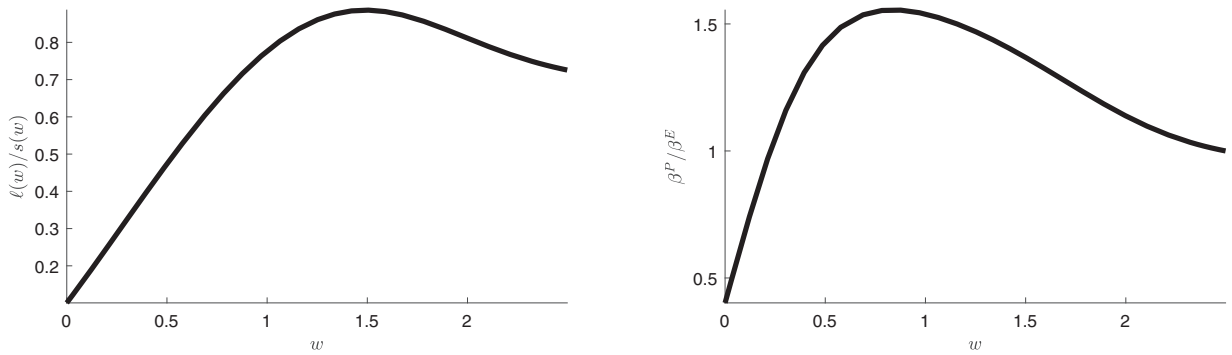


Fig. 3. Numerical illustration of the dynamics of β^p/β^E and ℓ/s . The parameters are $\alpha = 0.25$, $\sigma_K = 0.25$, $\sigma_X = 0.2$, $\rho = 0$, $\mu = 0.025$, $r = 0.046$, $\gamma = 0.048$, $\delta = 0.125$, $\lambda_s = \lambda_\ell = 1$, and $R = 0.2$.

asymmetric pay. We assume throughout the section that the correlation ρ between short- and long-run shocks is zero. For clarity of exposition, we focus on a specification in which the investment cost C is linear:

$$C(s, \ell) = \alpha \lambda_s s + \mu \lambda_\ell \ell. \quad (22)$$

As a consequence, investment follows a bang-bang solution (i.e., either full or no investment is optimal: $s \in \{0, s_{\max}\}$ and $\ell \in \{0, \ell_{\max}\}$). Equivalently, one could also specify that there is a linear adjustment cost to short-run (respectively long-run) investment up to some threshold s_{\max} (respectively ℓ_{\max}) and an infinite adjustment cost afterward. The Online Appendix shows that the results derived in this section also apply when the investment cost is convex.

Corner levels of investment are the only relevant cases in a model with binary effort choice (i.e., $s \in \{0, s_{\max}\}$ or $\ell \in \{0, \ell_{\max}\}$), as in He (2009), or in a model with effort cost functions that are linear in effort levels, as in Biaisi et al. (2007) and DeMarzo et al. (2012). As a result, considering a linear cost function C allows us to directly compare our results with those in the models in which moral hazard is solely over the long or the short run and to clarify what outcomes are unique and novel to our model featuring both types of moral hazard. Finally, we assume that full short- and long-run investment is always optimal so that $s(w) = s_{\max}$ and $\ell(w) = \ell_{\max}$ for all w . Thus with the linear investment cost, the dynamics of optimal incentives are not confounded by changes to investment levels.

When the investment cost is linear, incentive compatibility requires

$$\beta^s \geq \lambda_s \text{ and } \beta^\ell \geq \lambda_\ell.$$

The objective of the principal when choosing the manager's exposure to firm performance is to maximize the value derived from the firm, given a promised payment w to the manager. To do so, the principal equivalently minimizes the agent's exposure to shocks, while maintaining incentive compatibility (see Eq. (11)). Minimizing risk exposure amounts to minimizing the instantaneous variance of the scaled promised payments:

$$\Sigma(w) = (\beta^s \sigma_X)^2 + (\beta^\ell - w)^2 \sigma_K^2 \text{ subject to } \beta^s \geq \lambda_s \text{ and } \beta^\ell \geq \lambda_\ell.$$

This leads to the following result:

Proposition 8 (Asymmetric pay in executive compensation). *When investment costs are linear and full investment is optimal (i.e., $s = s_{\max}$ and $\ell = \ell_{\max}$), we have that*

- (i) Incentives are given by $\beta^s = \lambda_s$ and $\beta^\ell = \lambda_\ell + \max\{0, w - \lambda_\ell\}$.
- (ii) $\beta^\ell(w) > \lambda_\ell$ arises only if $\sigma_X > 0$ and $\sigma_K > 0$.
- (iii) Assume $\sigma_X > 0$ and $\sigma_K > 0$. If $\gamma - r$, ℓ_{\max} or λ_ℓ is sufficiently low, $\bar{w} > \lambda_\ell$ and $\beta^\ell(w) > \lambda_\ell$ for $w \in (\lambda_\ell, \bar{w}]$.

The finding that the incentive compatibility constraint $\beta^s \geq \lambda_s$ in Proposition 8 binds is standard and intuitive. The principal needs to expose the agent to firm performance, but this is costly because this increases the risk of inefficient liquidation. Thus, the principal optimally exposes the agent to as little short-run risk as possible.

The finding that the incentive compatibility constraint $\beta^\ell \geq \lambda_\ell$ does not necessarily bind stems from the fact that the principal optimally wants to expose the manager's continuation payoff to long-run, permanent shocks. Indeed, and as noted above, a positive permanent shock $dZ^K > 0$ has two effects. First, the agent is rewarded for good performance and is promised higher future payments W , which increases the stake w by $\beta^\ell \sigma_K dZ^K$. Second, firm size K grows more than expected, thereby reducing the agent's stake in the firm by $-w \sigma_K dZ^K$. This second effect implies that the agent's stake w is exposed to dilution risk, which can be alleviated using long-run incentives β^ℓ .

When $w > \lambda_\ell$, the principal can fully eliminate dilution risk by setting $\beta^\ell = w$ while maintaining incentive compatibility. Under these circumstances, long-run incentives are effectively costless and the manager is exposed to long-run shocks beyond the level needed to incentivize long-term investment. By contrast, incentive compatibility prevents the principal from eliminating long-run risk when $w < \lambda_\ell$ and $\beta^\ell = \lambda_\ell$. Importantly, when $\lambda_\ell = 0$, there is no agency conflict over the long run and the agent is paid for productivity shocks without any incentive motive, just as in Hoffmann and Pfeil (2010) and DeMarzo et al. (2012).

An important implication of Proposition 8 is that, in our model with dual moral hazard, the agent receives asymmetric performance pay. In particular, the agent is provided minimal long-run incentives $\beta^\ell = \lambda_\ell > w$ for low w and higher-powered long-run incentives $\beta^\ell = w > \lambda_\ell$ after positive past performance, in which case sufficient

slack w has been accumulated. In this region, incentives have option-like features and increase after positive performance. Our findings are consistent with evidence on the asymmetry of pay-for-performance in executive compensation (see, for example, [Garvey and Milbourn, 2006](#); [Francis et al., 2013](#)). In contrast with the suggested explanations, the asymmetry in pay-for-performance is part of an optimal contract and is not due to managerial entrenchment.¹⁷

It should be stressed that the model predicts an asymmetry in pay sensitivity to long-run shocks but not to short-run shocks. With the implementation of the optimal contract using stock prices and earnings (see [Section 5](#)), the asymmetric sensitivity applies to stock prices but not to earnings.

Remarkably, asymmetric performance pay and strong long-run incentives $\beta^\ell \geq \lambda_\ell$ can only arise when $\sigma_X > 0$ and $\sigma_K > 0$ and there is a moral hazard over both time horizons, the short and long run. When $\sigma_X = 0$, the principal does not grant the agent a stake w larger than λ_ℓ in that payouts $dc > 0$ are made before the agent's stake can grow sufficiently large.

To close this section, note that our findings differ from those in [He \(2009\)](#). First, in his model the incentive condition—which corresponds to $\beta^\ell \geq \lambda_\ell$ in our paper's notation—is always tight. This occurs because of $w \leq \bar{w} < \lambda_\ell$ in that the payout boundary cannot exceed λ_ℓ . Intuitively, if it were to happen that $w \geq \lambda_\ell$, the firm could profit by paying the agent some small amount dc . This is because the subsequent increase in liquidation risk is negligible (i.e., of order $o(dc^2)$), while the cost of delaying payments—or equivalently the benefits of paying the agent earlier—are of order $o(dc)$. While in our model risks associated with permanent cash flow shocks dZ^K are also negligible whenever $w \simeq \lambda_\ell$, the firm remains exposed to substantial transitory cash flow shocks. As a consequence, it can be optimal to accumulate even more slack and to eliminate permanent cash flow risk if possible in that $\bar{w} > \lambda_\ell$ and $\beta^\ell = w > \lambda_\ell$ for $w \in (\lambda_\ell, \bar{w}]$.

Second, in [He \(2009\)](#) all risk from permanent cash-flow shocks is eliminated after sufficiently strong past performance only in the extreme case of an equally patient agent and principal. This implies that the firm eventually becomes riskless and the agent works forever. As a result, the first-best outcome can be achieved. In contrast, all risk from permanent cash flow shocks can be eliminated in our model even under the assumption that the agent is more impatient than the principal $\gamma > r$; yet the firm remains exposed to transitory cash flow shocks. As a consequence, only long-run agency conflicts may be temporarily harmless. Indeed, sufficiently adverse cash flow shocks may lower w , drive it below λ_ℓ , and even trigger liquidation, implying that first best will never be reached.

¹⁷ In our model, the agent is essentially paid more for a positive shock than he is punished after a negative shock of the same size. Obviously, this statement is mathematically not exact since the agent's sensitivity to shocks dZ^K is locally symmetric but carries some meaning for shocks over a larger time interval. For a stark intuition, imagine, however, that at time t scaled continuation value equals $w_t = \lambda_\ell - \varepsilon$, and let $\Delta = 2\varepsilon > 0$. A shock $Z_{t+dt}^K - Z_t^K = \Delta > 0$ raises w_{t+dt} beyond λ_ℓ and therefore increases the agent's value by $W_{t+dt} - W_t > 2\varepsilon\lambda_\ell$. In contrast, a shock $Z_{t+dt}^K - Z_t^K = -\Delta < 0$ decreases the agent's value by $2\varepsilon\lambda_\ell$.

7. Robustness and extensions

7.1. Agent's limited wealth

The Online Appendix solves the model under the assumption that the agent has limited wealth and shows that our findings remain qualitatively unchanged in this alternative setting.

7.2. Private investment cost

In the model, we assume that the principal bears the investment cost C while the agent can divert funds for her private consumption. Alternatively, we could also assume that the effort (investment) cost C is private to the manager. In this alternative setting, incentivizing investment s, ℓ requires compensating this private cost to the manager by increasing the growth rate of the agent's scaled continuation value w . Hence, ignoring all other effects, increasing s, ℓ makes w drift up and therefore reduces the likelihood of termination. As a consequence, additional investment/effort cost C is actually beneficial for the principal when $p'(w) > 0$ or, equivalently, when w is low. As shown in [DeMarzo et al. \(2014\)](#) and [Szydlowski \(2019\)](#), this beneficial private cost effect may lead to overinvestment. For completeness, we solve our model with private investment cost in the Appendix and demonstrate that short- and long-termism can arise in this model as well.

In the baseline version of our model, the manager does not finance investment expenditures from her own pockets, and agency conflicts arise because of a misallocation or appropriation of funds allocated to investment. We believe that this setup is more realistic for most real-world environments. In addition, it allows us to clearly identify the drivers of short- and long-termism, compared to a model in which the cost of investment is private (see the Online Appendix for details).

8. Conclusion

We develop a continuous-time agency model in which the agent controls current earnings via short-term investment and firm growth via long-term investment. In this multitasking model, the principal optimally balances the costs and benefits of incentivizing the manager over the short or the long term. As shown in the paper, this can lead to optimal short- or long-termism (i.e., to short- or long-term investment levels above first best levels), depending on the severity of agency conflicts and firm characteristics. The model predicts that the nature of the risks facing firms is key in determining the corporate horizon. We show, for example, that the correlation between shocks to earnings and to firm value leads to externalities between investment choices, which are necessary to generate short-termism. We additionally predict that firm performance should be positively related to the corporate horizon. In particular, firms should become more short-termist after bad performance.

Incentives are provided in the optimal contract by making the agent's compensation contingent on firm performance via exposure to the firm's stock price and earnings.

Because the firm is subject to long-run, permanent shocks, it is optimal to introduce exposure to long-run volatility that is not needed to incentivize effort in the contract. In our model with multitasking, however, the principal needs to incentivize the manager to exert long-run effort. This generates the distinct prediction that extra pay-for-performance is introduced and the manager's wealth is fully exposed to permanent shocks only when her stake in the firm is large enough. Notably, when her stake is low, the extra pay-for-performance effect is shut down and the incentive compatibility constraint is binding. In other words, positive permanent shocks lead to additional pay-for-performance, and negative permanent shocks eventually eliminate this extra sensitivity to performance implied by the optimal contract. Our model therefore provides a rationale for the asymmetry of pay-for-performance observed in the executive compensation data.

Our model can be extended in several directions. First, we assume a risk-neutral manager who is optimally compensated by lumpy wages. Incorporating risk aversion would imply smooth, but time-varying, wage payments, which is arguably more consistent with empirical observations. Second, in our model, the outcome of long-term investment realizes instantaneously. It would be interesting to study a setup in which the impact of the manager's long-term investment decisions gradually realizes over time, giving rise to more involved incentive structures.

Appendix

Without loss in generality, we consider throughout the whole appendix that the depreciation rate of capital δ equals zero. To ensure the problem is well-behaved, we impose the following regularity conditions:

(a) Square integrability of the payout process $\{C\}$:

$$\mathbb{E} \left[\left(\int_0^\tau e^{-\gamma s} dC_s \right)^2 \right] < \infty.$$

- (b) The processes $\{s\}$ and $\{\ell\}$ are of bounded variation.
(c) The sensitivities $\{\beta^s\}$ and $\{\beta^\ell\}$ are almost surely bounded so that there exists $M > 0$ with $\mathcal{P}(\beta_t^K < M) = 1$ for each $t \geq 0$ and $K \in \{s, \ell\}$. We make this assumption for purely technical reasons and can choose $M < \infty$ arbitrarily large (enough) such that this constraint never binds at the optimum.

A.1. Proof of Proposition 1

Proof. The first-best investment levels (s^{FB}, ℓ^{FB}) maximize

$$\hat{p}(s, \ell) = \frac{1}{r + \delta - \mu\ell} [\alpha s - C(s, \ell)].$$

For the case of quadratic cost, straightforward calculations lead to the desired expressions for s^{FB} , ℓ^{FB} and $p^{FB} \equiv \hat{p}(s^{FB}, \ell^{FB})$, where ℓ^{FB} satisfies the relation $\mu p^{FB} = C_\ell(s^{FB}, \ell^{FB})$. \square

A.2. Proof of Proposition 2

A.2.1. Auxiliary results

We first show that each investment path $(\{s\}, \{\ell\})$ induces a probability measure under certain conditions. To begin with, fix a probability measure \mathcal{P}^0 such that

$$dX_t = \sigma_X K_t d\tilde{W}_t^X \text{ and } dK_t = \sigma_K K_t d\tilde{W}_t^K$$

with correlated standard Brownian motions $\{\tilde{W}_t^X\}, \{\tilde{W}_t^K\}$ under this measure, both progressive with respect to \mathbb{F} . The measure \mathcal{P}^0 corresponds to perpetual zero investment. Define $\tilde{\mathbf{W}}_t \equiv (\tilde{W}_t^X, \tilde{W}_t^K)'$ and let the (unconditional) covariance matrix of $\tilde{\mathbf{W}}_t$ under \mathcal{P}^0 be¹⁸

$$\mathbb{V}^0(\tilde{\mathbf{W}}_t) = \mathbb{E}^0(\tilde{\mathbf{W}}_t \tilde{\mathbf{W}}_t') = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \times t \equiv \mathbf{C}t.$$

In this equation, $\mathbb{V}^0(\cdot)$ denotes the variance operator with respect to the measure \mathcal{P}^0 . Let us employ a Cholesky decomposition to write $\mathbf{M}^{-1}(\mathbf{M}^{-1})' = \mathbf{C}$ or equivalently $\mathbf{M}'\mathbf{M} = \mathbf{C}^{-1}$ for an invertible, deterministic matrix \mathbf{M} . Observe that

$$\begin{aligned} \mathbb{V}^0(\mathbf{M}\tilde{\mathbf{W}}_t) &= \mathbf{M}\mathbb{E}^0(\tilde{\mathbf{W}}_t \tilde{\mathbf{W}}_t')\mathbf{M}' \\ &= \mathbf{M}\mathbf{C}\mathbf{M}' \cdot t = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}' \cdot t = \mathbf{I} \cdot t, \end{aligned}$$

where $\mathbf{I} \in \mathbb{R}^{2 \times 2}$ denotes the identity matrix. Because the two components of $\tilde{\mathbf{W}}_t$ are jointly normal and uncorrelated, they are also independent in that the process $\{\tilde{\mathbf{W}}_t^T\} \equiv \{\mathbf{M}\tilde{\mathbf{W}}_t\}$ follows a bidimensional standard Brownian motion. We can now apply Girsanov's theorem to $\{\tilde{\mathbf{W}}_t^T\}$, where all components, by definition, are mutually independent.

As a first step, we define

$$\Theta_t = \Theta_t(s, \ell) \equiv \left(\frac{\alpha s_t}{\sigma_X}, \frac{\mu \ell_t}{\sigma_K} \right)' \text{ and}$$

$$\tilde{\Theta}_t = \tilde{\Theta}_t(s, \ell) \equiv \mathbf{M}\Theta_t(s, \ell).$$

Further, let

$$\Gamma'_t = \Gamma'_t(s, \ell) \equiv \exp \left(\int_0^t \tilde{\Theta}_u \cdot d\tilde{\mathbf{W}}_u^T - \frac{1}{2} \int_0^t \|\tilde{\Theta}_u\|^2 du \right),$$

where $\|\cdot\|$ denotes the Euclidean norm in \mathbb{R}^2 and

$$\int_0^t \tilde{\Theta}_u \cdot d\tilde{\mathbf{W}}_u^T = \int_0^t \sum_{i=1,2} \tilde{\Theta}_{u,i} d\tilde{W}_{u,i}^T = \sum_{i=1,2} \int_0^t \tilde{\Theta}_{u,i} d\tilde{W}_{u,i}^T.$$

Throughout the paper, we will assume that the processes $\{s\}, \{\ell\}$ are such that the Novikov condition is satisfied in that

$$\mathbb{E}^0 \left[\exp \left(\frac{1}{2} \int_0^\tau \|\tilde{\Theta}_t\|^2(s, \ell) dt \right) \right] < \infty.$$

In fact, our regularity conditions imply the Novikov condition. Then, $\{\Gamma'\}$ follows a martingale under \mathcal{P}^0 rather than just a local martingale. Due to $\mathbb{E}^0 \Gamma'_t = \mathbb{E}^0 \Gamma'_0 = 1$, the process $\{\Gamma'\}$ is a progressive density process and defines

¹⁸ For a matrix-valued random variable $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^{n \times k}$, we denote the transposed random variable by $\mathbf{Y}' : \Omega \rightarrow \mathbb{R}^{k \times n}$.

the probability measure $\mathcal{P}^{s,\ell}$ via the Radon-Nikodym derivative

$$\left(\frac{d\mathcal{P}^{s,\ell}}{d\mathcal{P}^0}\right)\Big|_{\mathcal{F}_t} = \Gamma_t'$$

By Girsanov's theorem,

$$\left\{ \mathbf{Z}_t^T = \tilde{\mathbf{W}}_t^T - \int_0^t \tilde{\Theta}_u du : t \geq 0 \right\}$$

follows a bidimensional, standard Brownian motion under the measure $\mathcal{P}^{s,\ell}$. The linearity of the (Riemann) integral implies

$$\begin{aligned} \mathbf{M} \left(\begin{pmatrix} Z_t^X \\ Z_t^K \end{pmatrix} \right) &\equiv \mathbf{Z}_t^T = \mathbf{M} \left(\tilde{\mathbf{W}}_t - \int_0^t \Theta_u du \right) \\ &= \mathbf{M} \left(\begin{pmatrix} \tilde{W}_t^X \\ \tilde{W}_t^K \end{pmatrix} + \begin{pmatrix} \int_0^t \Theta_{u,1} du \\ \int_0^t \Theta_{u,2} du \end{pmatrix} \right). \end{aligned}$$

Therefore, for each $t \geq 0$,

$$dZ_t^X \equiv \frac{dX_t - K_t \alpha s_t dt}{K_t \sigma_X} \text{ and } dZ_t^K \equiv \frac{dK_t - K_t \mu \ell_t dt}{K_t \sigma_K}$$

are the increments of a standard Brownian motion under $\mathcal{P}^{s,\ell}$ with instantaneous correlation ρdt . In the following, we say the measure $\mathcal{P}^{s,\ell}$ is induced by the processes $\{s\}$, $\{\ell\}$. Note that all probability measures of the family $\{\mathcal{P}^{s,\ell}\}_{\{s\},\{\ell\}}$ are mutually equivalent in that they share the same null sets.

A.2.2. Proof of Proposition 2.1

Proof. Consider an incentive compatible contract $\Pi \equiv (\{C\}, \{s\}, \{\ell\}, \tau)$. Further, assume in the following without loss of generality that \mathbb{F} is the filtration generated by $\{X\}$, $\{K\}$ in that $\mathcal{F}_t = \sigma(X_s, K_s : 0 \leq s \leq t)$. Then, the agent's continuation utility at time t (under the principal's information) is defined by

$$\begin{aligned} W_t(\Pi) &\equiv \mathbb{E}_t^{s,\ell} \left[\int_t^\tau e^{-\gamma(z-t)} dC_z + \int_t^\tau e^{-\gamma(z-t)} K_z (C(s_z, \ell_z) \right. \\ &\quad \left. - C(\hat{s}_z, \hat{\ell}_z)) dz \right], \end{aligned}$$

where $\mathbb{E}_t^{s,\ell}(\cdot)$ denotes the conditional expectation given \mathcal{F}_t , taken under the probability measure $\mathcal{P}^{s,\ell}$ induced by $\{s\}$ and $\{\ell\}$. Define for $t \leq \tau$:

$$\begin{aligned} \Gamma_t(\Pi) &\equiv \mathbb{E}_t^{s,\ell} [W_0(\Pi)] \\ &= \int_0^t e^{-\gamma z} dC_z + \int_0^t e^{-\gamma z} K_z (C(s_z, \ell_z) - C(\hat{s}_z, \hat{\ell}_z)) dz \\ &\quad + e^{-\gamma t} W_t(\Pi). \end{aligned} \tag{A.1}$$

By construction, $\{\Gamma_t(\Pi) : 0 \leq t \leq \tau\}$ is a square-integrable martingale under $\mathcal{P}^{s,\ell}$, progressive with respect to \mathbb{F} . In the following, we will invoke incentive compatibility (i.e., $s_t = \hat{s}_t$, $\ell_t = \hat{\ell}_t$) whenever no confusion is likely to arise.

Next, observe that any sigma-algebra is invariant under an injective transformation of its generator. In particular, let $\mathbf{M} \in \mathbb{R}^{2 \times 2}$ an invertible, deterministic matrix with $\det(\mathbf{M}) \neq 1$, and note that

$$\begin{aligned} \mathcal{F}_t &= \sigma(X_s, K_s : s \leq t) = \sigma(Z_s^1, Z_s^2 : s \leq t) = \sigma(\mathbf{Z}_s : s \leq t) \\ &= \sigma(\mathbf{M} \cdot \mathbf{Z}_s : s \leq t) \end{aligned}$$

with $\mathbf{Z}_t \equiv (Z_t^1, Z_t^2)'$. Here,

$$dZ_t^1 \equiv \frac{dX_t - K_t \alpha s_t dt}{K_t \sigma_X} \text{ and } dZ_t^2 \equiv \frac{dK_t - K_t \mu \ell_t dt}{K_t \sigma_K} \tag{A.2}$$

are the increments of a standard Brownian motion under the probability measure $\mathcal{P}^{s,\ell}$. Note that $dZ_t^1 = dZ_t^X$ and $dZ_t^2 = dZ_t^K$ whenever $a_t = \hat{a}_t$ for all $a \in \{s, \ell\}$.

As in the previous section, let the covariance matrix $\mathbb{V}(\mathbf{Z}_t) = \mathbf{C}t$ and employ a Cholesky decomposition $\mathbf{M}'\mathbf{M} = \mathbf{C}^{-1}$. We have already shown that $\{\mathbf{Z}_t^T \equiv \mathbf{M}\mathbf{Z}_t : 0 \leq t \leq \tau\}$ follows a bidimensional, standard Brownian motion under $\mathcal{P}^{s,\ell}$, where both components are mutually independent. By the martingale representation theorem (see, e.g., [Shreve, 2004](#)), there exists a bidimensional process $\{\mathbf{b}_t\}_{t \geq 0}$, progressively measurable with respect to \mathbb{F} such that

$$\begin{aligned} d\Gamma_t(\Pi) &= e^{-\gamma t} \mathbf{b}_t' \cdot d\mathbf{Z}_t^T = e^{-\gamma t} \mathbf{b}_t' \cdot \mathbf{M}\mathbf{M}^{-1} \cdot d\mathbf{Z}_t^T \\ &= e^{-\gamma t} K_t (\beta_t^s \sigma_X dZ_t^1 + \beta_t^\ell \sigma_K dZ_t^2), \end{aligned}$$

where we exploit the linearity of the Itô integral—i.e., $d(\mathbf{M}\mathbf{Z}_t^T) = \mathbf{M}d\mathbf{Z}_t^T$ —and set $(\beta_t^s \sigma_X, \beta_t^\ell \sigma_K) \equiv \mathbf{b}_t' \mathbf{M} / K_t$ for all t . Combining with [Eq. \(A.1\)](#), one can verify that

$$\begin{aligned} d\Gamma_t(\Pi) &= e^{-\gamma t} K_t (\beta_t^s \sigma_X dZ_t^1 + \beta_t^\ell \sigma_K dZ_t^2) \\ &= e^{-\gamma t} - \gamma e^{-\gamma t} W_t(\Pi) dt + e^{-\gamma t} dW_t(\Pi), \end{aligned}$$

and thus [Eq. \(7\)](#) holds after rearranging. Indeed, since the right-hand side of [Eq. \(7\)](#) satisfies a Lipschitz condition under the usual regularity conditions (i.e., square integrability of $\{C\}$ and $\{s\}$, $\{\ell\}$ of bounded variation), $\{W\}$ is the unique strong solution to the stochastic differential [Eq. \(7\)](#).

Next, we provide necessary and sufficient conditions for the contract Π to be incentive compatible. For this purpose, let the recommended investment processes $\{s\}$ and $\{\ell\}$ and the expected payoff of the agent at time t be W_t when following the recommended strategy from time t onwards. Further, let $\{\hat{s}\}$ and $\{\hat{\ell}\}$ represent the actual investment processes, which may in principle differ from $\{s\}$ and $\{\ell\}$. We have

$$W_t \equiv \mathbb{E}_t^{s,\ell} \left[\int_t^\tau e^{-\gamma(z-t)} dC_z \right].$$

Recall that $\mathbb{E}_t^{s,\ell}$ denotes the expectation, conditional on the filtration \mathcal{F}_t , taken under the probability measure $\mathcal{P}^{s,\ell}$. As shown above, the process $\{W\}$ solves the stochastic differential equation:

$$dW_t = \gamma W_t dt + \beta_t^s (dX_t - K_t \alpha s_t dt) + \beta_t^\ell (dK_t - K_t \mu \ell_t dt) - dC_t.$$

We can rewrite this stochastic differential equation as

$$\begin{aligned} dW_t + dC_t &= \gamma W_t dt + K_t \beta_t^s [\alpha (\hat{s}_t - s_t) dt + \sigma_X dZ_t^X] \\ &\quad + K_t \beta_t^\ell [\mu (\hat{\ell}_t - \ell_t) dt + \sigma_K dZ_t^K] \end{aligned}$$

with

$$dZ_t^X \equiv \frac{dX_t - K_t \alpha \hat{s}_t dt}{K_t \sigma_X} \text{ and } dZ_t^K \equiv \frac{dK_t - K_t \mu \hat{\ell}_t dt}{K_t \sigma_K}.$$

Girsanov's theorem implies now that $dZ_t^X \equiv \frac{dX_t - K_t \alpha \hat{s}_t dt}{K_t \sigma_X}$ and $dZ_t^K \equiv \frac{dK_t - K_t \mu \hat{\ell}_t dt}{K_t \sigma_K}$ are the increments of a standard Brownian motion under the measure $\mathcal{P}^{s,\ell}$ induced by $(\{\hat{s}\}, \{\hat{\ell}\})$.

Next, define the auxiliary gain process

$$\begin{aligned} g_t &= g_t(\{\hat{s}\}, \{\hat{\ell}\}) \\ &\equiv \int_0^t e^{-\gamma z} dC_z - \int_0^t e^{-\gamma z} K_z(C(\hat{s}_z, \hat{\ell}_z) - C(s_z, \ell_z)) dz \\ &\quad + e^{-\gamma t} W_t, \end{aligned}$$

and recall that $W_\tau = 0$. Now, note that the agent's actual expected payoff under the strategy $(\{\hat{s}\}, \{\hat{\ell}\})$ reads

$$\begin{aligned} W'_0 &\equiv \max_{\{\hat{s}\}, \{\hat{\ell}\}} \mathbb{E}^{s, \ell} \left[\int_0^\tau e^{-\gamma z} dC_z - \int_0^\tau e^{-\gamma z} K_z(C(\hat{s}_z, \hat{\ell}_z) \right. \\ &\quad \left. - C(s_z, \ell_z)) dz \right] \\ &= \max_{\{\hat{s}\}, \{\hat{\ell}\}} \mathbb{E}^{s, \ell} [g_\tau(\{\hat{s}\}, \{\hat{\ell}\})]. \end{aligned}$$

We obtain

$$\begin{aligned} e^{\gamma t} dg_t &= K_t [C(s_t, \ell_t) - C(\hat{s}_t, \hat{\ell}_t)] dt \\ &\quad + K_t [\alpha \beta_t^s (\hat{s}_t - s_t) + \mu \beta_t^\ell (\hat{\ell}_t - \ell_t)] dt \\ &\quad + K_t [\beta_t^s \sigma_X dZ_t^X + \beta_t^\ell \sigma_K dZ_t^K] \\ &\equiv \mu_t^g dt + K_t [\beta_t^s \sigma_X dZ_t^X + \beta_t^\ell \sigma_K dZ_t^K]. \end{aligned}$$

It is now easy to see that when choosing $\hat{s}_t = s_t$, $\hat{\ell}_t = \ell_t$, the agent can always ensure that $\mu_t^g = 0$, in which case $\{g_z\}_{z \geq 0}$ follows a martingale under $\mathcal{P}^{s, \ell}$. Hence,

$$W'_0 = \max_{\{\hat{s}\}, \{\hat{\ell}\}} \mathbb{E}^{s, \ell} [g_\tau(\{\hat{s}\}, \{\hat{\ell}\})] \geq \mathbb{E}^{s, \ell} [g_\tau(\{s\}, \{\ell\})] = W_0.$$

The inequality is strict if and only if there exist processes $\{\hat{s}\}, \{\hat{\ell}\}$ and a stopping time τ' with $\mathcal{P}^{s, \ell}(\tau' < \tau) > 0$ such that $\mu_{\tau'}^g > 0$. This arises because then there also exists a set $\mathcal{A} \subseteq [0, \tau] \times \Omega$ with

$$\mu_t^G(\omega) > 0 \text{ for all } (t, \omega) \in \mathcal{A} \text{ and } \mathcal{L} \otimes \mathcal{P}^{s, \ell}(\mathcal{A}) > 0,$$

where \mathcal{L} is the Lebesgue measure on the Lebesgue sigma-algebra in \mathbb{R} . Because $\mathcal{P}^{s, \ell}(\tau < \infty)$ for all admissible $\{\hat{s}\}, \{\hat{\ell}\}$ it follows that $e^{-\gamma t} \mu_t^G(\omega) > 0$ for all $(t, \omega) \in \mathcal{A}$. Hence,

$$W'_0 \geq \int_{\mathcal{A}} e^{-\gamma z} \mu_z^G(\omega) d(\mathcal{L}(z) \otimes \mathcal{P}^{s, \ell}(\omega)) + W_0 > W_0.$$

In case $W'_0 > W_0$, either $\hat{s}_z(\omega) \neq s_z(\omega)$ or $\hat{\ell}_z(\omega) \neq \ell_z(\omega)$ on the set \mathcal{A} , which has positive measure so that Π is not incentive compatible.

Hence, for Π to be incentive compatible, it must, for all $t \geq 0$ (almost surely), hold that

$$\max_{\hat{s}_t, \hat{\ell}_t} \{\alpha \beta_t^s (\hat{s}_t - s_t) + \mu \beta_t^\ell (\hat{\ell}_t - \ell_t) + [C(s_t, \ell_t) - C(\hat{s}_t, \hat{\ell}_t)]\} = 0,$$

or equivalently

$$\begin{aligned} (s_t, \ell_t) \in \arg \max_{\hat{s}_t, \hat{\ell}_t} \{ &\alpha \beta_t^s (\hat{s}_t - s_t) + \mu \beta_t^\ell (\hat{\ell}_t - \ell_t) \\ &+ [C(s_t, \ell_t) - C(\hat{s}_t, \hat{\ell}_t)] \}, \end{aligned}$$

for given β_t^s, β_t^ℓ . After going through the maximization, we obtain that this is satisfied if $C_s(s_t, \ell_t) = \beta_t^s \alpha$ and

$C_\ell(s_t, \ell_t) = \beta_t^\ell \mu$, in case $(s_t, \ell_t) \in (0, s_{\max}) \times (0, \ell_{\max})$. If $a_t \in \{s_t, \ell_t\}$ is not interior in that $a_t = a_{\max}$ for $a \in \{s, \ell\}$, then $a_t = \hat{a}_t$ solves the above maximization problem if and only if $\beta_t^s \alpha \geq C_s(s_t, \ell_t)$ if $a_t = s_t$, or $\beta_t^\ell \mu \geq C_\ell(s_t, \ell_t)$ if $a_t = \ell_t$. It evidently suffices here to consider first-order optimality so that the result follows. \square

A.2.3. Proof of Proposition 2.2

In this section, we proceed as follows. First, we represent $P(W, K)$ as a twice continuously differentiable solution of a HJB equation and then show that there exists a function $p \in C^2$ such that $P(W, K) = Kp(w)$ and $p(w)$ solves the scaled HJB Eq. (11). Second, we verify that $P(W, K)$ or equivalently $p(w)$ with corresponding payout threshold \bar{w} and $w_0 = w^*$ characterizes indeed the optimal contract Π^* . Since we focus on incentive compatible contracts, we will work in the following—unless otherwise mentioned—with the measure $\mathcal{P}^{s^*, \ell^*}$ induced by optimal investment $(\{s^*\}, \{\ell^*\})$. For convenience, we will denote this measure by \mathcal{P} if no confusion is likely to arise. We follow an analogous convention concerning the expectation operator where we just write $\mathbb{E}_t(\cdot)$ instead of $\mathbb{E}^{s^*, \ell^*}(\cdot | \mathcal{F}_t)$.

Scaling of the value function

Given the optimal control and stopping problem (6), suppose that the principal's value function $P(W, K)$ satisfies the HJB equation

$$\begin{aligned} rP &= \max_{s, \ell, \beta^s, \beta^\ell} \left\{ \alpha s K - KC(s, \ell) + P_W \gamma W + P_K \mu \ell K \right. \\ &\quad + \frac{1}{2} (P_{WW} [(\beta^s \sigma_X K)^2 + (\beta^\ell \sigma_K K)^2 + 2\rho \sigma_X \sigma_K K^2 \beta^s \beta^\ell] \\ &\quad \left. + P_{KK} (\sigma_K K)^2 + 2P_{WK} [(\sigma_K K)^2 \beta^\ell + \rho \sigma_X \sigma_K K^2 \beta^s] \right\} \end{aligned}$$

in some region $S \subset \mathbb{R}^2$, subject to the boundary conditions

$$P(0, K) = RK, P(W, 0) = 0, P_W(\bar{W}, K) = -1, P_{WW}(\bar{W}, K) = 0.$$

Here, $\bar{W} \equiv \bar{W}(K) = \bar{w}K$ parametrizes the boundary of S on which $W, K > 0$. Taking the guess $P(W, K) = p(W/K)K$ for some function $p \in C^2$, we obtain

$$\begin{aligned} P_W &= p'(w), P_K = p(w) - wp'(w), P_{WK} = -w/K p''(w), \\ P_{WW} &= p''(w)/K, p_{KK} = w^2/K p''(w), \end{aligned}$$

which implies the HJB Eq. (11) and its boundary conditions.

In the following, we will assume that Eq. (11) admits a unique, twice continuously differentiable solution $p(\cdot)$ on $[0, \bar{w}]$. A formal existence proof is beyond the scope of the paper and is therefore omitted.¹⁹

We first rewrite the principal's problem (6) in a convenient manner. Let

$$\Psi_t = (\rho \sigma_K t, \sigma_K t)' \text{ and } \tilde{\Psi}_t = \mathbf{M} \Psi_t,$$

¹⁹ Indeed, the possible discontinuities of the functions $s(\cdot), \ell(\cdot)$ cause technical complications. If s_{\max}, ℓ_{\max} are sufficiently large, this problem is not present anymore. Then, the existence and uniqueness of the solution follow from the Picard-Lindelöf theorem since the required Lipschitz condition is evidently satisfied.

where $\mathbf{M}'\mathbf{M} = \mathbf{C}^{-1}$ and \mathbf{C}_t is the covariance matrix of (Z_t^X, Z_t^K) . Next, define the equivalent, auxiliary probability measure $\tilde{\mathcal{P}}$ according to the Radon-Nikodym derivative

$$\left(\frac{d\tilde{\mathcal{P}}}{d\mathcal{P}}\right)\Big|_{\mathcal{F}_t} \equiv \exp\left\{\int_0^t \tilde{\Psi}_u du - \frac{1}{2}\int_0^t \|\tilde{\Psi}_u\|^2 du\right\}. \quad (\text{A.3})$$

By arguments similar to those in Appendix A.2.1, Girsanov's theorem implies that

$$\tilde{Z}_t^X = Z_t^X - \rho\sigma_K t \text{ and } \tilde{Z}_t^K = Z_t^K - \sigma_K t$$

are both standard Brownian motions with correlation ρt under $\tilde{\mathcal{P}}$. An application of Itô's lemma consequently yields that the scaled continuation value $\{w\}$ evolves according to

$$dw_t + dc_t = (\gamma - \mu\ell_t)w_t dt + \beta_t^s \sigma_X d\tilde{Z}_t^X + (\beta_t^\ell - w_t)\sigma_K d\tilde{Z}_t^K$$

under $\tilde{\mathcal{P}}$. Finally, for $\psi_t \equiv rt - \mu \int_0^t \ell_z dz$ we can rewrite the principal's problem (6) as

$$\max_{\{c\}, \{s\}, \{\ell\}, w^*} \mathbb{E}\left[\int_0^\tau e^{-\psi_t} (\alpha s_t - C(s_t, \ell_t)) dt - \int_0^\tau e^{-\psi_t} dc_t + e^{-\psi_\tau} R \Big| w_0 = w^*\right],$$

where the expectation $\mathbb{E}[\cdot]$ is taken under the equivalent, auxiliary measure $\tilde{\mathcal{P}}$. Here, $dc_t \equiv dC_t/K_t = \max\{w_t - \bar{w}, 0\}$. The stated integral expression is implied by following lemma.

Lemma 1. Suppose $\{w\}$ is the unique, strong solution to the stochastic differential equation

$$dw_t = \delta_t dt + \Delta_t w_t dt - dc_t + (\beta_t^\ell - w_t)\sigma_K dZ_t^K + \beta_t^s \sigma_X dZ_t^X$$

for $t \leq \tau$, standard Brownian motions $\{Z^X\}$, $\{Z^K\}$ with correlation ρ and progressive processes $\{\delta\}$, $\{\Delta\}$, $\{\beta^\ell\}$, $\{\beta^s\}$ of bounded variation.²⁰ Assume that $dw_t = 0$ for $t > \tau$ where $\tau = \min\{t \geq 0 : w_t = 0\}$. Furthermore, $dc_t = \max\{w_t - \bar{w}, 0\}$ with threshold $\bar{w} > 0$. Let now $g : [0, \bar{w}] \rightarrow \mathbb{R}$ of bounded variation. Then the twice continuously differentiable function $f : [0, \bar{w}] \rightarrow \mathbb{R}$ (i.e., $f \in C^2$) solves the differential equation

$$r_t f(w_t) = g(w_t) + f'(w_t)[\delta_t + \Delta_t w_t] + f''(w_t)\left[\sigma_K^2 (\beta_t^\ell - w_t)^2 + (\beta_t^s \sigma_X)^2 + 2\rho\sigma_X\sigma_K\beta_t^s (\beta_t^\ell - w_t)\right] \quad (\text{A.4})$$

with boundary conditions $f(0) = R$, $f'(\bar{w}) = -1$ if and only if

$$f(w) = \mathbb{E}\left[\int_0^\tau e^{-\int_0^t r_u du} g(w_t) dt - \int_0^\tau e^{-\int_0^t r_u du} dc_t + e^{-\int_0^\tau r_u du} R \Big| w_0 = w\right]$$

for a progressive discount rate process $\{r\}$ of bounded variation.

²⁰ We call a process $\{Y\}$ "of bounded variation" if it can be written as the difference of two almost surely increasing processes. Similarly, a function $F \in \mathbb{R}^{[a,b]}$ is called "of bounded variation" if it can be written as the difference of two increasing functions on the interval $[a, b]$.

Proof. Suppose $f(\cdot)$ solves Eq. (A.4). Define

$$h_t \equiv \int_0^t e^{-\int_0^s r_u du} g(w_z) dz - \int_0^t e^{-\int_0^s r_u du} dc_z + e^{-\int_0^t r_u du} f(w_t).$$

Applying Itô's lemma, we obtain

$$e^{\int_0^t r_u dt} dh_t = \left\{g(w_t) - r_t f(w_t) + \frac{f''(w_t)}{2}\left[\sigma_K^2 (\beta_t^\ell - w_t)^2 + (\beta_t^s \sigma_X)^2 + 2\rho\sigma_X\sigma_K\beta_t^s (\beta_t^\ell - w_t)\right] + f'(w_t)(\delta_t + \Delta_t w_t)\right\} dt - \left[(1 + f'(w_t))dc_t\right] + f'(w_t)\left[dZ_t^X \beta_t^s \sigma_X + dZ_t^K (\beta_t^\ell - w_t)\sigma_K\right].$$

The first term in curly brackets equals zero because $f(\cdot)$ solves Eq. (A.4). Since $f'(\bar{w}) = -1$ and $dc_t = 0$ for all $w_t \leq \bar{w}$, the second term in square brackets equals also zero, and therefore $\{h\}$ follows a martingale up to time τ . As a result, we have

$$f(w_0) = f(w) = h_0 = \mathbb{E}[h_\tau] = \mathbb{E}\left[\int_0^\tau e^{-\int_0^t r_s ds} g(w_t) dt - \int_0^\tau e^{-\int_0^t r_s ds} dc_t + e^{-\int_0^\tau r_s ds} R \Big| w_0 = w\right].$$

The result follows. □

Verification

Proof. Next, we verify the optimality of the contract Π^* among all contracts Π satisfying incentive compatibility. To do so, we show that the principal obtains under any contract $\Pi \in \mathbb{IC}$ at most (scaled) payoff $\tilde{G}(\Pi)/K \leq p(w^*)$ with equality if and only if $\Pi = \Pi^*$. Here, $p(\cdot)$ solves the HJB Eq. (11) with corresponding payout threshold \bar{w} and $w_0 = w^*$.

Consider any incentive-compatible contract $\Pi \equiv (\{C\}, \{s\}, \{\ell\}, \tau)$. For any $t \leq \tau$, define its auxiliary gain process G as

$$G_t(\Pi) = \int_0^t e^{-ru} (dX_u - C(s_u, \ell_u)) du - \int_0^t e^{-ru} dC_u + e^{-rt} P(W_t, K_t),$$

where the agent's continuation payoff evolves according to Eq. (7). Recall that $w_t = \frac{W_t}{K_t}$ and $P(W_t, K_t) = K_t p(w_t)$. Itô's lemma implies that for $t \leq \tau$

$$e^{rt} \frac{dG_t(\Pi)}{K_t} = \left[- (r - \mu\ell_t)p(w_t) + \alpha s_t - C(s_t, \ell_t) + p'(w_t)w_t(\gamma - \mu\ell_t) + \frac{p''(w_t)}{2}\left[(\beta_t^s \sigma_X)^2 + \sigma_K^2 (\beta_t^\ell - w_t)^2 + 2\rho\sigma_X\sigma_K\beta_t^s (\beta_t^\ell - w_t)\right] \right] dt - (1 + p'(w_t))dc_t + \sigma_K(p(w_t) + p'(w_t)(\beta_t^\ell - w_t))dZ_t^K + \sigma_X(1 + \beta_t^s p'(w_t))dZ_t^X.$$

Under the optimal investment and incentives, the first term in square brackets stays at zero always. Other investment and incentive policies will make this term negative

(owing to the concavity of p). The second term is non-positive since $p'(w_t) \geq -1$ but equal to zero under the optimal contract. Therefore, for the auxiliary gain process, we have

$$dG_t(\Pi) = \mu_G(t)dt + e^{-rt}K_t[\sigma_K(p(w_t) + p'(w_t)(\beta_t^\ell - w_t))dZ_t^K + \sigma_X(1 + \beta_t^s p'(w_t))dZ_t^X],$$

where $\mu_G(t) \leq 0$. Due to our assumption of bounded sensitivities $\{\beta^s\}, \{\beta^\ell\}$, it follows that

$$\begin{aligned} & \mathbb{E}\left(\int_0^t e^{-ru}(p(w_u) + p'(w_u)(\beta_u^\ell - w_u))dZ_u^K\right) \\ &= \mathbb{E}\left(\int_0^t e^{-ru}(1 + \beta_u^s p'(w_u))dZ_u^K\right) = 0, \end{aligned}$$

which implies that $\{G_t\}_{t \geq 0}$ follows a supermartingale. Furthermore, under Π , investors' expected payoff is

$$\begin{aligned} \tilde{G}(\Pi) &\equiv \mathbb{E}\left[\int_0^\tau e^{-ru}(dX_u - C(s_u, \ell_u))du\right. \\ &\quad \left. - \int_0^\tau e^{-ru}dC_u + e^{-r\tau}RK_\tau\right]. \end{aligned}$$

As a result, we have

$$\begin{aligned} \tilde{G}(\Pi) &= \mathbb{E}[G_\tau(\Pi)] \\ &= \mathbb{E}\left[G_{\tau \wedge t}(\Pi) + \mathbf{1}_{\{t \leq \tau\}}\left(\int_t^\tau e^{-rs}(dX_s - dC_s - C(s_s, \ell_s))ds\right.\right. \\ &\quad \left.\left.+ e^{-r\tau}RK_\tau - e^{-rt}P(W_t, K_t)\right)\right] \\ &= \mathbb{E}[G_{\tau \wedge t}(\Pi)] \\ &\quad + e^{-rt}\mathbb{E}\left[\mathbf{1}_{\{t \leq \tau\}}\mathbb{E}_t\left(\int_t^\tau e^{-r(s-t)}(dX_s - dC_s - C(s_s, \ell_s))ds\right.\right. \\ &\quad \left.\left.+ e^{-r(\tau-t)}RK_\tau - P(W_t, K_t)\right)\right] \\ &\leq G_0 + e^{-rt}\mathbb{E}[P^{FB}(K_t) - W_t - P(W_t, K_t)] \\ &\leq G_0 + e^{-rt}(p^{FB} - R)\mathbb{E}[K_t], \end{aligned}$$

where $p^{FB} \equiv \frac{P^{FB}(K_t)}{K_t}$ is the (scaled) first-best value. The inequalities follow from the supermartingale property of G_t , the fact that the value of the firm with agency is below first best, and the fact that $p^{FB} - w - p(w) \leq p^{FB} - R$. Since $\mu_{\ell_{\max}} < r$, it follows that $\lim_{t \rightarrow \infty} e^{-rt}\mathbb{E}[K_t] = 0$. Therefore, letting $t \rightarrow \infty$ yields $\tilde{G}(\Pi) \leq G_0 \equiv P(W_0, K_0) = p(w_0)K_0$ for all incentive compatible contracts. For the optimal contract Π^* , the investors' payoff $\tilde{G}(\Pi^*)$ achieves $P(W_0, K_0) = p(w_0)K_0$ since the above weak inequality holds in equality when $t \rightarrow \infty$. This completes the argument. \square

A.2.4. Proof of Proposition 2.3

Auxiliary results. In this section, we prove the following auxiliary lemma, which is key for establishing the concavity of the value function.

Lemma 2. Let $p(\cdot)$ the unique, twice continuously differentiable solution to the HJB Eq. (11) on the interval $[0, \bar{w}]$ subject to the boundary conditions $p(0) = R$, $p'(\bar{w}) = -1$, and $p''(\bar{w}) = 0$. Further, assume the processes $\{s\}, \{\ell\}$ are of bounded variation. Then it follows for any $w_1 \in (0, \bar{w})$ with $p'(w_1) = 0$, that $p'(w_1) < 0$ and the policy functions $s(\cdot), \ell(\cdot)$ are continuous in a neighborhood of w_1 .

Proof. We start with an important observation. Because the processes $\{s\}, \{\ell\}$ are, by hypothesis, of bounded variation, they can be written as the difference of two almost surely increasing processes such that $a_t = a_t^1 - a_t^2$ for all $t \geq 0$, $a \in \{s, \ell\}$ and a_t^j increases almost surely. By Froda's theorem,²¹ each of the processes $\{\hat{a}\}$ has no essential discontinuity and at most countably many jump-discontinuities with probability one. Since $\{w\}$ follows a Brownian semimartingale, this implies that any point of discontinuity of $a(\cdot)$ can neither be an essential discontinuity nor can the set of discontinuity points of $a(\cdot)$ be dense in $[0, \bar{w}]$ for all $a \in \{s, \ell\}$.

We first prove that $p'(w_1) < 0$. Let us suppose to the contrary $p'(w_1) \geq 0$ hence $w_1 < \bar{w}$. Note that for any $\delta > 0$ exists $z \in (w_1 - \delta, w_1 + \delta)$ such that $s(\cdot), \ell(\cdot)$ are continuous in a neighborhood of z because discontinuity points do not form a dense set. Since $p'(\cdot), p''(\cdot)$ are continuous, for any $\varepsilon > 0$, we can choose $\delta > 0$ and $z \in (w_1 - \delta, w_1 + \delta)$ such that $\min\{p'(z), p''(z)\} > -\varepsilon$. The HJB Eq. (11) and the fact that $\ell(z) = \ell^{FB}$ is not necessarily optimal imply

$$\begin{aligned} (r - \mu\ell^{FB})p(z) &\geq \max_{s \in [0, s_{\max}]} \left\{ \alpha s + p'(z)(\gamma - \mu\ell^{FB})z - C(s, \ell^{FB}) + p''(z)\Sigma(z) \right\} \\ &\geq \max_{s \in [0, s_{\max}]} \left\{ \alpha s - \varepsilon(\gamma - \mu\ell^{FB})z - C(s, \ell^{FB}) + \Sigma(z) \right\}. \end{aligned}$$

Sending $\varepsilon, \delta \rightarrow 0$ such that $s = s(z) = s_{\max} \geq s^{FB}$ and in particular for $z = w_1$

$$\alpha s - C(s, \ell^{FB}) \geq \alpha s^{FB} - C(s^{FB}, \ell^{FB}) \geq (r - \mu\ell^{FB})p^{FB}.$$

Hence, there exists $z \in [0, \bar{w}]$ such that $p(z) \geq p^{FB}$, a contradiction.

Next, let us prove that $\ell(\cdot)$ must be continuous in a neighborhood of w_1 , and assume to the contrary that there is no neighborhood of w_1 on which $\ell(\cdot)$ is continuous. Since the set of discontinuities of $\ell(\cdot)$ must be discrete (not dense), it is immediate that

$$\ell_- \equiv \lim_{w \uparrow w_1} \ell(w) \neq \lim_{w \downarrow w_1} \ell(w) \equiv \ell_+,$$

i.e., $\ell(\cdot)$ has a jump discontinuity at w_1 itself. Without loss of generality, we will assume that $\ell_- < \ell_+$ and $w_1 < \bar{w}$.²²

Note that for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $z \in (w_1, w_1 + \delta)$, it holds that $|\ell(z) - \ell_+| < \varepsilon$. The optimality of $\ell(z)$ requires that $\frac{\partial p(z)}{\partial \ell} \Big|_{\ell=\ell(z)} \geq 0$ with equality if $\ell(z)$ is interior. Due to the continuity of $p''(\cdot)$, the limit $\varepsilon \rightarrow 0$ yields $\Gamma_\ell(w_1) \geq 0$ for

$$\begin{aligned} \Gamma_\ell(w) &= p(w) - p'(w)w - C_\ell(s, \ell_+) \text{ with } C_\ell(s, \ell_+) \\ &= \frac{\partial C(s, \ell)}{\partial \ell} \Big|_{\ell=\ell_+}. \end{aligned}$$

In addition, for all $\varepsilon > 0$, it must be that there exists $\delta > 0$ such that for all $x \in (w_1 - \delta, w_1)$, it holds that $|\ell(x) - \ell_-| < \varepsilon$. Hence, for $\varepsilon > 0$ sufficiently small,

²¹ Froda's theorem states that each real-valued, monotone function has at most countably many points of discontinuity. It is clear that such a function cannot have an essential discontinuity (i.e., a point of oscillation).

²² Since $p(\cdot)$ is extended linearly to the right of \bar{w} , discontinuity to the right of \bar{w} is not an issue.

$\ell(x) < \ell_{\max}$ and therefore $\frac{\partial p(x)}{\partial \ell} |_{\ell=\ell(x)} = 0$, which implies together with the continuity of $p''(\cdot)$ that $\widehat{\Gamma}_\ell(w_1) = 0$ for $\widehat{\Gamma}_\ell(w) = p(w) - p'(w)w - C_\ell(s, \ell_-)$.

Next, observe that

$$0 \leq \Gamma_\ell(w_1) - \widehat{\Gamma}_\ell(w_1) = -\lambda_\ell(\ell_+ - \ell_-).$$

Then, it follows that $\ell_- \geq \ell_+$, a contradiction.

Finally, assume that there is no neighborhood of w_1 on which $s(\cdot)$ is continuous. Since the set of discontinuity points of $s(\cdot)$ is discrete, this is equivalent to $s_- \equiv \lim_{w \uparrow w_1} s(w) \neq \lim_{w \downarrow w_1} s(w) \equiv s_+$. Without loss of generality, suppose $s_+ > s_-$. Then, for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $z \in (w_1, w_1 + \delta)$, it holds that $|s(z) - s_+| < \varepsilon$. Optimality requires $\frac{\partial p(z)}{\partial s} |_{s=s(z)} \geq 0$. Taking the limit $\varepsilon \rightarrow 0$, we obtain $\Gamma_s(w_1) \geq 0$ for $\Gamma_s(w) = \alpha s(w) - C_s(s_+, \ell)$. Similarly, $\widehat{\Gamma}_s(w_1) = 0$ for $\widehat{\Gamma}_s(w) = s(w) + p'(w)C_s(s_-, \ell)$. Hence,

$$0 \leq \Gamma_s(w_1) - \widehat{\Gamma}_s(w_1) = -\lambda_s(s_+ - s_-).$$

Then, it follows that $s_- \geq s_+$, a contradiction. \square

Concavity of the value function

Proof. Since $p''(\cdot)$ is continuous on $[0, \bar{w}]$ and $\{s\}, \{\ell\}$ are of bounded variation, it follows that the mappings $s(\cdot), \ell(\cdot)$ are continuous on $[0, \bar{w}]$ up to a discrete set with (Lebesgue) measure zero. On the set, where $s(\cdot), \ell(\cdot)$ are continuous, the envelope theorem implies now that $p'''(\cdot)$ exists and is given by

$$p'''(w) = \frac{(r-\gamma)p'(w) - p''(w)(w(\gamma - \mu\ell) - \sigma_K^2(\beta^\ell - w) - \rho\sigma_X\sigma_K\beta^s)}{\frac{1}{2}((\beta^s\sigma_X)^2 + \sigma_K^2(\beta^\ell - w)^2 + 2\rho\sigma_X\sigma_K\beta^s(\beta^\ell - w))}.$$

We have to show that $p''(w) < 0$ for all $0 \leq w < \bar{w}$.

By Lemma 2 we know that $s(\cdot), \ell(\cdot)$ are continuous in a neighborhood of \bar{w} . Then, we observe that $p''(\bar{w}) \propto \gamma - r > 0$ due to $\beta^s \geq \lambda_s s > 0$, and thus $p'''(\cdot) > 0$ in a neighborhood of \bar{w} . Hence, $p''(w) < 0$ on an interval $(\bar{w} - \varepsilon, \bar{w})$ with appropriate $\varepsilon > 0$.

Next, suppose there exists $w_0 \in [0, \bar{w}]$ with $p''(w_0) > 0$, and define $w_1 \equiv \sup\{w \in [0, \bar{w}] : p''(w) \geq 0\}$. By the previous step and continuity, it follows that $p''(w_1) = 0$ and $w_1 < \bar{w}$. We obtain now from Lemma 2 that $s(\cdot), \ell(\cdot)$ are continuous in a neighborhood of w_1 and $p'(w_1) < 0$. However, this implies $p'''(w_1) > 0$ and therefore $p'''(\cdot) > 0$ in a neighborhood of w_1 . Thus, there exists $w' > w_1$ with $p''(w') > 0$, a contradiction to the definition of w_1 . This completes the proof. \square

Last, let us state the following corollary, which proves useful in some instances:

Corollary 1. If $\gamma - r$ and σ_K^2 are sufficiently small, then $p'''(w) > 0$ for any $w \in [0, \bar{w}]$.

Proof. It is immediate from the above given the expression of $p'''(w)$. \square

A.3. Proofs of Propositions 3 and 4

Proof. The expressions for $s = s(w), \ell = \ell(w)$ follow directly from the maximization of $p(w)$ over $s \in [0, s_{\max}]$ and

$\ell \in [0, \ell_{\max}]$ for a given w as indicated by the HJB Eq. (11). Interior levels $s(w), \ell(w)$ must solve the respective first-order conditions of maximization (i.e., $\frac{\partial p(w)}{\partial s} |_{s=s(w)} = 0$ and $\frac{\partial p(w)}{\partial \ell} |_{\ell=\ell(w)} = 0$). After rearranging the first-order conditions, one arrives at the desired expressions.

Due to $p''(w) < 0$ for all $w < \bar{w}$ and $p''(\bar{w}) = 0$, it is immediate to see that $s(w) \leq s^{FB}$, with the inequality holding as equality if and only if $w = \bar{w}$. When $\gamma - r$ and σ_K are sufficiently small, then $p'''(w) > 0$ (see Corollary 1) for all w , and due to

$$\text{sign}\left(\frac{\partial s(w)}{\partial w}\right) = \text{sign}(p'''(w)),$$

short-run investment increases in w under these circumstances.

Evaluating the HJB equation at the boundary under the optimal controls yields

$$(r - \mu\ell)p(\bar{w}) + (\gamma - \mu\ell)\bar{w} = \alpha s - C(s, \ell).$$

Hence, owing to $\gamma > r$ and agency-induced termination, $\mathcal{P}(\tau < \infty) = 1$:

$$p(\bar{w}) + \bar{w} < \left(\frac{\alpha s - C(s, \ell)}{r - \mu\ell}\right) \leq p^{FB}.$$

Since $C_\ell(s^{FB}, \ell^{FB}) = \mu p^{FB}$ and $C_\ell(s(\bar{w}), \ell(\bar{w})) = \mu(p(\bar{w}) + \bar{w})$, it is clear that $\ell(\bar{w}) < \ell^{FB}$ and therefore by continuity that $\ell(w) < \ell^{FB}$ in a left neighborhood of \bar{w} . \square

A.4. Proof of Proposition 5

We prove Proposition 5 i) in two parts. Part I shows that either $\sigma_X = 0$ or $\sigma_K = 0$ implies $\ell(w) < \ell^{FB}$. Part II shows that there exist parameter values so that $\ell(w) > \ell^{FB}$ once $\sigma_X, \sigma_K > 0$. We start the proof with an auxiliary lemma.

A.4.1. Proof of Proposition 5—auxiliary results

Lemma 3. Under the optimal contract for an arbitrary parameter $\theta \notin \{r, \mu\}$, it holds

$$\begin{aligned} \frac{\partial p(w)}{\partial \theta} &= \mathbb{E} \left\{ \int_0^\tau e^{-\pi t + \mu \int_0^t \ell_s ds} \left[\frac{\partial \alpha}{\partial \theta} S_t - \frac{\partial C(S_t, \ell_t)}{\partial \theta} \right. \right. \\ &\quad + p'(w_t) w_t \frac{\partial(\gamma - \mu \ell_t)}{\partial \theta} - p(w_t) \frac{\partial(r - \mu \ell_t)}{\partial \theta} \\ &\quad + \frac{p''(w_t)}{2} \frac{\partial}{\partial \theta} [(\beta_t^s \sigma_X)^2 + \sigma_K^2(\beta_t^\ell - w_t)^2 \\ &\quad \left. \left. + 2\rho\sigma_X\sigma_K\beta_t^s(\beta_t^\ell - w_t)] \right] dt \Big| w_0 = w \right\}. \end{aligned}$$

Proof. Let $w \in [0, \bar{w}]$, $\theta \notin \{r, \mu\}$ and $s = s(w), \ell = \ell(w), \beta^s = \beta^s(w), \beta^\ell = \beta^\ell(w)$ be determined by the HJB Eq. (11).²³ Then, by the envelope theorem,

$$\frac{\partial p(w)}{\partial s} \frac{\partial s(w)}{\partial \theta} = \frac{\partial p(w)}{\partial \ell} \frac{\partial \ell(w)}{\partial \theta} = 0,$$

²³ For convenience, we suppress the dependence of $p(\cdot), \bar{w}$ on θ in the notation.

and therefore total differentiation of the HJB Eq. w.r.t. θ yields

$$\begin{aligned} & (r - \mu\ell) \frac{\partial p(w)}{\partial \theta} + \frac{\partial(r - \mu\ell)}{\partial \theta} p(w) \\ &= \frac{\partial \alpha}{\partial \theta} s - \frac{\partial C(s, \ell)}{\partial \theta} + p'(w)w \frac{\partial(\gamma - \mu\ell)}{\partial \theta} + w(\gamma - \mu\ell) \frac{\partial}{\partial w} \frac{\partial p(w)}{\partial \theta} \\ &+ \frac{p''(w)}{2} \frac{\partial}{\partial \theta} \left[(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) \right] \\ &+ \frac{\partial^2}{\partial w^2} \frac{\partial p(w)}{\partial \theta} \left[(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) \right]. \end{aligned}$$

Note that we used

$$\frac{\partial^k}{\partial w^k} \frac{\partial p(w)}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{\partial^k p(w)}{\partial w^k} \text{ for } k \in \{1, 2\};$$

i.e., we changed the order of (partial) differentiation, which is possible since p is sufficiently smooth. The above ordinary differential equation admits a unique solution subject to the boundary conditions

$$\frac{\partial p(w)}{\partial \theta} \Big|_{w=0} = 0 \text{ and } \frac{\partial p'(w)}{\partial \theta} \Big|_{w=\bar{w}} = \frac{\partial}{\partial w} \frac{\partial p(w)}{\partial \theta} \Big|_{w=\bar{w}} = 0,$$

and we can invoke Lemma 1 to arrive at the desired expression. \square

A.4.2. Proof of Proposition 5 (i)–part I

Let us assume $\sigma_X = 0$ and state the following lemma:

Lemma 4. Assume $\sigma_X = 0$. Hence, short-run investment $s(w)$ is contractible and constant over time. Then, it must be that $\beta^\ell > w$.

Proof. The proof is split in several parts. Part (a) shows that $\beta^\ell(\bar{w}) \neq \bar{w}$. Part (b) shows that $\beta^\ell(w) \neq w$, and part (c) concludes by showing $\beta^\ell(w) > w$ for all $w \in [0, \bar{w}]$.

(a) Let us first show that $\beta^\ell(\bar{w}) = \lambda_\ell \ell(\bar{w}) \neq \bar{w}$. Define $\ell := \ell(\bar{w})$ and suppose to the contrary $\lambda_\ell \ell = \bar{w}$. Then

$$p(\bar{w}) = \frac{1}{r - \mu\ell} \left(\alpha s - \frac{1}{2} (\lambda_s^2 \alpha s + \lambda_\ell \ell^2 \mu) - \bar{w}(\gamma - \mu\ell) \right).$$

Let $\varepsilon > 0$ and consider the Taylor expansion of $p(\bar{w} - \varepsilon)$ around $p(\bar{w})$, given by $p(\bar{w} - \varepsilon) = p(\bar{w}) + \varepsilon + o(\varepsilon^3)$. Further, define $\ell_\varepsilon := \ell(\bar{w} - \varepsilon)$ and note that in optimum $\beta^\ell(\bar{w} - \varepsilon) = \lambda_\ell \ell_\varepsilon + o(\varepsilon)$ by continuity. Hence

$$\begin{aligned} & (r - \mu\ell_\varepsilon) p(\bar{w} - \varepsilon) \\ &= \alpha s - \frac{\lambda_s \alpha s^2}{2} - \frac{1}{2} \lambda_\ell \ell_\varepsilon^2 \mu + p'(\bar{w} - \varepsilon) ((\gamma - \mu\ell_\varepsilon)(\bar{w} - \varepsilon)) \\ &+ \frac{\sigma_K^2 (\lambda_\ell \ell_\varepsilon + o(\varepsilon) - \bar{w} + \varepsilon)^2}{2} p''(\bar{w} - \varepsilon) \\ &= \alpha s - \frac{\lambda_s \alpha s^2}{2} - \frac{1}{2} \lambda_\ell \ell_\varepsilon^2 \mu + (-1 + o(\varepsilon^2)) ((\gamma - \mu\ell_\varepsilon)(\bar{w} - \varepsilon)) \\ &+ \frac{\sigma_K^2 (\lambda_\ell \ell_\varepsilon - \bar{w} + o(\varepsilon))^2}{2} p''(\bar{w} - \varepsilon), \end{aligned}$$

where we used that $p'(\bar{w} - \varepsilon) = p'(\bar{w}) - \varepsilon p''(\bar{w}) + o(\varepsilon^2)$.

Combining the above and using the Taylor expansion for $p(\bar{w} - \varepsilon)$ around $p(\bar{w})$ yields

$$\begin{aligned} & p(\bar{w} - \varepsilon) \mu (\ell_\varepsilon - \ell) \\ &= \varepsilon(r - \mu\ell) + (\gamma - \mu\ell_\varepsilon)(\bar{w} - \varepsilon) - \bar{w}(\gamma - \mu\ell) \end{aligned}$$

$$\begin{aligned} & + \frac{1}{2} \mu \lambda_\ell (\ell_\varepsilon^2 - \ell^2) - \frac{\sigma_K^2 (\lambda_\ell \ell_\varepsilon - \bar{w} + o(\varepsilon))^2}{2} p''(\bar{w} - \varepsilon) \\ &+ o(\varepsilon^2) + o(\varepsilon^3). \end{aligned}$$

Next, note that $\ell = \ell_\varepsilon + \varepsilon \ell'(\bar{w} - \varepsilon) + o(\varepsilon^2)$, in case $\ell(\cdot)$ is differentiable, which is guaranteed for $\varepsilon > 0$ sufficiently small. This yields

$$\begin{aligned} & \mu p(\bar{w} - \varepsilon) (-\varepsilon \ell'(\bar{w} - \varepsilon)) \\ &= \varepsilon(r - \gamma) - \bar{w} \mu \varepsilon \ell'(\bar{w} - \varepsilon) + o(\varepsilon^2) \\ &\iff o(\varepsilon) - \mu(p(\bar{w} - \varepsilon) + \bar{w}) \ell'(\bar{w} - \varepsilon) = r - \gamma. \end{aligned}$$

If $\ell(\bar{w}) = \ell_{\max}$, then it must be either that $\ell'(\bar{w} - \varepsilon) = o(\varepsilon)$ for ε sufficiently small, which leads to $\gamma - r = o(\varepsilon)$ and thereby a contradiction, or $\lim_{w \uparrow \bar{w}} \ell'(w) > 0$.

If $\ell(\bar{w}) < \ell_{\max}$ or $\lim_{w \uparrow \bar{w}} \ell'(w) > 0$, then $\ell(\bar{w} - \varepsilon)$ solves the following first-order condition of maximization, $\frac{\partial p(\bar{w} - \varepsilon)}{\partial \ell} = 0$. Moreover, $\ell(\bar{w}) < \ell_{\max}$ also solves the FOC at $w = \bar{w}$:

$$\begin{aligned} & \mu p(\bar{w}) + \mu \bar{w} - \lambda_\ell \mu \ell(\bar{w}) = 0 \\ &\iff p(\bar{w}) + \bar{w} - \lambda_\ell \ell(\bar{w}) = 0. \end{aligned}$$

Invoking the implicit function theorem, we can differentiate the above identity with respect to $w = \bar{w}$ so as to obtain $\ell'(\bar{w}) = 0$ as well as $\ell''(\bar{w}) = 0$. Then, by Taylor's theorem, which is applicable owing to $p \in C^2$, we get $\ell'(\bar{w} - \varepsilon) = o(\varepsilon^2)$, which yields the desired contradiction. This concludes the proof.

(b) Let us assume that there exists now $w < \bar{w}$ with $\beta^\ell = \lambda_\ell \ell(w) = w$ optimal, in which case the HJB equation under the optimal control reads:

$$\begin{aligned} (r - \mu\ell(w)) p(w) &= \alpha s(w) - \frac{\lambda_s \alpha s(w)^2}{2} - \frac{\lambda_\ell \mu \ell(w)^2}{2} \\ &+ p'(w) w (\gamma - \mu\ell(w)). \end{aligned}$$

Due to $p'(w) \geq -1$ (i.e., since scaled payouts at rate $w(\gamma - \mu\ell)$ and this way keeping $w_t = w$ constant for all future times t is always an option but not necessarily optimal), it follows that

$$\begin{aligned} p(w) &\geq \frac{1}{r - \mu\ell(w)} \left(\alpha s(w) - \frac{\lambda_s \alpha s(w)^2}{2} - \frac{\lambda_\ell \ell(w)^2 \mu}{2} \right. \\ &\left. - w(\gamma - \mu\ell(w)) \right). \end{aligned}$$

Likewise, due to the fact that $\ell(w) \lambda_\ell = w$ is optimal, it also must hold that

$$p(w) \geq \max_{s, \ell} \frac{1}{r - \mu\ell} \left(\alpha s - \frac{\lambda_s \alpha s^2}{2} - \frac{\lambda_\ell \ell^2 \mu}{2} - w(\gamma - \mu\ell) \right).$$

Then

$$\begin{aligned} & p(w) < p(\bar{w}) - (w - \bar{w}) \\ &= \max_{s, \ell} \frac{1}{r - \mu\ell} \left(\alpha s - \frac{\lambda_s \alpha s^2}{2} - \frac{\lambda_\ell \ell^2 \mu}{2} - \bar{w}(\gamma - r) \right. \\ &\quad \left. - w(r - \mu\ell) \right) \\ &< \max_{s, \ell} \frac{1}{r - \mu\ell} \left(\alpha s - \frac{\lambda_s \alpha s^2}{2} - \frac{\lambda_\ell \ell^2 \mu}{2} - w(\gamma - r) \right. \\ &\quad \left. - w(r - \mu\ell) \right) \end{aligned}$$

$$= \max_{s, \ell} \frac{1}{r - \mu \ell} \left(\alpha s - \frac{\lambda_s \alpha s^2}{2} - \frac{\lambda_\ell \ell^2 \mu}{2} - w(\gamma - \mu \ell) \right),$$

where the first inequality is due to strict concavity and the second one due to $w < \bar{w}$. This yields the desired contradiction.

(c) Eventually, let us assume that to the contrary $\beta^\ell = \ell(w)\lambda_\ell < w$ for at least one point w and define for this sake the function $\chi(w) = \beta^\ell(w) - w$. In a neighbourhood of $w = 0$ (i.e., on $(0, \varepsilon)$ for appropriate $\varepsilon > 0$), it is evident that $\chi(w) > 0$ because $0 < R = p(0) \Rightarrow \ell(0) > 0$.

If there is w' such that $\chi(w') < 0$, then there exists also w such that $\chi(w) = 0$ by continuity. If $\chi(w) = 0$ for some $w > 0$, then it must either be that $\chi(\bar{w}) = 0$, which contradicts part a), or $\chi(w) = 0$ for $0 < w < \bar{w}$, which contradicts part b). Hence, $\beta^\ell = \ell(w)\lambda_\ell > w$ for all $w \in [0, \bar{w}]$, which eventually proves the lemma. □

A.4.3. Proof of Proposition 5 (i)–part II

Proof. Here, we prove that $\sigma_K = 0$ and $\sigma_X > 0$ imply that $\ell(w) < \ell^{FB}$ provided investment is not at the corner.

For interior levels, $\ell = \ell(w)$ solves the first-order condition of maximization $\frac{\partial p(w)}{\partial \ell} = 0$ so that

$$\mu(p(w) - p'(w)w) - \lambda_\ell \mu \ell = 0.$$

Because of $p(w) - wp'(w) < p^{FB}$ and ℓ^{FB} solves $\mu p^{FB} - \lambda_\ell \mu \ell = 0$, it is immediate to see that $\ell(w) < \ell^{FB}$ for all $w \in [0, \bar{w}]$. For corner levels, a similar argument applies, which readily yields $\ell(w) \leq \ell^{FB}$ with the inequality being strict if $\ell_{\max} > \ell^{FB}$. □

A.4.4. Proof of Proposition 5 (ii)

Proof. Let θ denote an arbitrary set of model parameters and denote the family of solutions to the principal's problem by $\{p_\theta, \bar{w}_\theta\}_\theta$. By Berge's maximum theorem, \bar{w}_θ is continuous w.r.t. (the value of) θ , in the standard Euclidean metric space on \mathbb{R} and p_θ is continuous in θ on $\mathbf{A}^{\mathbf{B}}$ with respect to the topology, induced by the norm $\|\cdot, \cdot\|_\infty$ where

$$\|f\|_\infty = \sup_{x \in \mathbf{A}} |f(x)|.$$

Here, \mathbf{A}, \mathbf{B} are some compact subsets of \mathbb{R} that satisfy all necessary regularity conditions and possibly depend on θ . We choose \mathbf{A} sufficiently large so that $\bar{w}_\theta \in \mathbf{A}$ and $0 \in \mathbf{A}$ for all considered θ . We may choose \mathbf{B} so that $p_\theta(w) \in \mathbf{B}$ for all $w \in [0, \bar{w}_\theta]$ for all considered θ . For brevity, we omit a formal introduction of the sets \mathbf{A}, \mathbf{B} and the associated notation in the following.

Without loss of generality, we assume throughout that the constraint $\ell \leq \ell_{\max}$ is never tight. The proof goes through as long as the first-best level is interior (i.e., $\ell_{\max} > \ell^{FB}$). Formally dealing within the proof with corner levels would merely complicate the notation.

Let us start by considering the limit case $\mu \rightarrow 0$ holding the remaining parameters fixed. That is, we study the family $\{p_\mu, \bar{w}_\mu\}_{\mu \geq 0}$ and take the limit $\mu \rightarrow 0$. The model in the limit case $\mu \rightarrow 0$ is well behaved and features a value function p_0 with reflecting boundary $\bar{w}_0 > 0$. Due to

continuity in μ , it follows that $p_\mu \rightarrow p_0$ and $\bar{w}_\mu \rightarrow \bar{w}_0$ as $\mu \rightarrow 0$. As a consequence,

$$\ell(w) \rightarrow \frac{-wp_0''(w)\lambda_\ell \sigma_K^2}{-p_0''(w)(\lambda_\ell \sigma_K)^2} = \frac{w}{\lambda_\ell} \wedge \ell_{\max} > 0 \text{ for } \bar{w}_0 > w > 0,$$

where we omit for simplicity indexing for the optimal controls (e.g., for $\ell = \ell_\mu$). It can be verified for $w < \bar{w}_\mu$ that

$$\nabla(dw) = (\beta^s \sigma_X)^2 dt + (\beta^\ell - w)^2 \sigma_K^2 dt = o(\sigma_X^2) dt,$$

when $\mu \rightarrow 0$ because $\beta^\ell \rightarrow \lambda_\ell \ell(w) = w$. If it were $\ell(w) = \ell_{\max}$, then it is easy to verify that $\beta^\ell = w$ becomes optimal.

As a consequence, the joint limit $\sigma_X, \mu \rightarrow 0$ would lead to a solution where $\{w\}$ has no volatility and accordingly $\bar{w}_0 = 0$. That is, $\lim_{\sigma_X \rightarrow 0} \bar{w}_0 = 0$. Hence, $\sigma_X = 0 \Rightarrow \bar{w}_0 = 0$. Since the limit case $\mu \rightarrow 0$ corresponds (effectively) to the model of DeMarzo and Sannikov (2006), we know that $\sigma_X > 0 \Leftrightarrow \bar{w}_0 > 0$.

To take the limit $\lim_{\mu \rightarrow 0} \ell^{FB}$, we have to use L'Hospital's rule, which yields

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \frac{1}{\mu} \left[r - \sqrt{r^2 - \frac{\mu \alpha}{\lambda_s \lambda_\ell}} \right] \\ &= \lim_{\mu \rightarrow 0} \frac{1}{2\sqrt{r^2 - \frac{\mu \alpha}{\lambda_s \lambda_\ell}}} \frac{\alpha}{\lambda_s \lambda_\ell} = \frac{\alpha}{2\lambda_s \lambda_\ell r}. \end{aligned}$$

To avoid cluttering with subscripts, we omit indexing model quantities by μ when it does not cause confusion.

We prove now the claim regarding γ . Let us fix all parameters except γ and consider the (continuous) family of solutions $\{p_\gamma, \bar{w}_\gamma\}_{\gamma > r \geq 0}$. We evaluate the HJB equation at the boundary:

$$\begin{aligned} & (r - \mu \ell(\bar{w}_\gamma)) p_\gamma(\bar{w}_\gamma) \\ &= \frac{\alpha}{2\lambda_s} - (\gamma - \mu \ell(\bar{w}_\gamma)) \bar{w}_\gamma - \frac{\lambda_\ell \mu \ell(\bar{w}_\gamma)^2}{2}, \end{aligned}$$

and totally differentiate with respect to γ . Using $dp(\bar{w}_\gamma)/d\gamma = p'(\bar{w}_\gamma) \partial \bar{w}_\gamma / \partial \gamma + \partial p(\bar{w}_\gamma) / \partial \gamma$ and the boundary condition $p'(\bar{w}_\gamma) = -1$, we obtain

$$\frac{\partial \bar{w}_\gamma}{\partial \gamma} = -\frac{1}{\gamma - r} \left(\bar{w}_\gamma + (r - \mu \ell(\bar{w}_\gamma)) \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} \right),$$

where, by Lemma 3,

$$\begin{aligned} \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} &= \mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} p'_\gamma(w_t) w_t dt \mid w_0 = \bar{w}_\gamma \right) \\ &\geq -\mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} w_t dt \mid w_0 = \bar{w}_\gamma \right), \end{aligned}$$

where the inequality uses $p'_\gamma \geq -1$.

Let us assume that $\frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} \geq 0$. Then

$$\begin{aligned} \mathcal{A}(\gamma) &:= \bar{w}_\gamma + (r - \mu \ell(\bar{w}_\gamma)) \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} \\ &\geq \frac{\bar{w}_\gamma \mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} dt \mid w_0 = \bar{w}_\gamma \right)}{\mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} dt \mid w_0 = \bar{w}_\gamma \right)} \\ &\quad + (r - \mu \ell_{\max}) \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} \end{aligned}$$

$$\begin{aligned} &\geq (r - \mu \ell_{\max}) \bar{w}_\gamma \mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} dt \mid w_0 = \bar{w}_\gamma \right) \\ &\quad + (r - \mu \ell_{\max}) \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} \\ &\geq (r - \mu \ell_{\max}) \mathbb{E} \left(\int_0^\tau e^{-rt} (\bar{w}_\gamma - w_t) dt \mid w_0 = \bar{w}_\gamma \right) > 0, \end{aligned}$$

where we use that $p'_\gamma \geq -1$ as well as $\mathcal{P}(\tau < \infty) = 1$. In case $\frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} < 0$, we obtain similarly

$$\begin{aligned} A(\gamma) &= \bar{w}_\gamma + (r - \mu \ell(\bar{w}_\gamma)) \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} \\ &\geq \bar{w}_\gamma - r \mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} w_t dt \mid w_0 = \bar{w}_\gamma \right) \\ &\geq r \mathbb{E} \left(\int_0^\tau e^{-rt} (\bar{w}_\gamma - w_t) dt \mid w_0 = \bar{w}_\gamma \right) \\ &\geq (r - \mu \ell_{\max}) \mathbb{E} \left(\int_0^\tau e^{-rt} (\bar{w}_\gamma - w_t) dt \mid w_0 = \bar{w}_\gamma \right) > 0 \end{aligned}$$

for all $\gamma > r$.

It follows that $\bar{w}_\gamma + (r - \mu \ell(\bar{w}_\gamma)) \frac{\partial p_\gamma(\bar{w}_\gamma)}{\partial \gamma} > 0$, so that $\frac{\partial \bar{w}_\gamma}{\partial \gamma} < 0$. Since the process $\{w\}$ has—because of $\sigma_X > 0$ and $\beta^S(\bar{w}) = \lambda_s s^{FB}$ —strictly positive volatility at the boundary \bar{w}_γ , the payout boundary \bar{w}_γ cannot constitute an absorbing (or attracting) state. This holds true for any $\gamma > r$. Due to that and the fact that the stochastic process $\{w\}$ possesses strictly positive volatility almost everywhere on $(0, \bar{w}_\gamma)$, it cannot be that the above expectation $\mathcal{E}(\gamma) := \mathbb{E}(\int_0^\tau e^{-rt} (\bar{w}_\gamma - w_t) dt \mid w_0 = \bar{w}_\gamma)$ tends to zero, as $\gamma \rightarrow r$ so that $\mathcal{E}(\gamma) \notin o(\gamma - r)$, and therefore $\lim_{\gamma \downarrow r} \mathcal{E}(\gamma) > 0$. By continuity, the aforementioned limit exists but possibly takes value ∞ . From there it follows that

$$\begin{aligned} \frac{\partial \bar{w}_\gamma}{\partial \gamma} &= \frac{A(\gamma)}{-(\gamma - r)} \\ &\leq \frac{(r - \mu \ell_{\max}) \mathbb{E}(\int_0^\tau e^{-rt} (\bar{w}_\gamma - w_t) dt \mid w_0 = \bar{w}_\gamma)}{-(\gamma - r)}, \end{aligned}$$

and accordingly

$$\begin{aligned} \lim_{\gamma \downarrow r} \frac{\partial \bar{w}_\gamma}{\partial \gamma} &= \lim_{\gamma \downarrow r} \frac{A(\gamma)}{\gamma - r} \\ &\leq \lim_{\gamma \downarrow r} \frac{(r - \mu \ell_{\max}) \mathbb{E}(\int_0^\tau e^{-rt} (\bar{w}_\gamma - w_t) dt \mid w_0 = \bar{w}_\gamma)}{-(\gamma - r)} \\ &= -\infty. \end{aligned}$$

Thus, the function $\gamma \mapsto \bar{w}_\gamma$, which is defined on $[r, \infty)$, has a singularity (pole) at $\gamma = r$, which implies $\lim_{\gamma \downarrow r} \bar{w}_\gamma = \infty$.

By continuity, for any $\varepsilon > 0$ and for any $\mu > 0$, there exists $\gamma > r$ sufficiently low so that the payout threshold $\bar{w}_{\mu, \gamma}$ (dependent on μ, γ) satisfies $\frac{\bar{w}_{\mu, \gamma}}{\lambda_\ell} > \frac{\alpha}{2\lambda_s \lambda_\ell r} + \varepsilon$. Under these circumstances, there is a value $0 < \tilde{w}_\gamma^\mu < \bar{w}_{\mu, \gamma}$ with

$$\frac{\tilde{w}_\gamma^\mu}{\lambda_\ell} > \frac{\alpha}{2\lambda_s \lambda_\ell r} + \varepsilon.$$

Taking the limit of the above constructed sequence yields (by construction)

$$\lim_{\mu \rightarrow 0} \ell(\tilde{w}_\gamma^\mu) > \lim_{\mu \rightarrow 0} \ell^{FB}.$$

By continuity, there exist now $\mu > 0$ and $\gamma > r$ sufficiently low and $w \in (0, \bar{w}_{\mu, \gamma})$ so that $\ell(w) > \ell^{FB}$.

Let $w^H \equiv \sup\{w : \ell(w) > \ell^{FB}\}$ and $w^L \equiv \sup\{w : \ell(w) > \ell^{FB}\}$. Since $\ell(\bar{w}_{\mu, \gamma}) < \ell^{FB}$ for any $\mu > 0, \gamma > r$, it must be that $w^H < \bar{w}_{\mu, \gamma}$ with $\lim_{\mu \rightarrow 0, \gamma \rightarrow r} w^H = \bar{w}_{\mu, \gamma}$. In addition, $\lim_{\mu \rightarrow 0, \gamma \rightarrow r} w^L = \frac{\alpha}{2\lambda_s r}$. It follows then that in the limit $\mu \rightarrow 0, \gamma \rightarrow r$ it must be that the set $\{w : \ell(w) = w/\lambda_\ell + o(\mu) \wedge \ell_{\max} > \ell^{FB}\}$ is convex. By continuity, there exist $\mu > 0$ and $\gamma > r$, ensuring the set $\{w : \ell(w) > \ell^{FB}\}$ is convex, thereby concluding the proof. \square

A.4.5. Proof of Proposition 5 (iii)

Proof. We consider parameters are such that $\sup\{\ell(w) : w \in [0, \bar{w}]\} = \ell^{FB}$, and let $\mu > 0$ so that $\ell(w) > 0$ for any $w \in (0, \bar{w}]$. Let us evaluate the HJB equation at the boundary:

$$(r - \mu \ell(\bar{w})) p(\bar{w}) = \frac{\alpha}{2\lambda_s} - (\gamma - \mu \ell(\bar{w})) \bar{w} - \frac{\lambda_\ell \mu \ell(\bar{w})^2}{2}.$$

Differentiating this identity w.r.t. σ_i for $i \in \{X, K\}$ leads to

$$\frac{\partial \bar{w}}{\partial \sigma_i} = \frac{r}{r - \gamma} \frac{\partial p(\bar{w})}{\partial \sigma_i}.$$

Lemma 3 then implies:

$$\begin{aligned} &\frac{\partial p(\bar{w})}{\partial \sigma_K} \\ &= \mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} p''(w_t) (\beta_t^\ell - w_t)^2 \sigma_K dt \mid w_0 = \bar{w} \right) < 0, \\ &\frac{\partial p(\bar{w})}{\partial \sigma_X} \\ &= \mathbb{E} \left(\int_0^\tau e^{-rt + \mu \int_0^t \ell_u du} (p''(w_t) (\beta_t^S)^2 \sigma_X) dt \mid w_0 = \bar{w} \right) < 0 \end{aligned}$$

so that \bar{w} increases in σ_i for $i \in \{X, K\}$. The claim follows due to continuity in parameter values $\{\sigma_X, \sigma_K\}$. \square

A.5. Proof of Proposition 6 and 7

We prove the two propositions separately. In both cases claim i) is trivial since $\sigma_K = 0$ precludes risk externalities between short- and long-run incentives.

A.5.1. Proof of Proposition 6 (ii)

Proof. The proof of Proposition 6 (ii) is split in two parts. The first part of the proof shows that there is short-termism, $s(w) > s^{FB}$, for σ_X sufficiently small; the second one points out under which circumstances $\{w : s(w) > s^{FB}\}$ is convex.

Let us assume that correlation ρ is negative. Let us fix all parameters and consider the family of solution $\{p_{\sigma_X}, \bar{w}_{\sigma_X}\}$, which is—by Berge’s maximum theorem—continuous in σ_X w.r.t. an appropriate topology, already discussed before. In the limit case $\sigma_X \rightarrow 0$, we have $s(w) \rightarrow s^{FB}$ for all $w \in [0, \bar{w}_{\sigma_X}]$. In addition, for any $\sigma_X \geq 0$,

including the limit case $\sigma_X \rightarrow 0$, we have $p''_{\sigma_X}(0) < 0$, as $\bar{w}_{\sigma_X} > 0$ due to $\sigma_K > 0$. For notational convenience, we omit indexing model quantities by σ_X when no confusion is likely to arise.

We can write

$$\begin{aligned} s(w) &= \frac{\alpha + p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell\ell(w) - w)}{\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2} \\ &= \frac{\alpha + p''(w)\rho\sigma_X\sigma_K\lambda_s\lambda_\ell\ell(w)}{\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2} + o(w). \end{aligned}$$

From there it follows immediately that

$$\frac{\partial s(w)}{\partial p''(w)} = o(\sigma_X).$$

Thus

$$\begin{aligned} \frac{ds(w)}{d\sigma_X} &= \frac{\partial s(w)}{\partial \sigma_X} + \frac{\partial s(w)}{\partial p''(w)} \frac{\partial p''(w)}{\partial \sigma_X} \\ &\propto [\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2] p''(w)\rho\sigma_K\lambda_s(\lambda_\ell\ell(w) - w) \\ &\quad + 2[\alpha + p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell\ell(w) - w)] \\ &\quad \times p''(w)\lambda_s^2\sigma_X + o(\sigma_X) \\ &= p''(w)\rho\sigma_K\lambda_s^2(\lambda_\ell\ell(w) - w)\alpha + o(\sigma_X) \\ &= p''(w)\rho\sigma_K\lambda_s^2\lambda_\ell\ell(w)\alpha + o(\sigma_X) + o(w), \end{aligned}$$

where \propto means “has the same sign as.”

Because of $R > 0$, we have $\ell(0) > 0$. This implies $\ell(w) > 0$ close to zero and $\ell(w) \notin o(w)$. Hence, it holds that $\lambda_\ell\ell(w) > w$ in a neighborhood of zero, implying short-run investment $s(w)$ increases in σ_X , provided $\sigma_X > 0$ and w are sufficiently close to zero. This follows from $\lim_{\sigma_X \rightarrow 0} p'' \neq 0$ and $p''(0) < 0$ because $\sigma_K > 0$ guarantees a nontrivial boundary $\bar{w} > 0$, even in the limit $\sigma_X \rightarrow 0$. Because of $s(w) = s^{FB}$, if $\sigma_X = 0$, there exists $\sigma_X > 0$ and $w \in [0, \bar{w}]$ so that $s(w) > s^{FB}$, which concludes the first part of the proof.²⁴

The second part of the proof establishes the convexity of the set $\{w: s(w) > s^{FB}\}$ under certain parameters conditions. Let us calculate

$$\begin{aligned} \frac{\partial s(w)}{\partial w} &\equiv s'(w) \propto p'''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell\ell(w) - w) \\ &\quad + p''(w)\rho\sigma_X\sigma_K\lambda_s \frac{\partial(\lambda_\ell\ell(w) - w)}{\partial w} + o(\sigma_X^2). \end{aligned}$$

If $\gamma - r$ (and possibly σ_K^2) is sufficiently small, then $p'''(w) \geq 0$ (see [Corollary 1](#)) so that the first term is negative for $w < \lambda_\ell\ell(w)$ (i.e., for w close to zero). If λ_ℓ is sufficiently small, then

$$\frac{\partial(\lambda_\ell\ell(w) - w)}{\partial w} = \lambda_\ell\ell'(w) - 1 < 0$$

so that the second term is also negative. The remainder is negligible for σ_X sufficiently small. Under these conditions, $s'(w) < 0$ for $w \leq \lambda_\ell\ell(w)$.

Let us conclude the proof by demonstrating $\{w: s(w) > s^{FB}\}$ must be a convex set, containing zero when in addition to σ_X also λ_ℓ and $\gamma - r$ are sufficiently small so as to ensure $\partial s(w)/\partial w < 0$ for $w < \lambda_\ell\ell(w)$. Without loss of

generality, assume that $\{w: s(w) > s^{FB}\}$ is nonempty. If the set is not convex, it must be that there exists $w' \in [0, \bar{w}]$ with $s(w') = s^{FB}$ and $s'(w') > 0$ such that $w' < \bar{w}$. Next, let us take a look at

$$s(w) = \frac{\alpha + p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell\ell(w) - w)}{\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2},$$

and notice that for $s(w') \geq s^{FB}$ being optimal it is necessary that $\lambda_\ell\ell(w') > w'$, as $\rho < 0$ and $p''(w') < 0$. This implies $s'(w') < 0$ when λ_ℓ and $\gamma - r$ are sufficiently small, a contradiction.

Next, assume the set does not contain zero (i.e., $s(0) \leq s^{FB}$). It follows that $s'(\hat{w}) > 0$ for $\hat{w} = \inf\{w \geq 0 : s(w) > s^{FB}\}$. By continuity $s(\hat{w}) = s^{FB}$. However, due to $s(\bar{w}) = s^{FB}$, it must be that $\hat{w} < \bar{w}$. For $s(\hat{w}) = s^{FB}$ being optimal, it must be that $\lambda_\ell\ell(\hat{w}) > \hat{w}$. This implies $s'(\hat{w}) < 0$ when λ_ℓ and $\gamma - r$ are sufficiently small, a contradiction. This concludes the proof. \square

A.5.2. Proof of Proposition 7 (ii)

Proof. Fix $\sigma_X > 0$ and consider $\gamma - r$ sufficiently small, such that there exists $w < \bar{w}$ with $w > \lambda_\ell\ell(w)$. This is possible as $\bar{w} \rightarrow \infty$ for $\gamma \rightarrow r$ and because there exists a left neighborhood of \bar{w} , where $\ell(w) < \ell^{FB} < \infty$ for any $\gamma > r$.

Note that this holds for any $\sigma_X > 0$. Therefore, we can choose σ_X sufficiently small and $\gamma - r$ sufficiently small so that there exists $w < \bar{w}$ with $p''(w) < 0$ and $s(w) > s^{FB}$ because of

$$s(w) = \frac{\alpha + p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell\ell(w) - w)}{\lambda_s\alpha + o(\sigma_X)^2},$$

that is, because the incentive cost of short-run investment is of order σ_X^2 , while the incentive externality is of order σ_X . Taking the limit $\sigma_X \rightarrow 0$ is innocuous only because $\sigma_K > 0$ guarantees a nontrivial solution in this limit. To be more rigorous, one could mimic and adapt the argument of the proof of [Proposition 6\(ii\)](#).

Since in the limit $\mu \rightarrow 0$ for arbitrary $\sigma_X > 0$, long-term investment satisfies $\ell(w) \rightarrow \frac{w}{\lambda_\ell}$, it follows that $s(w) \rightarrow \hat{s}(w) < s^{FB}$ for $w < \bar{w}$, as $\mu \rightarrow 0$. From there, it follows readily that there exist $\mu > 0$, $\gamma - r$, and σ_X sufficiently small such that $\{w: s(w) > s^{FB}\}$ is nonempty and convex with its infimum exceeding zero and its supremum equal to \bar{w} . \square

A.6. Proof of Proposition 8

Proof. Claim i) is straightforward, directly follows from the HJB equation, and is already explained in the main text.

Claim ii) is implied by the proof of [Proposition 5\(i\)](#), where we show that $\lambda_\ell\ell(w) > w$ for all w when $\sigma_X = 0$. The proof can be easily adjusted for linear cost (compare, e.g., [He, 2009](#)).

Claim (iii) relies on the premise that \bar{w} increases in $1/(\gamma - r)$ with $\lim_{\gamma \downarrow r} \bar{w} = \infty$ and can be proven to mimic the argument of the proof of [Proposition 5\(ii\)](#). Moreover, the limit $\lambda_\ell \rightarrow 0$ leads to a well-behaved solution with a strictly positive payout threshold. Hence, it follows by continuity of the solution $\{p_{\lambda_\ell}, \bar{w}_{\lambda_\ell}\}_{\lambda_\ell \geq 0}$ that there exists w with $\beta^\ell(w) = w > \lambda_\ell$ for $0 < w < \bar{w}_{\lambda_\ell}$ when λ_ℓ is sufficiently small. \square

²⁴ If we did not have $R > 0$, the proof is still valid as long as there exists a point $w < \bar{w}$ satisfying $\ell(w) > w/\lambda_\ell$. The existence of such a point can be ensured by appropriate λ_ℓ .

References

- Abel, A., Eberly, J., 2011. How q and cash flow affect investment without frictions: an analytic explanation. *Rev. Econ. Stud.* 78, 1179–1200.
- Aghion, P., Stein, J., 2008. Growth versus margins: destabilizing consequences of giving the market what it wants. *J. Financ.* 63, 1025–1057.
- Ai, H., Li, R., 2015. Investment and CEO compensation under limited commitment. *J. Financ. Econ.* 116 (3), 452–472.
- Albuquerque, R., Hopenhayn, H., 2004. Optimal lending contracts and firm dynamics. *Rev. Econ. Stud.* 71, 285–315.
- Asker, J., Farre-Mensa, J., Ljunqvist, A., 2015. Corporate investment and stock market listing: a puzzle. *Rev. Financ. Stud.* 28, 342–390.
- Barton, D., Manyika, J., Williamson, S., 2017. The data: where long-termism pays off. *Harv. Bus. Rev.* 95 (3), 67–68.
- Bernstein, S., 2015. Does going public affect innovation? *J. Financ.* 70 (1), 1365–1403.
- Bettis, J.C., Bizjak, J., Coles, J.L., Kalpathy, S., 2018. Performance-vesting provisions in executive compensation. *J. Account. Econ.* 66 (1), 194–221.
- Biais, B., Mariotti, T., Plantin, G., Rochet, J.C., 2007. Dynamic security design: convergence to continuous time and asset pricing implications. *Rev. Econ. Stud.* 74 (2), 345–390.
- Biais, B., Mariotti, T., Rochet, J.-C., Villeneuve, S., 2010. Large risks, limited liability, and dynamic moral hazard. *Econometrica* 78 (1), 73–118.
- Bolton, P., Chen, H., Wang, N., 2011. A unified theory of Tobin's q , corporate investment, financing, and risk management. *J. Financ.* 66 (5), 1545–1578.
- Bolton, P., Scheinkman, J., Xiong, W., 2006. Executive compensation and short-termist behavior in speculative markets. *Rev. Econ. Stud.* 73 (1), 577–610.
- Bolton, P., Wang, N., Yang, J., 2019. Optimal contracting, corporate finance, and valuation with inalienable human capital. *J. Financ.* 74 (3), 1363–1429.
- Brunnermeier, M., Sannikov, Y., 2014. A macroeconomic model with a financial sector. *Am. Econ. Rev.* 104, 379–421.
- Carlson, M., Fisher, A., Giammarino, R., 2006. Corporate investment and asset price dynamics: implications for SEO event-studies and long-run performance. *J. Financ.* 61, 1009–1034.
- Chang, X., Dasgupta, S., Wong, G., Yao, J., 2014. Cash-flow sensitivities and the allocation of internal cash flow. *Rev. Financ. Stud.* 27 (12), 3628–3657.
- Cox, J., Ingersoll, J., Ross, S., 1985. An intertemporal general equilibrium model of asset prices. *Econometrica* 53, 363–384.
- Décamps, J.-P., Gryglewicz, S., Morellec, E., Villeneuve, S., 2017. Corporate policies with permanent and transitory shocks. *Rev. Financ. Stud.* 30 (1), 162–210.
- DeMarzo, P., Fishman, M., He, Z., Wang, N., 2012. Dynamic agency and the q theory of investment. *J. Financ.* 67 (6), 2295–2340.
- DeMarzo, P., Sannikov, Y., 2006. Optimal security design and dynamic capital structure in a continuous-time agency model. *J. Financ.* 61 (6), 2681–2724.
- DeMarzo, P.M., Livdan, D., Tchisti, A., 2014. Risking other people's money: gambling, limited liability, and optimal incentives. Unpublished working paper. Stanford GSB.
- Edmans, A., Gabaix, X., Jenter, D., 2017. Executive compensation: a survey of theory and evidence. In: Hermalin, B.E., Weisbach, M.S. (Eds.), *The Handbook of the Economics of Corporate Governance*. Elsevier, Amsterdam, pp. 383–539.
- Francis, B., Iftekhar, H., Kose, J., Zenu, S., 2013. Asymmetric benchmarking of pay in firms. *J. Corp. Financ.* 23, 39–53.
- Garvey, G., Milbourn, T., 2006. Asymmetric benchmarking in compensation: executives are rewarded for good luck but not penalized for bad. *Journal of Financial Economics* 82 (1), 197–225.
- Gertler, M., Kiyotaki, N., 2010. Financial intermediation and credit policy in business cycle analysis. In: Friedman, B., Woodford, M. (Eds.), *Handbook of Monetary Economics*. Elsevier, Amsterdam, pp. 547–599.
- Giannetti, M., Yu, X., 2018. Adapting to radical change: the benefits of short-horizon investors. Unpublished working paper. Stockholm School of Economics.
- Gryglewicz, S., Hartman-Glaser, B., 2019. Investment timing and incentive costs. *Rev. Financ. Stud.* 33 (1), 309–357.
- Gryglewicz, S., Mancini, L., Morellec, E., Schroth, E.J., Valta, P., 2019. Cash flow shocks and financing policies. Unpublished working paper. Swiss Finance Institute.
- Gutierrez, G., Philippon, T., 2017. Investment-less growth: an empirical investigation. Unpublished working paper. New York University.
- Hackbarth, D., Rivera, A., Wong, T.Y., 2018. Optimal short-termism. Unpublished working paper. Boston University.
- He, Z., 2009. Optimal executive compensation when firm size follows geometric Brownian motion. *Rev. Financ. Stud.* 22 (2), 859–892.
- He, Z., 2011. A model of dynamic compensation and capital structure. *J. Financ. Econ.* 100 (2), 351–366.
- Hoffmann, F., Pfeil, S., 2010. Reward for luck in a dynamic agency model. *Rev. Financ. Stud.* 23 (9), 3329–3345.
- Hoffmann, F., Pfeil, S., 2018. Dynamic multitasking and managerial investment incentives. Unpublished working paper. Erasmus University Rotterdam.
- Kogan, L., 2004. Asset prices and real investment. *J. Financ. Econ.* 73, 411–432.
- Lee, D., Shin, H., Stulz, R., 2018. Why does capital no longer flow more to the industries with the best growth opportunities. Unpublished working paper. Ohio State University.
- Leland, H., 1994. Corporate debt value, bond covenants, and optimal capital structure. *J. Financ.* 49 (4), 1213–1252.
- Malenko, A., 2019. Optimal dynamic capital budgeting. *Rev. Econ. Stud.* 86 (4), 1747–1778.
- Marinovic, I., Varas, F., 2019. CEO horizon, optimal pay duration, and the escalation of short-termism. *J. Financ.* 74 (4), 2011–2053.
- Miao, J., Rivera, A., 2016. Robust contracts in continuous time. *Econometrica* 84, 1405–1440.
- Morellec, E., Schürhoff, N., 2011. Corporate investment and financing under asymmetric information. *J. Financ. Econ.* 99 (2), 262–288.
- Narayanan, M., 1985. Managerial incentives for short-term results. *J. Financ.* 40 (6), 1469–1484.
- Rampini, A., Viswanathan, S., 2013. Collateral and capital structure. *J. Financ. Econ.* 109, 466–492.
- Sannikov, Y., 2008. A continuous-time version of the principal-agent problem. *Rev. Econ. Stud.* 75 (3), 957–984.
- Shreve, S., 2004. *Stochastic Calculus for Finance II: Continuous-time Models*. Springer Verlag, Berlin.
- Stein, J., 1989. Efficient capital markets, inefficient firms: a model of myopic corporate behavior. *Q. J. Econ.* 104 (4), 655–669.
- Strebulaev, I., 2007. Do tests of capital structure theory mean what they say? *J. Financ.* 62 (4), 1747–1787.
- Szydlowski, M., 2019. Incentives, project choice, and dynamic multitasking. *Theor. Econ.* 14 (3), 813–847.
- Thakor, R., 2018. A theory of efficient short-termism. Unpublished working paper. University of Minnesota.
- Varas, F., 2017. Managerial short-termism, turnover policy, and the dynamics of incentives. *Rev. Financ. Stud.* 31 (9), 3409–3451.
- Von Thadden, E.L., 1995. Long-term contracts, short-term investment and monitoring. *Rev. Econ. Stud.* 62 (6), 557–575.
- Zhu, J., 2012. Optimal contracts with shirking. *Rev. Econ. Stud.* 80 (2), 812–839.
- Zhu, J., 2018. Myopic agency. *Rev. Econ. Stud.* 85, 1352–1388.