

DISCUSSION PAPER SERIES

DP12720

**AGENCY CONFLICTS OVER THE SHORT
AND LONG RUN: SHORT-TERMISM,
LONG-TERMISM, AND PAY-FOR-LUCK**

Sebastian Gryglewicz, Simon Mayer and Erwan
Morellec

FINANCIAL ECONOMICS



AGENCY CONFLICTS OVER THE SHORT AND LONG RUN: SHORT-TERMISM, LONG-TERMISM, AND PAY-FOR-LUCK

Sebastian Gryglewicz, Simon Mayer and Erwan Morellec

Discussion Paper DP12720
Published 13 February 2018
Submitted 13 February 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Sebastian Gryglewicz, Simon Mayer and Erwan Morellec

AGENCY CONFLICTS OVER THE SHORT AND LONG RUN: SHORT-TERMISM, LONG-TERMISM, AND PAY-FOR-LUCK

Abstract

We develop a dynamic agency model in which the agent controls current earnings via short-term effort and firm growth via long-term effort and the firm is subject to both short- and long-run shocks. Under the optimal contract, agency conflicts can induce both over- and underinvestment in short- and long-term efforts compared to first best, leading to short- or long-termism in corporate policies. Exposure to long-run shocks introduces pay-for-luck in incentive compensation but only after sufficiently good performance due to incentive compatibility, thereby rationalizing the asymmetric benchmarking observed in the data. Correlated short- and long-run shocks to earnings and firm size lead to externalities in incentive provision over different time horizons.

JEL Classification: N/A

Keywords: N/A

Sebastian Gryglewicz - gryglewicz@ese.eur.nl
Erasmus University Rotterdam

Simon Mayer - mayer@ese.eur.nl
Erasmus University Rotterdam

Erwan Morellec - erwan.morellec@epfl.ch
EPFL and CEPR

Acknowledgements

We thank Bruno Biais, Peter DeMarzo, Zhiguo He, Julien Hugonnier, Lukas Schmid, Stéphane Villeneuve, and Neng Wang for useful comments and discussions.

Agency Conflicts over the Short and Long Run: Short-Termism, Long-Termism, and Pay-for-Luck*

Sebastian Gryglewicz[†] Simon Mayer[‡] Erwan Morellec[§]

January 19, 2018

Abstract

We develop a dynamic agency model in which the agent controls current earnings via short-term effort and firm growth via long-term effort and the firm is subject to both short- and long-run shocks. Under the optimal contract, agency conflicts can induce both over- and underinvestment in short- and long-term efforts compared to first best, leading to short- or long-termism in corporate policies. Exposure to long-run shocks introduces pay-for-luck in incentive compensation but only after sufficiently good performance due to incentive compatibility, thereby rationalizing the asymmetric benchmarking observed in the data. Correlated short- and long-run shocks to earnings and firm size lead to externalities in incentive provision over different time horizons.

Keywords: Agency conflicts; Multi-tasking; Pay-for-luck; Optimal short- and long-termism.

*First draft: July 4, 2017. We thank Bruno Biais, Peter DeMarzo, Zhiguo He, Julien Hugonnier, Lukas Schmid, and Stéphane Villeneuve for useful comments and discussions. Erwan Morellec acknowledges financial support from the Swiss Finance Institute.

[†]Erasmus University Rotterdam. Email: gryglewicz@ese.eur.nl

[‡]Erasmus University Rotterdam. E-mail: mayer@ese.eur.nl.

[§]EPF Lausanne, Swiss Finance Institute, and CEPR. E-mail: erwan.morellec@epfl.ch.

Most corporate decisions are affected by agency conflicts between firm owners and managers due to the separation between ownership and control. Because managers have effective control over corporate policies, they can take actions that are in their own but not necessarily in the interest of the shareholders. In general, providing optimal incentives to management is difficult because many actions executives take are unobservable and their effects on firm performance unravel over different horizons. Indeed, many managerial decisions may not have immediate consequences—such as whether to maintain an R&D division or to launch a new product line—but may be key for future performance. By contrast, other actions are mainly of short-lived relevance, such as diverting output as private benefits or exerting effort to improve production. As a result, agency conflicts over both the short and the long run are likely to arise, with both types of conflicts having different implications for optimal incentives, effort provision, corporate policies, and firm performance.

Although the empirical literature emphasizes the importance of both short- and long-run incentives in managerial compensation (see Jensen and Murphy (1990) or Frydman and Saks (2010)), principal-agent models generally focus either on short-run agency conflicts, assuming that the manager can affect current but not future firm performance as in Holmstrom and Milgrom (1987), DeMarzo and Sannikov (2006), and DeMarzo, Fishman, He, and Wang (2012), or on long-run agency conflicts, assuming that managerial decisions exclusively have long-term effects as in He (2009, 2011). In this paper, we advance the literature by developing a dynamic agency model in which the agent can exert unobservable effort to affect both current earnings and firm growth. We then examine the implications of long- and short-run agency conflicts for optimal compensation and corporate policies and use the model to shed light on the relation between pay-for-luck, short-termism, and firm performance.

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) or Bolton, Wang, and Yang (2017), earnings are proportional to firm size, which is stochastic and governed by a geometric Brownian Motion. In contrast with these models, earnings are also subject to moral hazard and to short-term shocks that do not necessarily affect (or correlate with) long-term prospects (i.e. shocks to firm value). The agent controls the drifts of both the earnings process and the firm size process through unobservable effort.

As a result, the agent can exert effort for the short run to stimulate current earnings and/or effort for the long run to increase the rate at which the firm grows.

Exerting short- or long-term effort is costly, which requires the compensation contract to provide sufficient incentives to the agent, as reflected by sufficient exposure to firm performance. 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 and the liquidation of the firm are triggered, which occurs once the agent's expected wealth (or continuation payoff) falls to zero. Since liquidation is inefficient and generates deadweight costs, maintaining incentive compatibility is costly. To lower the likelihood of liquidation, the optimal contract postpones payments to the manager (thereby increasing the growth rate of manager's continuation payoff) and makes cash payments only after sufficiently good past performance. Deferring compensation is however costly as the manager is less patient than the investor. Based on these tradeoffs, the paper derives an incentive compatible contract that maximizes the value that the principal derives from owning the firm while satisfying the limited liability condition and the participation constraint of the agent. The optimal contract specifies as a function of past performance (1) the manager's compensation, (2) the level of short- and long-run effort, and (3) whether the contract is terminated and the firm is liquidated.

To solve the model, we use the agent's continuation payoff—formally introduced in Spear and Srivastava (1987)—and firm size as state variables. These two state variables completely summarize the relevant history of the firm. Moreover, because of the size-homogeneity of the model, the agent's continuation payoff per unit of firm size becomes sufficient for the contract-relevant history of the firm, as in e.g. He (2009) or DeMarzo, Fishman, He, and Wang (2012). An important consequence of this property of the model is that we are able to characterize the optimal contract by means of an ordinary differential equation, as in models that focus either on short-term or on long-term agency conflicts (see the discussion above). Based on this characterization, we then analytically examine the implications of the optimal contract for long- and short-run incentives and for managerial effort provision.

We highlight our main findings. Considering first effort choice, we depart substantially from contributions with one-dimensional moral hazard, such as He (2009), DeMarzo and Sannikov (2006) or DeMarzo, Fishman, He, and Wang (2012), in that first-best effort is not

always implemented. In our model, the principal optimally balances the costs and benefits of incentivizing the manager over the short or the long run, leading to both overinvestment, i.e. effort above the first-best level, and underinvestment, i.e. effort below the first-best level. Our analysis demonstrates for example that short-term effort is at its first-best level at the payout boundary where the manager receives direct cash payments, while long-term effort is below first-best. Below the payout boundary, underinvestment can arise because incentivizing the manager is costly, which increases the cost of effort to the principal and thus reduces the optimal level of effort. Yet, overinvestment in short-term effort can also arise when the firm is financially weak and decides to promote short-term effort at the expense of long-term effort to reduce the likelihood of inefficient termination. Relatedly, overinvestment in long-term effort can arise when the firm becomes financially stronger because it is then optimal to favor firm growth at the expense of current earnings.

Our model therefore predicts that a firm may optimally exhibit short- and long-termism, depending on whether short- and long-term effort are above or below first-best. This contrasts with the literature on managerial short-termism (Stein (1989) and Shleifer and Vishny (1990)), in that we show that the same firm can find it optimal at times to be short-termist and at other times to be long-termist. Our results are therefore in line with the views recently expressed in Harvard Business Review¹ that companies “feel they need to strike the right “balance” between the short term and long term” and in the Financial Times² that “while there little doubt that too much short-termism has negative effects, one should not assume that it follows that extreme long-termism is always for the best.”

Corporate short-termism has been the subject of debate among academics, with much of the discussion focusing on whether it destroys value.³ A recent study by Barton, Manyika, and Williamson (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-

¹See “Long-termism is just as bad as short-termism,” September 25, 2014.

²See “Cultish long-termism can hobble investors,” September 12, 2017.

³Asker, Farre-Mensa, and Ljunqvist (2015) show that closely held firms tend to invest more than similar publicly listed companies, consistent with the idea that managers in public firms tend to take actions that can deliver immediate returns to shareholders. Bernstein (2015) finds that when firms go public their best inventors tend to leave and those who remain produce fewer patents. Gutierrez and Philippon (2017) find that the more companies are held by institutional investors, the less they tend to invest. All these papers suggest that publicly held companies may focus on short-term goals at the expense of long-term value creation. By contrast, Kaplan (2017) shows that the presumed short-term orientation of managers does not show up in corporate profits, price-earnings ratios, or venture capital investments and returns.

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.⁴ Notably, firms with a high risk of liquidation (i.e. firms that perform worse) find it optimal to focus on the short term in our model and while firms with a low risk of liquidation find it optimal to focus on the long term. Our paper additionally finds that short-termism is more likely to be optimal when the volatility of the firm value process is high and the correlation between shocks to earnings and firm value is positive, i.e. when stock return volatility is high consistent with the evidence in Brochet, Loumioti, and Serafeim (2012, 2013). By contrast, long-termism is more likely to be optimal when the volatility of the firm value process is low and the correlation between shocks to earnings and firm value is negative, i.e. when stock return volatility is low.

As in prior contributions, incentives are provided in the optimal contract by making the agent’s compensation contingent on firm performance, in particular by rewarding (punishing) the agent when outcomes are better (worse) than expected. As shown in Hoffmann and Pfeil (2010) or DeMarzo et al. (2012), long-run, permanent shocks make it optimal to introduce pay-for-luck in incentive compensation, i.e. exposure to volatility that is not needed to incentivize effort. Indeed, a positive permanent shock makes liquidation more inefficient. As a result, the agent’s promised wealth under the optimal contract increases in response to a positive shock in order to reduce the likelihood of termination. While our model shares this feature with prior contributions, the principal additionally needs to incentivize the manager to exert long-run effort in our framework with multi-tasking. This generates the distinct prediction that pay-for-luck is only present when the agent’s stake in the firm is large enough. Notably, we show that when the continuation payoff of the manager is low, pay-for-luck does not suffice to provide incentives to the manager so that pay-for-luck is shut down. When the continuation payoff is high enough, positive shocks increase pay-for-luck and negative shocks decrease and eventually eliminate it. Our model therefore provides a rationale for the asymmetry of pay-for-luck observed in the executive compensation data (see e.g. Garvey and Milbourn (2006) and Francis, Iftekhhar, Kose, and Zenu (2013)).

⁴Interestingly, 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 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.”

We also show that if luck shocks are not informative about future performance, it is suboptimal to make the agent's payments contingent on these shocks, in accordance with Holmstrom (1979). This is the case in our model of short-term cash flow shocks that are uncorrelated with the permanent shocks to firm size i.e. of shocks that do not affect long-term prospects. As in prior contributions (see e.g. DeMarzo and Sannikov (2006)), the incentive compatibility constraint with respect to short-term effort is therefore always binding, in that minimum incentives (or equivalently exposure to short-term shocks) are provided.

Another unique feature of incentive compensation in our model is that it is based on performance measures that are both in terms of levels and in terms of growth rates, due to the dual nature of agency conflicts. In our model, performance measures in levels relate to short-run incentives while performance measures in growth rates relate to long-run incentives. As documented in Edmans, Gabaix, and Jenter (2017), stock compensation, the dominant form of incentive pay in the recent years, has been increasingly performance-based. Performance-vesting is now more popular than time-vesting. Usually multiple performance criteria are used, some in terms of levels and some in growth rates, consistent with our findings.

Lastly, we find that the correlation between short- and long-run shocks is a key determinant of incentive provision, effort choice, and firm performance. Notably, a positive correlation, by increasing the volatility of the continuation payoff of the manager, leads to externalities between effort choices with higher correlation leading to decreased long-term effort and incentives. More generally, we demonstrate that short-run and long-run effort are either substitutes (for positive correlation) or complements (for negative correlation) in the firm's production function. This in turn implies that, depending on firm characteristics, short-term profits may not come at the expense of firm growth.

Our paper is related to the growing literature on dynamic agency conflicts. 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 DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin, and Rochet (2007), Sannikov (2008), Zhu (2012), Malenko (2016), Miao and Rivera (2016), or Nikolov, Schmid, and Steri (2017). 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. In these last two models, earnings are not subject to short-term moral hazard. Our analysis demonstrates

that while in models with one-dimensional moral hazard the firm never overinvests, this need not be the case in our model in which the principal optimally balances the costs and benefits of incentivizing the manager over the short- or long-term, leading to overinvestment and short-termism when the likelihood of termination is high or long-termism when it is low.

Building on the insights of Hoffmann and Pfeil (2010) and DeMarzo, Fishman, He, and Wang (2012), we show that exposure to long-run, persistent shocks leads to pay-for-luck in incentive compensation. In our model however, pay-for-luck may not be sufficient to incentivize the manager to exert long-term effort. Pay-for-luck therefore only arises after sufficiently good performance, thereby rationalizing the asymmetric benchmarking in executive compensation observed in the data. DeMarzo, Fishman, He, and Wang (2012) and Biais, Mariotti, Rochet, and Villeneuve (2010) analyze corporate investment under agency frictions over the short run. In these papers, investment is observable and therefore not subject to moral hazard. In our model, investment externalities imply that short-term moral hazard feeds back into investment decisions. Similarly, Gryglewicz and Hartman-Glaser (2017) and Gryglewicz, Hartman-Glaser, and Zheng (2017) analyze lumpy investment assuming that the decision to invest is at the discretion of the shareholders. Sannikov (2014) considers dynamic moral hazard over one task that has delayed persistent consequences. Szydlowski (2015) studies the optimal choice of investment projects in a dynamic moral hazard model with multitasking. His model features multiple agency conflicts that are exclusively over the short run. In contemporaneous work, Hackbarth, Rivera, and Wong (2017) use a dynamic agency model to show that shareholder-debtholder conflicts may make short-termism optimal for shareholders. Their analysis focuses on the agency costs of debt and does not analyze optimal long-termism or pay-for-luck. Lastly, our modeling of cash flows with permanent and transitory shocks is similar to that in Décamps, Gryglewicz, Morellec, and Villeneuve (2017). Their model does not feature agency conflicts.

The paper is organized as follows. Section 1 presents the model and its solution. Section 2 analyzes the implications of the model for optimal effort and incentives. Section 3 concludes. Technical developments are gathered in the Appendix.

1 The Setting and the Optimal Contract

1.1 The Model

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 this model, there is separation between ownership and control and agency problems arise because the agent can take hidden actions that affect current earnings as well as the growth rate of the firm.

Specifically, we consider a firm that employs physical capital for production and denote by K_t the level of the capital stock at time $t \geq 0$. Earnings are proportional to the capital stock K_t (i.e. the firm employs an “AK” technology) and depend on the agent's unobservable effort choices. Earnings are also 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. Notably, following He (2009) and Bolton, Wang, and Yang (2017), we consider that the firm's capital stock (firm size) $\{K\} = \{K_t\}_{t \geq 0}$ evolves according to the controlled geometric Brownian motion process:⁵

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

where $\mu > 0$ is a constant, $\delta > 0$ is the expected rate of depreciation, $\sigma_K > 0$ is a constant volatility parameter, $\ell_t \in [0, \ell_{\max}]$ is the manager's long-term effort choice, and $\{Z^K\} = \{Z_t^K\}_{t \geq 0}$ is a standard Brownian motion. 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:

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

where α and σ_X are strictly positive constants, $s_t \in [0, s_{\max}]$ is the manager's short-term

⁵This 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, Ingersoll, and Ross (1985) or Kogan (2004)), corporate finance (e.g. Abel and Eberly (2011) or Bolton, Wang, and Yang (2017)), or macroeconomics (e.g. Gertler and Kiyotaki (2010) or Brunnermeier and Sannikov (2014)).

effort 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 that:⁶

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

While $\{K\}$ and $\{X\}$ are observable and contractible, the manager's effort choices $\{s\}$ and $\{\ell\}$ are not. There are therefore two sources of agency conflicts as the principal cannot disentangle the manager's actions from the Brownian shocks in equations (1) and (2).

As shown by equations (1) and (2), long-term effort improves firm growth while short-term effort improves current earnings.⁷ Exerting effort however is costly for the manager. Notably, the cost of effort is convex in the effort choice and given by:⁸

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

where λ_s and λ_ℓ are positive constants. This specification implies that the cost of effort (or, equivalently, the cost of investment) increases with firm size as administering a larger firm requires more effort; a similar assumption is made for example in He (2009). An important difference between the two models is that while the return on invested capital is constant in He (2009), in that $\frac{dX_t}{K_t} = \alpha dt$, this is not the case in our model in which it is subject to short-term moral hazard and short-run shocks, in that $\frac{dX_t}{K_t} = s_t \alpha dt + \sigma_X dZ_t^X$.

The principal discounts future cash-flows at the rate $r > \mu \ell_{\max} > 0$. As in DeMarzo and Sannikov (2006), Biais et al. (2007), or DeMarzo et al. (2012), the agent is more impatient and has a discount rate $\gamma > r$. This implies that the principal cannot indefinitely postpone payments to the agent. The agent possesses an outside option (her next best employment

⁶In general, the correlation coefficient ρ between short-term and permanent cash flow shocks can be positive or negative. Examples of a negative correlation include decisions to invest in R&D 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 existing supply of a commodity or improving technology for the production and discovery of a commodity. Chang, Dasgupta, Wong, and Yao (2014) estimate that for firms listed in the Compustat Industrial Annual files between 1971 and 2011, the correlation between short-term and permanent cash flow shocks is negative.

⁷Alternatively, short-term effort choice measures how effectively the manager is using the firm's capital while long-term effort choice can be thought of as how much effort the manager puts in the selection/adoption of investment projects. We thank Lukas Schmid for suggesting this interpretation.

⁸Note that our specification implies that the standard convex adjustment cost function for capital investment used in the neoclassical investment literature can be seen as a cost of incentivizing the manager. Our model can easily incorporate additional costs unrelated to managerial effort.

opportunity) normalized to zero and is protected by limited liability, ruling out negative wages.⁹ Without loss of generality, the manager cannot maintain a savings account.¹⁰ The manager's employment starts at time $t = 0$ and is terminated once the firm is liquidated at an endogenous stopping time τ . At the time of liquidation, the principal recovers a fraction $R \geq 0$ of assets, valued as RK_τ . Liquidation is inefficient and generates deadweight losses. It is however necessary to incentivize the manager who is protected by limited liability.

To maximize firm value, the investor offers a contract to the agent at time $t = 0$ and commits to a compensation scheme $\{C\}$, recommended effort processes $\{\hat{s}\}$ and $\{\hat{\ell}\}$, and a termination time τ . Because the agent has limited liability, the process $\{C\}$ is non-decreasing. We let $\Pi \equiv (\{C\}, \{\hat{s}\}, \{\hat{\ell}\}, \tau)$ represent the contract, where all elements are progressively measurable with respect to \mathbb{F} , and assume that the processes $\{\hat{s}\}$ and $\{\hat{\ell}\}$ are of bounded variation and that Π satisfies the standard square integrability condition:¹¹

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

We call a contract incentive compatible if $a_t = \hat{a}_t$ for $a \in \{s, \ell\}$ for all $\tau \geq t \geq 0$. Since we focus without loss of generality on incentive compatible contracts, we will not formally distinguish between actual and prescribed effort levels.

Before proceeding, note that the modeling of cash flows in (1) and (2) encompasses two popular frameworks as special cases. When $\sigma_K = 0$, we obtain the stationary environment of the dynamic agency models of DeMarzo and Sannikov (2006) and DeMarzo, Fishman, He, and Wang (2012). Models analyzing the effects of financing frictions on corporate decisions, such as Bolton, Chen, and Wang (2011), Décamps, Mariotti, Rochet, and Villeneuve (2011) or Hugonnier, Malamud, and Morellec (2015), also employ this cash-flow environment. Since there is no noise to hide the agent's long-run effort choice, the long-run agency conflict is irrelevant in that case. By contrast, when $\sigma_X = 0$, there are no short-run shocks and we obtain the cash-flow environment used in the dynamic capital structure (e.g. Leland (1994)

⁹As in Albuquerque and Hopenhayn (2004) or Rampini and Viswanathan (2013), we could assume that the manager is able to 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 .

¹⁰As in DeMarzo and Sannikov (2006), it is possible to show that savings cannot be part of the optimal contract. Indeed, because $\gamma > r$, it is cheaper for the principal to save and make direct payments to the agent. As a consequence, the manager consumes all payments she receives immediately.

¹¹We will refer to these conditions collectively as the 'usual regularity condition'.

or Strebulaev (2007)) and real options literatures (e.g. Carlson, Fisher, and Giammarino (2006) or Morellec and Schürhoff (2011)) as well as in the dynamic agency models of He (2009, 2011). Since there is no noise to hide the agent's short-run effort choice, the short-run agency conflict is irrelevant in that case. Section 2.4 below analyzes incentives and effort choice when moral hazard only affects either short-run effort *or* long-run effort.

1.2 The Contracting Problem

Consider a contract Π , fix the effort processes $\{s\}$ and $\{\ell\}$, and define the agent's expected payoff at time $t = 0$ as

$$W(\Pi) \equiv \mathbb{E}^{s,\ell} \left[\int_0^\tau e^{-\gamma t} dC_t - \int_0^\tau e^{-\gamma t} \frac{1}{2} K_t (\lambda_s s_t^2 \alpha + \lambda_\ell \ell_t^2 \mu) dt \right],$$

where the expectation operator $\mathbb{E}^{s,\ell}[\cdot]$ is under the probability measure induced by $(\{s\}, \{\ell\})$. Since we focus on incentive compatible contracts, we suppress the dependence on the effort choices $(\{s\}, \{\ell\})$ and just write $\mathbb{E}[\cdot]$ in the following. Intuitively, $W = W(\Pi)$ equals the promised value the agent gets if she follows the recommended path from time $t = 0$ onwards, net of the cost of implementing the recommended effort level.

The investor receives the firm cash flows and pays the compensation to the manager. As a result, given the contract Π , the investor's expected payoff can be written as:

$$P(W_0, K_0) \equiv \mathbb{E} \left[\int_0^\tau e^{-rt} dX_t + e^{-r\tau} RK_\tau - \int_0^\tau e^{-rt} dC_t \right].$$

The objective of the principal is therefore 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 (see e.g. DeMarzo, Fishman, He, and Wang (2012) for alternative specifications). Denote the set of incentive compatible contracts by \mathbb{IC} . Given an initial promise to the manager W_0 and an initial firm size K_0 , the investor's optimization problem reads

$$\max_{\Pi \in \mathbb{IC}} P(W_0, K_0) \text{ such that } W(\Pi) = W_0 \geq 0, \text{ and } P(W_0, K_0) \geq RK_0. \quad (3)$$

We denote the solution to this problem as $\Pi^* \equiv (C^*, s^*, \ell^*, \tau^*)$.

1.3 Benchmark Case: Effort Choice under First-Best

We start by deriving the value of the firm and the optimal effort levels absent agency conflicts. This is the case when there is no noise to hide the agent's action, so 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

$$P^{FB}(K_0) = \frac{K_0}{r + \delta - \mu\ell} \left[\alpha s - \frac{1}{2} (\lambda_s \alpha s^2 + \lambda_\ell \mu \ell^2) \right],$$

where the short- and long-term effort choice maximize the value of the firm. The first order condition with respect to s implies $s^{FB} = \frac{1}{\lambda_s}$ (when effort choice is interior). After imposing this first best choice of short term effort, the first order condition with respect to ℓ reads

$$\frac{\mu}{2(r + \delta - \mu\ell)^2} \left(\frac{\alpha}{\lambda_s} - \lambda_\ell \mu \ell^2 \right) - \frac{\lambda_\ell \mu \ell}{r + \delta - \mu\ell} = 0,$$

which can be rewritten as

$$\Phi(\ell) \equiv \frac{\alpha}{\lambda_s} + \lambda_\ell \mu \ell^2 - 2(r + \delta) \lambda_\ell \ell = 0.$$

Due to the second order condition of the maximization, it must be that $\frac{\partial^2 P^{FB}(K_0)}{\partial \ell^2} < 0$, which implies $\Phi'(\ell^{FB}) < 0$. Solving yields the following result:

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

1. *The first-best short-term effort level satisfies: $s^{FB} = \frac{1}{\lambda_s}$.*
2. *The first-best long-term effort level satisfies: $\ell^{FB} = \frac{1}{\mu} \left[r + \delta - \sqrt{(r + \delta)^2 - \frac{\mu\alpha}{\lambda_\ell \lambda_s}} \right]$.*
3. *The first-best firm value is given by:*

$$P^{FB}(K_0) = \frac{K_0}{r + \delta - \mu\ell^{FB}} \left[\frac{\alpha}{2\lambda_s} - \frac{1}{2} \lambda_\ell \mu (\ell^{FB})^2 \right].$$

1.4 Model Solution

We now solve the model with agency conflicts over the short and long run, that is assuming $\sigma_K \lambda_\ell > 0$ and $\sigma_X \lambda_s > 0$. In our model, λ_s (λ_ℓ) reflects the potential gains for the agent of deviating from the recommended effort choice over the short run (long run). In addition, σ_X and σ_K reflects the difficulty for the principal of detecting shirking. When cash-flow and/or firm size are volatile, it becomes difficult for the principal to judge whether adverse outcomes are due to bad luck or to the agent shirking. As a consequence, if either λ_s (λ_ℓ) or σ_X (σ_K) is sufficiently small, firm value is hardly affected by moral hazard over the short run (long run) because providing incentives is relatively easy.

Recall that the compensation contract of the manager Π specifies a compensation scheme $\{C\}$, effort processes $\{s\}$ and $\{\ell\}$, and a termination time τ . Denote by W_t the continuation payoff of the agent at time $t \geq 0$ under this contract. The optimal contract relies on cash payments dC_t and changes in the value of future payments dW_t . The manager will continue within the firm if and only if promised future transfers exceed the value of her outside option. To compensate the agent for her time preference and effort cost, incremental compensation $dC_t + dW_t$ must equal $(\gamma W_t + \mathcal{C}(s_t, \ell_t))dt$ on average:

$$\mathbb{E}[dC_t + dW_t] = \left[\gamma W_t + \frac{1}{2} K_t (\lambda_s s_t^2 \alpha + \lambda_\ell \ell_t^2 \mu) \right] dt. \quad (4)$$

Equation (4) corresponds to the ‘Promise Keeping Condition’ in the discrete time formulation of DeMarzo and Fishman (2007). While this condition determines how much the agent should earn on average, her compensation must also be sufficiently sensitive to firm performance, as captured by dX_t and dK_t , to maintain incentive compatibility. By punishing (rewarding) the manager if either asset growth or cash-flow is worse (better) than expected—i.e. falls short of (exceeds) its expectation—incentive compatibility is maintained. Using the martingale representation theorem, this sensitivity of the manager’s incremental payoff on output and productivity changes can be formalized as follows (see Appendix A):

$$dW_t + dC_t = \gamma W_t dt + \frac{1}{2} K_t (\lambda_s s_t^2 \alpha + \lambda_\ell \ell_t^2 \mu) dt + \beta_t^s (dX_t - \alpha s_t K_t dt) + \beta_t^\ell (dK_t - (\mu \ell_t - \delta) K_t dt), \quad (5)$$

where the sensitivities β_t^s and β_t^ℓ are used to satisfy incentive compatibility conditions with respect to the short- and long-run efforts, respectively. To understand why such a compensa-

tion scheme may lead the agent to exert effort, suppose that the agent decides to marginally deviate from the recommended choice \hat{s}_t over the time interval $[t, t + dt)$. By doing so, she saves cost $\hat{s}_t \alpha K_t \lambda_s dt$. However, as a consequence of her behavior, realized earnings fall below their expectation and thus her compensation is reduced by $\alpha K_t \beta_t^s dt$. Incentivizing the manager to exert effort therefore requires that $\beta_t^s = \lambda_s \hat{s}_t$ if \hat{s}_t is interior and $\beta_t^s \geq \lambda_s \hat{s}_t$ if $\hat{s}_t = s_{\max}$. Similarly, $\beta_t^\ell = \lambda_\ell \hat{\ell}_t$ if $\hat{\ell}_t$ is interior and $\beta_t^\ell \geq \lambda_\ell \hat{\ell}_t$ if $\hat{\ell}_t = \ell_{\max}$. Both incentive compatibility constraints require that the agent has enough *skin in the game*, 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 continuation payoff of the agent W and firm value K , the scale invariance of the firm's environment allows us to write $P(W, K) = Kp(w)$ and to reduce the problem to a single state variable: $w \equiv \frac{W}{K}$, the scaled promised payments to the agent as in He (2009) or DeMarzo, Fishman, He, and Wang (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 $p'(w) \geq -1$. That is, the marginal cost of delaying payouts can never exceed the cost of an immediate transfer. As shown by equation (5), deferring compensation increases the growth rate of w and produces direct benefits by lowering the risk of liquidation when w is close to zero. As a result, the optimal contract will set $dc \equiv \frac{dC}{K}$ to zero for low values of w . However, due to the relative impatience of the agent ($\gamma > r$), postponing payments is costly. Since the benefits of delaying payouts decrease while the cost increases in w , we may conjecture that $p(w)$ is concave and that there exists a threshold \bar{w} above which it is optimal to directly pay the manager, i.e. such that

$$p'(\bar{w}) = -1 \text{ and } dc = \max\{0, w - \bar{w}\}, \quad (6)$$

where the optimal payout boundary is determined by the smooth-pasting condition:

$$p''(\bar{w}) = 0. \quad (7)$$

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

$$p(0) = R. \quad (8)$$

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

$$(r + \delta)p(w) = \max_{s, \ell, \beta^s, \beta^\ell} \left\{ \alpha s + \mu \ell p(w) + p'(w)w(\gamma + \delta - \mu \ell) + \frac{p'(w)}{2} (\lambda_\ell \mu \ell^2 + \lambda_s \alpha s^2) \right. \\ \left. + \frac{p''(w)}{2} [(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w)] \right\}, \quad (9)$$

subject to the incentive compatibility constraints on β^s and β^ℓ . Evaluating the above HJB-equation at the payout boundary \bar{w} and using equations (6) and (7) yields

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

Observe that postponing cash-payments by increasing \bar{w} reduces the risk of termination. Delaying compensation is efficient until the investor's and manager's required returns $(r + \delta - \mu \ell)p(\bar{w}) + \bar{w}(\gamma + \delta - \mu \ell)$ exhaust the available expected net cash-flow $\alpha s - \frac{1}{2} (\lambda_s s^2 \alpha + \lambda_\ell \ell^2 \mu)$. Note that the required total rate of return (of both the investor and the manager) is reduced by the growth rate $\mu \ell$ since both benefit from the expanding firm size, leading to effective discount rates $\gamma + \delta - \mu \ell$ and $r + \delta - \mu \ell$. Consequently, the prospect of higher future payments makes the manager and the investor loosely speaking more patient.

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

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

with unique solution $w_0 = w^*$ satisfying

$$p'(w^*) = 0. \quad (10)$$

As a consequence, the principal initially promises the agent utility $w^* K_0$ and ex-ante expects

payoff $P(K_0 w^*, K_0) = p(w^*)K_0$.¹² To close this section, we summarize our results about the optimal contract in the following Proposition. Its proof is deferred to Appendix A.

Proposition 2 (Firm value and optimal compensation under agency). *Suppose that $\lambda_i \sigma_j > 0$ for $j \in \{X, K\}$ and $i \in \{s, \ell\}$ and let $\Pi^* \equiv (C^*, s^*, \ell^*, \tau^*)$ denote the optimal contract solving problem (3). 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 (5). The optimal contract is incentive compatible in that $\beta^s = \lambda_s s(w)$ for $s(w) < s_{\max}$ and $\beta^s \geq \lambda_s s(w)$ for $s(w) = s_{\max}$ and $\beta^\ell = \lambda_\ell \ell(w)$ for $\ell(w) < \ell_{\max}$ and $\beta^\ell \geq \lambda_\ell \ell(w)$ for $\ell(w) = \ell_{\max}$.*
2. *The investor's value function $P(W, K)$ is proportional to firm size and satisfies $P(W, K) = Kp(w)$, where $p(w)$ is the unique solution to equation (9) subject to (6), (7), and (8) 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 $[0, \bar{w}]$.*

2 Model Analysis

This section examines the features of the optimal compensation policy and the implications of agency conflicts for long- and short-term effort choice. For clarity of exposition, we assume in following two subsections that the correlation ρ between short- and long-run shocks is zero. Section 2.3 analyzes the effects of non-zero correlation on effort choice and incentive provision. Section 2.4 analyzes incentives and effort choice when moral hazard only affects either short-run effort *or* long-run effort. Section 2.5 provides a quantitative analysis of the effects of short- and long-term agency conflicts on investment and incentive provision.

2.1 Incentive Provision and Effort Choice

We start by analyzing optimal effort choice $\{s(w), \ell(w)\}$. We will show in this section that short- and long-run efforts are determined differently by agency frictions. Furthermore,

¹²For convenience, we might drop or normalize K_0 to unity if no confusion is likely to arise. Therefore, we sometimes refer to $p(w^*)$ as expected payoff instead of scaled expected payoff.

moral hazard can lead to either under- and over-investment in both $s(w)$ and $\ell(w)$ relative to the first-best levels s^{FB} and ℓ^{FB} . In our model with short- and long-run agency conflicts, the optimal effort choice $\{s(w), \ell(w)\}$ is determined by $s(w) = \min\{s_{\max}, s^*(w)\}$ and $\ell(w) = \min\{\ell_{\max}, \ell^*(w)\}$, where we respectively allow s and ℓ to take values between $[0, s_{\max}]$ and $[0, \ell_{\max}]$. The value $s^*(w)$ is pinned down using the incentive-compatibility conditions and taking the first-order condition in equation (9). This leads to the following result:

Proposition 3 (Optimal short-term effort). *Optimal short-term effort is given by $s(w) = \min\{s_{\max}, s^*(w)\}$. If the marginal cost of short-term effort to the principal is negative in that $p'(w)\lambda_s\alpha + p''(w)(\lambda_s\sigma_X)^2 > 0$, the optimal short-term effort choice is given by*

$$s(w) = s_{\max}.$$

By contrast, if $p'(w)\lambda_s\alpha + p''(w)(\lambda_s\sigma_X)^2 < 0$, the marginal cost of effort equals its marginal benefit at the optimum and the optimal short-term effort is interior and given by

$$s(w) = \frac{\overbrace{\alpha}^{\text{Direct benefit of effort}}}{\underbrace{-p'(w)\lambda_s\alpha}_{\text{Direct cost of effort}} - \underbrace{p''(w)(\lambda_s\sigma_X)^2}_{\text{Volatility cost of effort}}}. \quad (11)$$

When optimal short-term effort is interior, as determined by equation (11), the marginal benefit of short-run effort is simply the cash flow rate α and is at the first-best level. The direct (marginal) cost of effort for the manager is $\lambda_s\alpha$. Since the investor compensates the manager for this cost by increasing her continuation utility, the cost scales by $-p'(w)$ from the principal's perspective. In the region where $p'(w) > 0$ (that is, for small w), this direct cost effect is negative as increasing w benefits the investor by reducing the risk of inefficient liquidation. When the continuation payoff of the agent w is large, the direct-cost effect is positive. At $w = \bar{w}$, where $p'(\bar{w}) = -1$, the effect is exactly at the first best-level. The second term in the denominator of equation (11), the volatility cost of effort, relates to the concavity of the investor's value function. Because of this concavity, it is costly to increase effort as this requires additional incentives β^s , which increases the volatility of w . The effect is strongest when $p''(w)$ is the largest in absolute value and disappears at $w = \bar{w}$ where $p''(\bar{w}) = 0$. The top two panels of Figure 1 show how the direct cost of effort and the volatility cost of effort

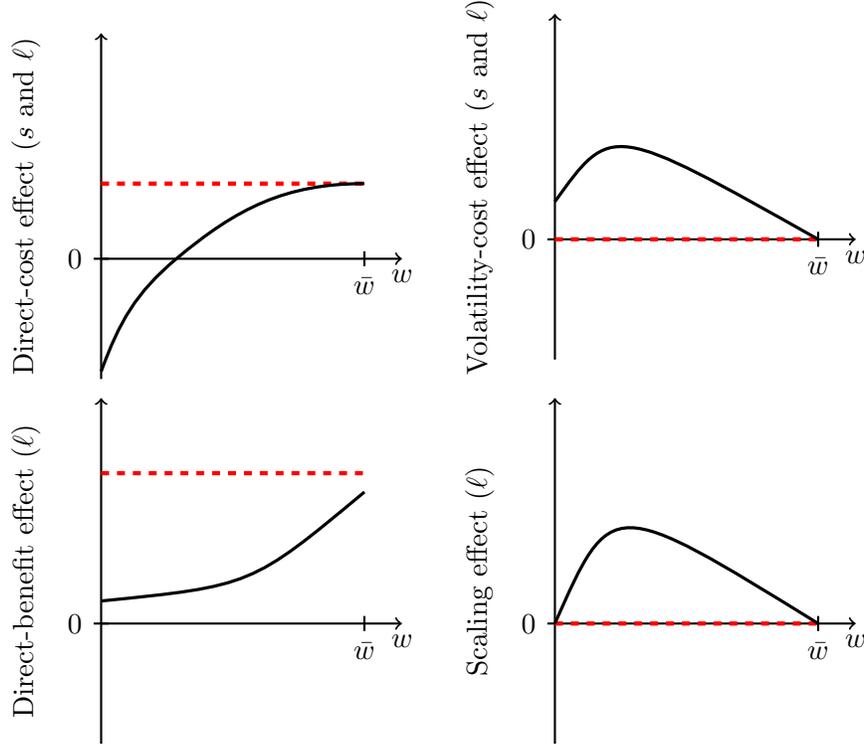


Figure 1: **The determinants of the optimal short- and long-run efforts.** The solid curves represent the four effects in equations (11) and (12) in the typical case in which $p'(0) > 0$. The dashed lines represent the effects in the first-best problem. The direct-benefit effect in the choice of s (not presented) is constant in w and at the first-best level.

vary with the (scaled) promised payments to the agent w .

Similarly, the first order condition with respect to long-term effort in the Hamilton-Jacobi-Bellman equation (9) leads to the following result:

Proposition 4 (Optimal long-term effort). *Optimal long-term effort is given by $\ell(w) = \min\{\ell_{\max}, \ell^*(w)\}$. If the cost of long-term effort to the principal is negative in that $p'(w)\lambda_\ell\mu + p''(w)(\lambda_\ell\sigma_K)^2 > 0$, the optimal long-term effort choice is given by*

$$\ell(w) = \ell_{\max}.$$

By contrast, if $p'(w)\lambda_\ell\mu + p''(w)(\lambda_\ell\sigma_K)^2 < 0$, the marginal cost of long-term effort equals its

marginal benefit at the optimum and the optimal long-term effort is interior and given by

$$\ell(w) = \frac{\overbrace{\mu(p(w) - p'(w)w)}^{\text{Direct benefit of effort}} \overbrace{-p''(w)w\lambda_\ell\sigma_K^2}^{\text{Scaling}}}{\underbrace{-p'(w)\lambda_\ell\mu}_{\text{Direct cost of effort}} \underbrace{-p''(w)(\lambda_\ell\sigma_K)^2}_{\text{Volatility cost of effort}}}. \quad (12)$$

As shown by Propositions 3 and 4, there are two structural differences between equations (11) for short-term effort $s(w)$ and (12) for long-term effort $\ell(w)$. A first difference is that optimal $\ell(w)$ has an additional benefit of effort compared to $s(w)$: the scaling effect. Since $p''(w) \leq 0$, this scaling effect unambiguously increases long-run effort. To understand the source of this effect, note that a positive permanent shock $dZ^K > 0$ has two opposing consequences. First, the agent is rewarded via the sensitivity β^ℓ and is promised higher future payments W . This increases $w = \frac{W}{K}$ by $\beta^\ell dZ^K$, which equals $\lambda_\ell \ell(w) dZ^K$ when the incentive-compatibility constraint is binding and the long-term effort choice is interior. Second, firm size K grows more than expected, thereby diluting the agent's stake $w = \frac{W}{K}$ by $-wdZ^K$ and increasing the likelihood of inefficient termination. Because these two effects move w in the opposite direction, the scaling/dilution effect reduces the volatility of the continuation payoff of the manager (and therefore the cost of incentivizing the manager), leading to an increase in optimal long-term effort.¹³ As for the volatility cost of effort, the strength of the scaling effect depends on how much volatility in w matters for the investor's value function. Therefore, it is the strongest when the concavity of the scaled value function is the largest. Conditional on interior effort, the effect disappears at $w = \bar{w}$ where $p''(\bar{w}) = 0$.

A second difference between optimal short- and long-run efforts is that the direct benefit of effort is constant in (11) while it is scaled by $p(w) - p'(w)w$ in (12). Two observations follow. First, $p(w) - p'(w)w$ is time-varying and increases with w . Second, $p(w) - p'(w)w$ is strictly less than p^{FB} , reducing the direct benefit of long-term effort compared to first best (p^{FB} scales the benefit of effort in the first-best problem). The bottom two panels of Figure 1 plot the two effects specific to the long-run effort as functions of w .

¹³We show below that when the incentive compatibility constraint is not binding because the manager's continuation payoff is large (section 2.2) or in the absence of long-run moral hazard (section 2.4), the principal optimally sets $\beta^\ell = w$ such that $(\beta^\ell - w)\sigma_K dZ^K = 0$. In such instances, the agent is incentivized for the long run without raising the threat of termination and incentive provision optimally involves pay-for-luck.

Using the results in Propositions 3 and 4, we can now analyze the patterns of under- and over-investment in effort relative to the first-best levels. We will refer to over-investment in short-run effort as short-termism and to over-investment in long-run effort as long-termism. Since the optimal effort levels with agency frictions depend on the promised payments w to the agent (and the first best levels do not), so do the investment patterns. In particular, at $w = \bar{w}$, the total marginal cost of short-run effort under moral hazard (the denominator of (11)) is exactly equal to the marginal cost of effort under the first-best. As the benefit of short-run effort is always at the first-best level, the optimal short-run effort under moral hazard is at the first-best level at $w = \bar{w}$, in that $s(\bar{w}) = \frac{1}{\lambda_s} = s^{FB}$.

While short-run agency conflicts are locally resolved at $w = \bar{w}$, this is not the case of long-run agency conflicts. At $w = \bar{w}$ and with interior long-run effort, the scaling effect is absent since $p''(\bar{w}) = 0$. In this case, the weaker direct-benefit effect under agency conflicts generates underinvestment in long-run effort. To understand this result, note that the benefit of investing in long-run effort when $w = \bar{w}$ is proportional to total firm value $p(\bar{w}) + \bar{w}$ (since $p'(\bar{w}) = -1$). This benefit is therefore strictly smaller than the benefit in the first-best firm, which is proportional to p^{FB} . Moreover, at $w = \bar{w}$, as in the case with the short-run effort, the marginal cost of long-run effort is equal to the marginal cost of effort under the first-best. This leads to under-investment in long-run effort in that $\ell(\bar{w}) < \ell^{FB}$. The following Proposition summarizes these results:

Proposition 5. *Assume the usual regularity conditions and that the bounds $(0, i_{\max})$ for $i \in \{s, \ell\}$ are such that first-best effort levels are interior. At the payout boundary, the firm always optimally invests in the short-run in that $s(\bar{w}) = s^{FB}$ and underinvests for the long run in that $\ell(\bar{w}) < \ell^{FB}$.*

Agency frictions generate additional costs and decrease benefits of effort and thus can lead to under-investment in both short- and long-term effort for $w < \bar{w}$. Interestingly, optimal over-investment can also occur for $w < \bar{w}$. This is due to two separate effects. First, the cost of effort under moral hazard can fall below the first-best cost. This is the case when the negative direct-cost effect dominates the volatility-cost effect. This can occur for small w and can affect both the short- and long-run efforts. Second, the benefits of long-run effort under moral hazard can exceed the first-best benefit of effort. This happens when the scaling effect compensates the weakened direct benefit of long-run effort (as shown in Figure 1 this

may arise for intermediate levels of w). This second effect is unique to long-run effort.

In summary, there are two opposing effects distinct to long-run effort. The positive scaling effect increases long-run effort compared to first-best. The negative direct-benefit effect decreases long-run effort compared to first-best. The two effects are dominant for different parameter values and for different levels of the state variable w , leading to either over- or underinvestment in long-run effort. Moreover, the model predicts that a firm may exhibit short-termism when close to distress because the cost to the principal of incentivizing the manager is low. Our results are therefore in contrast with those in Stein (1989) and Shleifer and Vishny (1990), in that short-termism arises as the optimal response to agency conflicts. They also differ from those in models with one-dimensional moral hazard, in which overinvestment in short- or long-term effort never arises. Section 2.5 presents a quantitative analysis of the occurrence of long- and short-termism under agency frictions.

2.2 Pay-for-luck

We next turn to incentive provision. When the level of effort is interior, the incentive-compatibility constraint directly maps effort to incentives since $\beta^s = \lambda_s s(w)$ and $\beta^\ell = \lambda_\ell \ell(w)$. Higher effort means higher incentives and the discussion of effort and incentives cannot be separated. This section therefore focuses on corner levels of effort, i.e. situations in which $s(w) = s_{\max}$ and $\ell(w) = \ell_{\max}$. These cases arise whenever the costs of efforts are low (low $\lambda_s \sigma_X$ or low $\lambda_\ell \sigma_K$) or the maximum effort level is low (low s_{\max} or low ℓ_{\max}). Corner levels of efforts are the only relevant cases in two variants of our model; namely, in a model with only binary effort choices ($s \in \{0, s_{\max}\}$ or $\ell \in \{0, \ell_{\max}\}$) as in e.g. He (2009), or in a model with effort cost functions that are linear in effort levels as in e.g. Biais, Mariotti, Plantin, and Rochet (2007) or DeMarzo, Fishman, He, and Wang (2012).

The objective of the principal when choosing the manager's exposure to firm performance is to maximize the value he derives 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 equation (9)). An application of Itô's formula implies that the dynamics of scaled promised payments are given by:

$$dw = \left[(\gamma + \delta - \mu\ell)w + \sigma_K^2(w - \beta^\ell) + \frac{1}{2} (\lambda_s s^2 \alpha + \lambda_\ell \ell^2 \mu) \right] dt + \beta^s \sigma_X dZ^X + (\beta^\ell - w) \sigma_K dZ^K.$$

Minimizing risk exposure amounts to minimizing the instantaneous variance of the scaled promised payments:

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

This leads to the following result:

Proposition 6 (Pay-for-luck and asymmetric benchmarking). *Assume the usual regularity conditions and that the bounds $(0, i_{\max})$ for $i \in \{s, \ell\}$ are such that the optimal effort levels are at the corner. Then, we have $\beta^s = \lambda_s s_{\max}$ and $\beta^\ell = \max\{\lambda_\ell \ell_{\max}, w\}$ at the optimum.*

The finding that the incentive compatibility constraint $\beta^s \geq \lambda_s s_{\max}$ in Proposition 6 is tight is standard and intuitive. The principal wants 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 $\beta^\ell \geq \lambda_\ell \ell_{\max}$ is not necessarily tight stems from the fact that the principal optimally wants to expose the manager's continuation payoff to long-term, permanent shocks. Indeed, as noted in Hoffmann and Pfeil (2010) and DeMarzo, Fishman, He, and Wang (2012), a positive permanent shock makes liquidation more inefficient. As a result, under the optimal contract, the agent's promised wealth increases in response to a positive shock in order to reduce the likelihood of inefficient liquidation. More precisely, 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 $\beta^\ell dZ^K$. Second, firm size K grows more than expected, thereby reducing the agent's stake in the firm by $-wdZ^K$. If $w > \beta^\ell$, the scaling/dilution effect outweighs the absolute reward due to good performance and the relative stake w decreases following a positive permanent shock, thereby increasing the likelihood of inefficient termination. To eliminate this negative effect, and thus to make w less volatile (which is beneficial because of the concavity of $p(w)$), the investor can increase the sensitivity of future payments to changes in firm size to $\beta^\ell = w > \lambda_\ell \ell(w)$. In such instances, we have $(\beta^\ell - w)\sigma_K dZ^K = 0$ and the effects from scaling/dilution and performance based compensation exactly cancel out. As a result, the agent is incentivized for the long run *without* raising the threat of termination. This requires introducing pay-for-luck in the manager's incentive

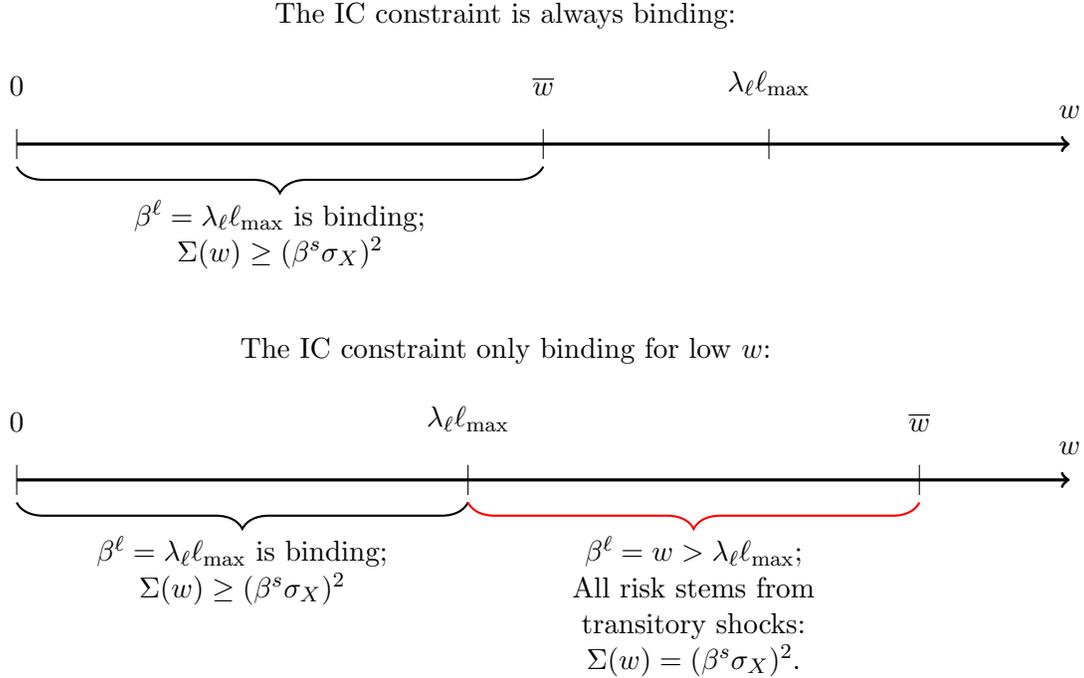


Figure 2: **Binding and non-binding incentive-compatibility constraint for the long-run effort in the case of $\rho = 0$.**

package, i.e. exposure to volatility that is not needed to incentivize long-term effort.

Importantly, Hoffmann and Pfeil (2010) and DeMarzo, Fishman, He, and Wang (2012) also find that exposure to permanent shocks leads to optimal pay-for-luck in managerial compensation. In contrast with these papers, the principal needs to incentivize the manager to exert long-run effort in our framework with multi-tasking. This generates the distinct prediction that pay-for-luck is only present when the agent's stake in the firm is large enough. Notably, when w is low, setting $\beta^\ell = w$ may not suffice to provide incentives. This is why the pay-for-luck effect is shut down and the contract implies instead that $\beta^\ell = \lambda\ell_{\max} > w$. The resulting pattern of pay-for-luck is thus asymmetric in that excessive exposure to permanent shocks is the largest at the highest $w = \bar{w}$ and vanishes as w falls below $\lambda\ell_{\max}$. In other words, positive shocks increase pay-for-luck and negative shocks decrease and eventually eliminate it. This is consistent with evidence on the asymmetry of pay-for-luck in executive compensation (see for example Garvey and Milbourn (2006) and Francis et al. (2013)). In contrast with the suggested explanations, the asymmetry in pay-for-luck is part of an optimal contract and is not due to managerial entrenchment.

Accordingly, two different scenarios can occur in our model. First, when $\lambda_\ell \ell_{\max} > \bar{w}$, incentive-compatibility constraints are always tight and there is no pay-for-luck, as in related models by DeMarzo and Sannikov (2006) or He (2009). This case is depicted in the top panel of Figure 2. Second, a novelty of our model is that \bar{w} can exceed $\lambda_\ell \ell_{\max}$ and thus we can have $\beta^\ell = w > \lambda_\ell \ell_{\max}$ for all $w \in (\lambda_\ell \ell_{\max}, \bar{w}]$. Hence, the incentive-compatibility constraint corresponding to the long-run will be loose in some states and incentive provision will involve pay-for-luck, as illustrated in the bottom panel of Figure 2. Whenever $w > \lambda_\ell \ell_{\max}$, all risk the agent is exposed to stems exclusively from operating cash-flow and incentive provision for the long run becomes effectively costless.

Remarkably, the effects described above critically depends on the presence of both short-run and long-run moral hazard, as we show in the following Proposition:

Proposition 7. *Assume the usual regularity conditions. If long-run effort is observable, long-run incentives satisfy $\beta^\ell = w$ for all $w \in [0, \bar{w}]$ and the manager is always rewarded for long-term luck shocks. If short-run effort is observable, long-run incentives satisfy $\beta^\ell > w$ for all $w \in [0, \bar{w}]$ and the manager is never rewarded for long-term luck shocks.*

Section 2.4 goes back to this issue by providing a detailed analysis of the model with only short- or long-run moral hazard. To close this section note that our findings differ from those in He (2009). In contrast to his work, 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 $\gamma > r$. In He (2009), this can only happen in the extreme case of equally patient agent and principal, where the resulting agency conflict does not affect firm value. This implies that the firm eventually becomes riskless and the agent works forever. As a result, the first-best outcome can be achieved. By contrast, when the firm is exposed to both permanent and transitory shocks, *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.

2.3 Correlated Short- and Long-run Shocks

Assume now that the short- and long-run shocks are correlated. In such instances, optimal short- and long-run efforts, if interior, are given by

$$s(w) = \frac{\alpha + \overbrace{p''(w)\rho\sigma_X\sigma_K\lambda_s}^{\text{Correlation effect}}(\lambda_\ell\ell(w) - w)}{-p'(w)\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2} \quad (13)$$

and

$$\ell(w) = \frac{\mu(p(w) - p'(w)w) + p''(w)\overbrace{(\rho\sigma_X\sigma_K\lambda_\ell\lambda_s s(w))}^{\text{Correlation effect}} - w\lambda_\ell\sigma_K^2}{-p'(w)\lambda_\ell\mu - p''(w)(\lambda_\ell\sigma_K)^2}. \quad (14)$$

Compared to equations (11) and (12), new terms appear that affect both optimal effort levels and incentives. With non-zero correlation, the agent's exposure $\beta^s = \lambda_s s(w)$ to transitory cash flow shocks poses an externality on the choice of $\beta^\ell = \lambda_\ell \ell(w)$. Intuitively, when the two sources of risk are positively correlated, exposing the manager's continuation payoff W to both transitory and permanent shocks creates excess volatility and is therefore costly. As a result, leaving β^s unchanged, a higher correlation between short- and long-run shocks ρ makes it optimal for the principal to decrease β^ℓ —and thus the volatility of w —to limit the risk of inefficient liquidation. With interior effort levels, the sensitivities β^s and β^ℓ directly relate to $s(w)$ and $\ell(w)$ and there are therefore first-order externalities of short-run effort on long-run effort and vice versa. This externality of $s(w)$ on $\ell(w)$ is negative (positive) if $\rho > 0$ ($\rho < 0$) and is captured in equation (14) by

$$p''(w)\rho\sigma_X\sigma_K\lambda_\ell\beta^s = p''(w)\rho\sigma_X\sigma_K\lambda_\ell\lambda_s s(w).$$

The magnitude of the externality scales with the curvature of the value function $p''(w)$ and is therefore relatively weaker once the firm has accumulated sufficient financial slack (i.e. once w is sufficiently high). In other words, optimal short- and long-run efforts are essentially interrelated only when w is low and the likelihood of inefficient liquidation is high.

Equation (13) demonstrates that the choice of long-run effort $\ell(w)$ also feeds back in the choice of short-run effort $s(w)$. However, the correlation effect in the numerator of $s(w)$ in

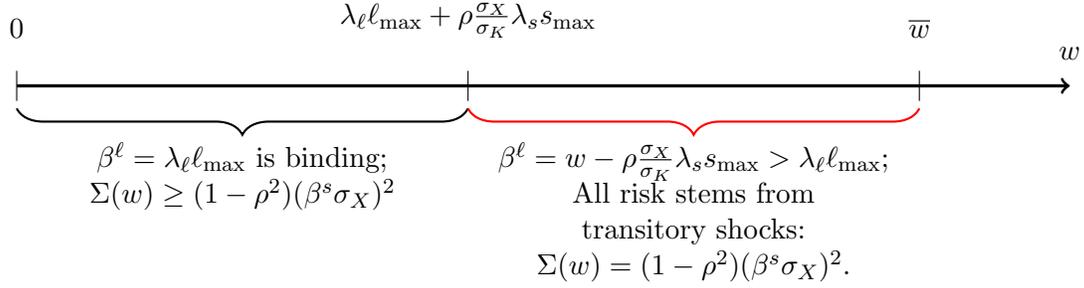


Figure 3: **Non-binding incentive-compatibility constraint for the long-run effort in the case of $\rho \neq 0$.**

equation (13) has an additional term and therefore consists of two separate components:

$$p''(w)\rho\sigma_X\sigma_K\lambda_s(\lambda_\ell\ell(w) - w) = \underbrace{p''(w)\rho\sigma_X\sigma_K\lambda_s\lambda_\ell\ell(w)}_{\text{Externality}} - \underbrace{p''(w)\rho\sigma_X\sigma_K\lambda_s w}_{\text{Scaling}}.$$

This decomposition shows that when the correlation between shocks is non-zero, 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 discussed in section 2.1, the principal counteracts this scaling/dilution effect by tying the manager's compensation to permanent shocks when there is no correlation. When the two sources of cash-flow risk are correlated, this can be done by changing either long-run or short-run incentives. This leads to lower short-run effort and incentives when $\rho < 0$ and to higher short-run effort and incentives when $\rho > 0$. The overall pattern is thus that negative ρ promotes long-run effort at the expense of short-run effort and positive ρ promotes short-run effort at the expense of long-run effort.

Another important change brought about by non-zero correlation relates to pay-for-luck when long-run effort is in the corner. Under the conditions stated in Appendix D, the incentive compatibility condition with respect to long-run incentives is non-binding and the agent is rewarded for luck when w exceeds $\lambda_\ell\ell_{\max} + \rho\frac{\sigma_X}{\sigma_K}\lambda_s s(w)$, in which case all long-run risk as well the portion of short-run risk correlated with the long-run risk are removed from the continuation utility. Figure 3 illustrates this case. Non-zero correlation has two effects. First, it allows to remove the portion of short-run risk correlated with the long-run risk. As expected, this is independent of the sign of the correlation coefficient. Second, the level of w that allows for a non-binding incentive compatibility condition (i.e. pay-for-luck) changes

with ρ . Negative ρ decreases this level compared to the case of zero correlation and positive ρ increases it. Hence, our model predicts that CEO's in industries in which transitory and permanent cash-flow shocks are positively correlated are less frequently rewarded for luck, in that their compensation is more closely tied to actual performance.

2.4 Effort and Incentives when Moral Hazard is One-dimensional

To get a better understanding of how agency conflicts interact to determine optimal incentive provision and effort choices, it is instructive to see how the sensitivities to short- and long-run performance β^s and β^ℓ as well as the effort levels $s(w)$ and $\ell(w)$ are optimally chosen when moral hazard only affects either short-run effort *or* long-run effort.

Assume first that moral hazard is only over the short-run, in that $\{\ell\} = \{\ell_t\}_{t \geq 0}$ is perfectly observable to the principal (equivalently, the Brownian motion $\{Z_t^K\}_{t \geq 0}$ is observable to the principal). Appendix F shows that in this case both short- and long-run effort choices are distorted due to moral hazard. Notably, because the benefits but not the cost of long-run investment depend on (scaled) firm value, the firm underinvests in long-run effort in that

$$\ell(w) = \frac{1}{\lambda_\ell} [p(w) - wp'(w)] \leq \ell^{FB},$$

for all $t \geq 0$, where the inequality is strict if $\ell^{FB} < \ell_{\max}$. In addition, if interior, optimal short-run effort and incentives are given by:

$$s(w) = \frac{\alpha}{-p'(w)\lambda_s\alpha - p''(w)(\lambda_s\sigma_X)^2(1 - \rho^2)},$$

and $\beta^s(w) = \lambda_s s(w)$. The optimal exposure to long-run shock in turn satisfies

$$\beta^\ell(w) = w - \underbrace{\rho \frac{\sigma_X}{\sigma_K}}_{\text{Externality}} \beta^s(w). \tag{15}$$

Equation (15) shows that even though there is no agency conflict over the long run, it is optimal to expose the agent to permanent cash-flow shocks that are beyond her influence, such that $\beta^\ell \neq 0$ and the agent is paid for luck as in Hoffmann and Pfeil (2010) and DeMarzo et al. (2012). Notably, the extent of pay-for-luck, i.e. β^ℓ , depends not only on the agent's stake w but also on the correlation between transitory and permanent shocks as well as the level

of short-run incentives. As shown in section 2.2, this result heavily depends on the absence of long-run moral hazard. In the presence of both moral hazard problems, the principal will only reward the agent for luck—if at all—after sufficiently strong past performance. That is, the optimal contract involves pay-for-luck in an asymmetric manner and implements it after good but not poor performance.

Whenever correlation is non-zero, making W contingent on short-run performance via $\beta^s > 0$ poses an externality on the choice of β^ℓ . Intuitively, if $\rho > 0$, exposing the agent to both sources of risk implies excess volatility and is particularly costly. This makes it optimal—leaving short-run incentives β^s unchanged—to decrease β^ℓ and therefore to pay the agent less for luck. By optimally choosing the sensitivity β^ℓ , it is possible to reduce the squared volatility of the agent’s continuation payoff w down to $\Sigma_s^* \equiv (\beta^s \sigma_X)^2 (1 - \rho^2)$.¹⁴

Next, suppose short-run effort is observable but long-run effort is not. That is, the moral hazard problem is only over the long run. In this case, short-run effort s is always at the first-best level $s^{FB} = \frac{1}{\lambda_s}$, while long-run effort $\ell(w)$ is distorted through the (usual) agency frictions. Notably, it is straightforward to establish that

$$\ell(w) = \frac{\mu(p(w) - p'(w)w) - p''(w)w\lambda_\ell\sigma_K^2(1 - \rho^2)}{-p'(w)\lambda_\ell\mu - p''(w)(\lambda_\ell\sigma_K)^2(1 - \rho^2)}$$

and

$$\beta^s(w) = \rho \frac{\sigma_K}{\sigma_X} (w - \beta^\ell).$$

As with observable long-run effort, the investor exposes the manager to transitory cash flow shocks beyond her influence and thereby decreases the squared volatility of w down to $\Sigma_\ell^0 \equiv (1 - \rho^2)\sigma_K^2(\beta^\ell - w)^2$. Importantly, pay-for-luck over the short run is proportional to ρ and therefore vanishes with zero correlation. This arises because through exposing the agent to transitory shocks the principal attempts to decrease overall risk. This is only possible when current cash-flow is also informative about future performance, in that $\rho \neq 0$. This result again illustrates the central role of the correlation between short- and long-term shocks on incentive provision and effort choice. If correlation were zero, it would be suboptimal to make the agent’s payments contingent on these shocks, in accordance with Holmstrom (1979).

¹⁴The model with observable permanent shocks $\{Z^K\}$ and parameters $(\sigma_X, \sigma_K, \rho)$ is isomorphic to a model with no Brownian permanent shocks and corresponding parameter vector given by $(\sigma_X \sqrt{1 - \rho^2}, 0, 0)$.

This is the case in our model of short-term cash flow shocks that are uncorrelated with the permanent shocks to firm size, i.e. of shocks that do not affect long-term prospects.¹⁵

2.5 Quantitative Analysis

2.5.1 Short-termism and Long-termism

In sections 2.1 and 2.3, we showed that both short-run and long-run efforts can exceed the first-best levels leading to short- and long-termism. Table 1 presents optimal short- and long-run effort choices for various levels of short-run risk σ_X , long-run risk σ_K , correlation between short- and long-run shocks ρ , and promised payments to the manager (likelihood of termination) w . The first-best effort choices are constant along these four dimensions and parameter values are such that the first-best levels of effort satisfy $s^{FB} = \ell^{FB} = 1$. Thus $s(w) > 1$ represents optimal short-termism and $\ell(w) > 1$ represents optimal long-termism.

In our quantitative analysis, we use the following input parameter values. We set the discount rate parameters to $r = 4.6\%$ and $\gamma = 5\%$ and the depreciation rate to $\delta = 12.5\%$, in line with DeMarzo et al. (2012). For convenience, we normalize first best effort levels to one, i.e. $s^{FB} = \ell^{FB} = 1$, and set $\mu = 13.2\%$ so that first best investment $\mu\ell^{FB}$ is as in Bolton et al. (2017). This implies the cost parameter values $\lambda_s = 1$ and $\lambda_\ell = 1.42$. The volatility parameter of the long term shock is set to $\sigma_K = 20\%$, in line with Bolton et al. (2017), 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 $\alpha = 30\%$, implying that the (expected) return on assets is 5.6% at first best, in line with the estimates in Morellec, Nikolov, and Schürhoff (2012) and Graham, Leary, and Roberts (2015).¹⁶ Finally, we assume the recovery value $R = 20\%$. In our analysis, we also investigate the effects of varying these parameters around their base case values.

Panel A of Table 1 shows that short-termism arises when σ_X is low, σ_K is high, ρ is positive, and w is close to zero. Note that σ_X and σ_K can be interpreted as costs of short- and long-run efforts to the principal, since high volatilities make it difficult for the principal

¹⁵The model with observable transitory shocks $\{Z^X\}$ and parameters $(\sigma_K, \sigma_X, \rho)$ is isomorphic to a model with no Brownian transitory shocks and corresponding parameter vector given by $(\sigma_K\sqrt{1-\rho^2}, 0, 0)$.

¹⁶One important difference between our model and prior work is that effort is generally not at first best and varies between 0 and i_{\max} for $i = s, \ell$. This is why we base our calibration of the return on assets and of the growth rate of capital on the first best case.

Table 1: Optimal short- and long-run effort levels. The constant parameter values are such that the first-best levels of s and ℓ are always equal to 1. The parameters are $\alpha = 0.3$, $\mu = 0.132$, $\sigma_X = 0.25$, $\sigma_K = 0.2$, $\lambda_s = 1$, $\lambda_\ell = 1.42$, $r = 0.046$, $\gamma = 0.05$, $s_{\max} = \ell_{\max} = 2$, and $R = 0.2$.

Panel A: Short-run effort, $s(w)$											
		$w = 0$					$w = w^*$				
		$\sigma_K =$					$\sigma_K =$				
		0.05	0.10	0.15	0.20	0.25	0.05	0.10	0.15	0.20	0.25
$\sigma_X = 0.15$	$\rho = -0.5$	0	0	0	0	0.001	0.793	0.861	0.952	0.99	1.004
	$\rho = 0$	0.223	0.227	0.228	0.229	0.229	0.811	0.901	1.003	1.038	1.049
	$\rho = 0.5$	0.718	0.886	1.40	2	2	1.017	1.145	1.219	1.221	1.234
$\sigma_X = 0.2$	$\rho = -0.5$	0	0	0	0	0	0.644	0.715	0.805	0.841	0.852
	$\rho = 0$	0.22	0.22	0.227	0.228	0.228	0.644	0.729	0.833	0.876	0.893
	$\rho = 0.5$	0.568	0.686	1.01	1.406	1.889	0.757	0.91	1.069	1.111	1.136
$\sigma_X = 0.25$	$\rho = -0.5$	0	0	0	0	0	0.603	0.619	0.691	0.722	0.739
	$\rho = 0$	0.22	0.221	0.226	0.227	0.228	0.517	0.605	0.704	0.751	0.774
	$\rho = 0.5$	0.260	0.585	0.829	1.112	1.442	0.588	0.751	0.923	0.988	1.028
$\sigma_X = 0.30$	$\rho = -0.5$	0	0	0	0	0	0.429	0.521	0.595	0.632	0.646
	$\rho = 0$	0.22	0.221	0.223	0.227	0.227	0.429	0.507	0.603	0.650	0.670
	$\rho = 0.5$	0.178	0.499	0.728	0.951	1.203	0.463	0.636	0.812	0.884	0.932

Panel B: Long-run effort, $\ell(w)$											
		$w = 0$					$w = w^*$				
		$\sigma_K =$					$\sigma_K =$				
		0.05	0.10	0.15	0.20	0.25	0.05	0.10	0.15	0.20	0.25
$\sigma_X = 0.15$	$\rho = -0.5$	0	0	0	0	0	1.023	0.847	0.585	0.448	0.364
	$\rho = 0$	0.054	0.011	0.003	0.001	0	0.840	0.628	0.376	0.275	0.223
	$\rho = 0.5$	0	0	0	0	0	0.407	0.143	0	0	0
$\sigma_X = 0.20$	$\rho = -0.5$	0	0	0	0	0	1.206	1.015	0.716	0.548	0.442
	$\rho = 0$	0.01	0.009	0.009	0.004	0.002	1.082	0.823	0.500	0.368	0.299
	$\rho = 0.5$	0	0	0	0	0	0.722	0.356	0.041	0	0
$\sigma_X = 0.25$	$\rho = -0.5$	0.001	0	0	0	0	1.25	1.135	0.813	0.622	0.509
	$\rho = 0$	0.051	0.043	0.023	0.009	0.005	1.25	0.968	0.594	0.442	0.365
	$\rho = 0.5$	0	0	0	0	0	1.003	0.531	0.130	0.035	0.018
$\sigma_X = 0.30$	$\rho = -0.5$	0.01	0.001	0	0	0	1.25	1.226	0.878	0.683	0.558
	$\rho = 0$	0.142	0.121	0.054	0.017	0.009	1.25	1.085	0.660	0.492	0.403
	$\rho = 0.5$	0	0	0	0	0	1.212	0.655	0.179	0.057	0.035

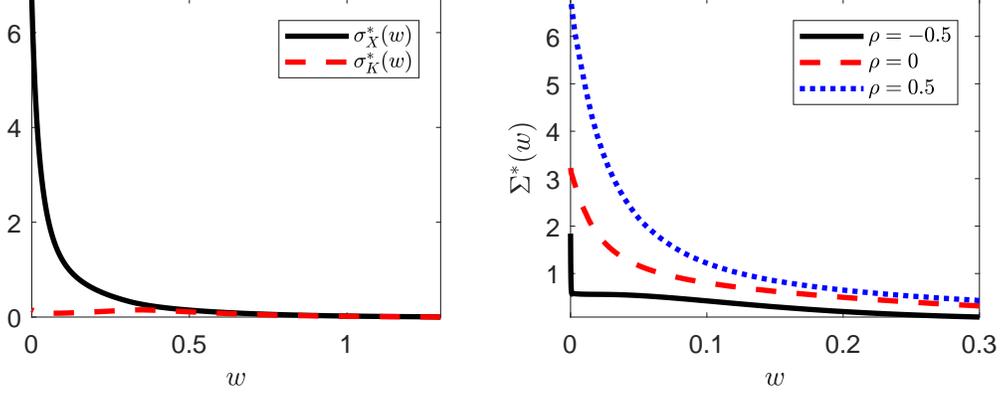


Figure 4: Return volatility components $\Sigma_X^*(w)$ and $\Sigma_K^*(w)$ for $\rho = 0$ (left panel) and aggregate volatility $\Sigma_P^*(w)$ for different values of ρ (right panel) in our base case environment. The parameters are $\alpha = 0.3$, $\mu = 0.132$, $\sigma_X = 0.25$, $\sigma_K = 0.2$, $\lambda_s = 1$, $\lambda_\ell = 1.42$, $r = 0.046$, $\gamma = 0.05$, $s_{\max} = \ell_{\max} = 2$, and $R = 0.2$.

to infer the manager's actions from the realizations of X and K . Therefore, consistent with economic intuition, the principal puts more emphasis on short-run effort when it is relatively less costly to incentivize short-term actions. Optimal short-termism also arises when the firm is close to termination and short- and long-run shocks are positively correlated.¹⁷ As discussed earlier, a positive correlation between shocks increases the risk of liquidation. When the likelihood of liquidation is high, the benefits of long-term growth are limited. By contrast, stimulating short-term effort increases the cash flow rate of the firm and therefore the continuation utility of the manager w , thereby reducing the risk of termination. Short-termism is therefore optimal in such instances. The above analysis implies that in our model short-termism arises exactly when the principal's return on the firm is very volatile. The right panel of Figure 4 plots the volatility of the return on the principal's stake $\Sigma_P^*(w)$ (derived in Appendix G) as a function of the manager's continuation payoff w for three different values of the correlation coefficient ρ . The left panel decomposes this volatility between transitory volatility $\Sigma_X^*(w)$ and permanent volatility $\Sigma_K^*(w)$ when $\rho = 0$. Because the principal's stake can be seen as the firm's equity, our results are consistent with the evidence in Brochet, Loumioti, and Serafeim (2012) and Brochet, Loumioti, and Serafeim (2013), who show that corporate short termism is associated with higher stock-return volatility.

¹⁷We say that a firm is 'close to termination' when w is sufficiently close to zero. In order to assess firm policies 'close to termination', we analyze effort choice 'just before' termination, given by $\lim_{w \downarrow 0} k(w)$ for $k \in \{s, \ell\}$. For convenience, we informally write $k(0) \equiv \lim_{w \downarrow 0} k(w)$, even though effort at $w = 0$ itself equals zero, in that in general $0 = k(0) \neq \lim_{w \downarrow 0} k(w)$ and $k(\cdot)$ is not right-continuous at $w = 0$.

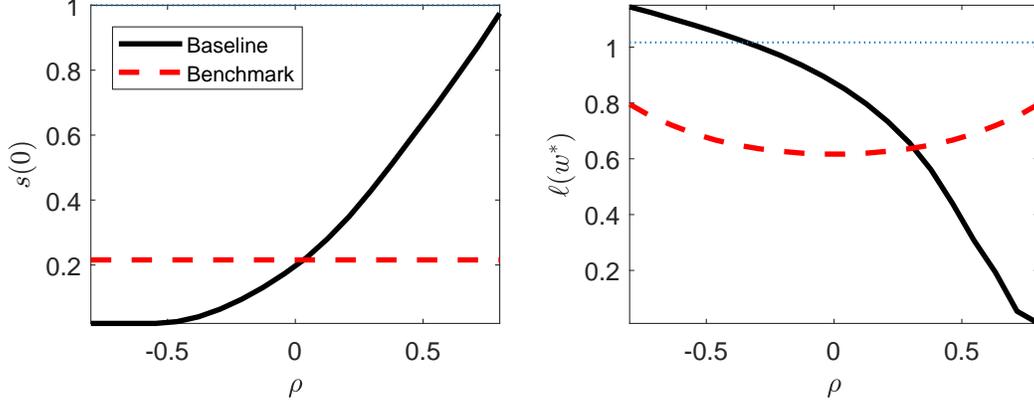


Figure 5: Low σ_K ($= 0.1$) with $s^{FB} = \ell^{FB} = 1$. The parameters are $\alpha = 0.3$, $\mu = 0.132$, $\sigma_X = 0.25$, $\sigma_K = 0.1$, $\lambda_s = 1$, $\lambda_\ell = 1.42$, $r = 0.046$, $\gamma = 0.05$, $s_{\max} = \ell_{\max} = 2$, and $R = 0.2$.

Panel B of Table 1 reports optimal long-run effort. Consistent with economic intuition, the principal puts more emphasis on long-run effort when it is less costly to incentivize (low σ_K) and short-run effort is more costly to incentivize (high σ_X). The table also shows that optimal long-termism primarily arises when ρ is negative (in which case long- and short-run efforts are complements) and the manager's continuation payoff is close to w^* solving equation (10). By the preceding discussion, our results suggest that excessive long-termism is more likely to arise in firms with low stock return volatility.

Next, we compare the optimal investment in efforts to another important benchmark in which long-run effort $\{\ell\}$ is publicly observable. This benchmark, analyzed in section 2.4, amounts to a version of the model of DeMarzo and Sannikov (2006) augmented with observable permanent shocks. The question we ask here is the following: What do agency conflicts over the long run bring to the standard dynamic short-run moral hazard model?

Figures 5 and 6 present short- and long-run effort levels in our baseline model, the benchmark model with observable long-run effort, and at first-best. We focus on the levels of the continuation payoff w at which over-investment is most likely, that is $w = 0$ when analyzing short-run effort $s(w)$ and w^* when analyzing long-run effort $\ell(w)$. For negative ρ and low σ_K , long-run investment tends to be higher in the baseline model compared to the benchmark, as shown in Figure 5. In contrast, short-run effort is generally higher than in the benchmark, whenever $\rho > 0$ and σ_K is large (see Figure 6). Importantly, in the benchmark model both short- and long-run investment tend to be below first-best, while this need not be the case in our model with dual moral hazard. That is, our more richer model yields the new

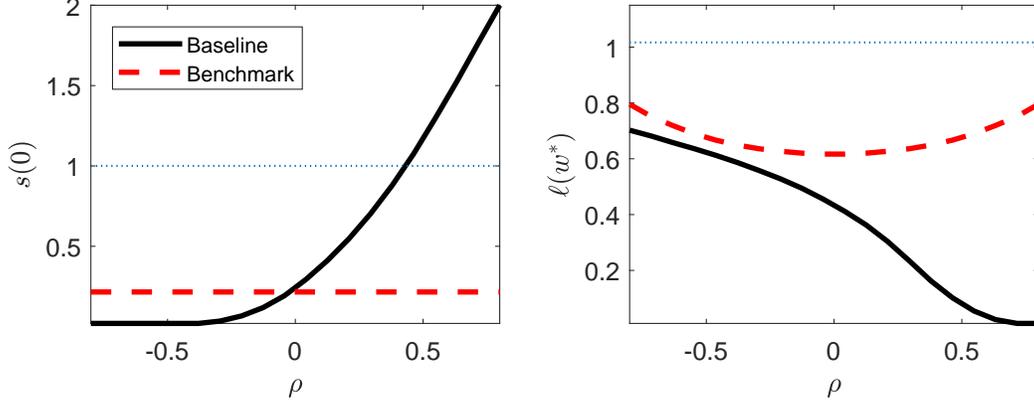


Figure 6: High σ_K ($= 0.2$) with $s^{FB} = \ell^{FB} = 1$. The parameters are $\alpha = 0.3$, $\mu = 0.132$, $\sigma_X = 0.25$, $\sigma_K = 0.2$, $\lambda_s = 1$, $\lambda_\ell = 1.42$, $r = 0.046$, $\gamma = 0.05$, $s_{\max} = \ell_{\max} = 2$, and $R = 0.2$.

insight that agency issues may lead to higher managerial effort and to optimal short-termism or long-termism.

2.5.2 Are Short- and Long-run Efforts Substitutes or Complements?

A common problem in firms is the trade-off between long-term growth and short-term profits. This issue is related to the critique of corporate short-termism raises a number of questions. For example, when do firms need to sacrifice short-term profits to stimulate long-term growth? When is profit-orientation compatible with high growth? Our model presents a new perspective on these questions based on long- and short-term agency conflicts. In terms of our model, these questions amount to determining whether short- and long-run efforts are substitutes or complements in the firm's production function.

The analysis in Sections 2.1 and 2.3 identified different forces that move short- and long-run effort $s(w)$ and $\ell(w)$ in the same as well as opposing directions. On the one hand, short- and long-run effort pose mutual risk externalities on each other, which are positive if and only if $\rho < 0$. On the other hand, it seems natural from an optimal multitasking perspective, that an increased focus on the short-run comes at the expense of the long-run and vice versa. According to our model, this is because providing incentives for both short- and long-term goals exposes the firm to a lot of risk and therefore reducing long-run effort ℓ 'frees up' some capacity to raise s .

Whereas our previous results unambiguously predict that s and ℓ are substitutes whenever correlation is non-negative, we do not have yet a sharp prediction for the case $\rho < 0$.

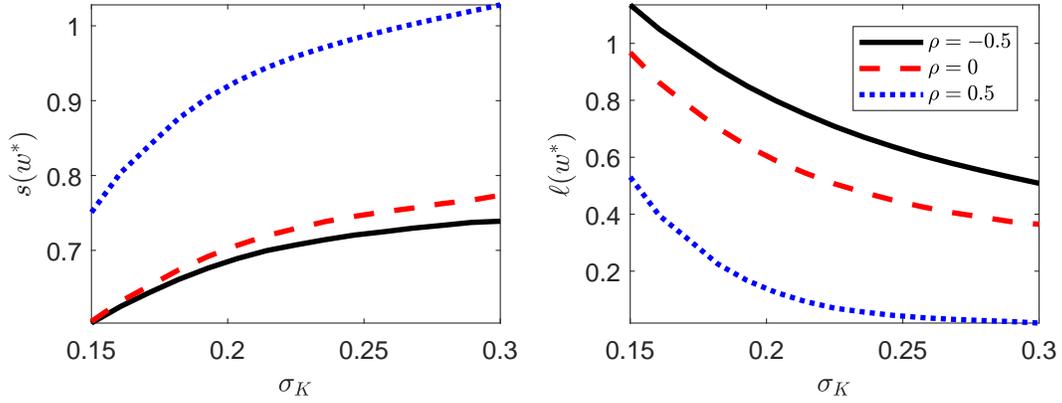


Figure 7: Short-run effort increases in σ_K .

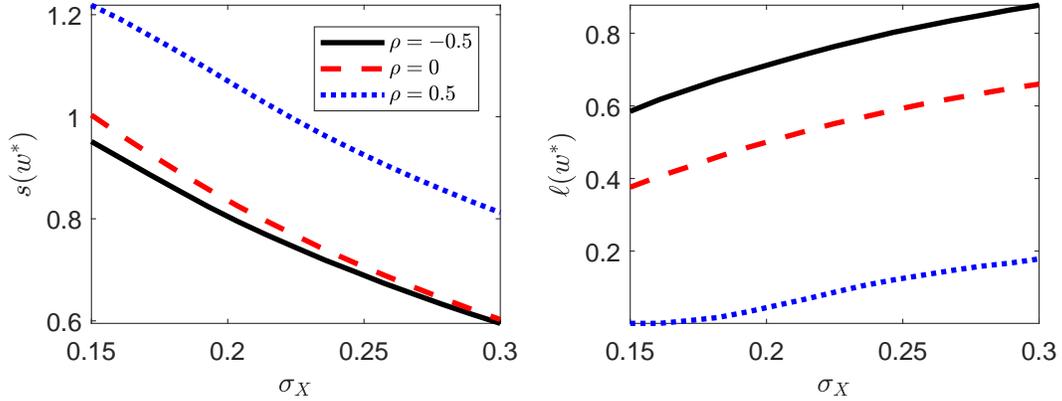


Figure 8: Long-run effort increases in σ_X

Therefore, Figures 7 and 8 present a numerical analysis for our benchmark parameters and varying levels of ρ at $w = w^*$.

Increasing the cost of long-run (short-run) effort through raising σ_K (σ_X) directly decreases $\ell(w)$ and increases $s(w)$, regardless of whether correlation is positive or negative. This means that the two inputs ℓ and s are substitutes. Hence, even if permanent and transitory shocks are negatively correlated, it remains optimal for a firm to sacrifice current cash-flow in order to grow faster, whenever growth opportunities improve. In that regard, the positive external effect between effort choices (when $\rho < 0$) is crowded out by multi-tasking considerations and therefore quantitatively of second order, at least in our baseline calibration.

3 Conclusion

We develop a continuous-time agency model in which the agent controls current earnings via short-term effort and firm growth via long-term effort. In this multi-tasking model, the principal optimally balances the costs and benefits of incentivizing the manager over the short- or the long-term, leading to both overinvestment, i.e. effort above the first-best level, and underinvestment, i.e. effort below the first-best level. As shown in the paper, this leads to optimal short-termism when the likelihood of inefficient termination is high and to optimal long-termism when it is low. The model additionally predicts that optimal short-termism is more likely to arise when the volatility of the firm value process is high and the correlation between shocks to earnings and firm value is positive, i.e. when stock return volatility is higher in line with the available evidence. By contrast, optimal long-termism is more likely to arise when the volatility of the firm value process is low and the correlation between short- and long-term shocks is negative, i.e. when stock return volatility is lower.

Incentives are provided in the optimal contract by making the agent's compensation contingent on firm performance, in particular by rewarding (punishing) the agent when outcomes are better (worse) than expected. Because the firm is subject to long-run, permanent shocks, it is optimal to introduce pay-for-luck in incentive compensation, i.e. exposure to volatility that is not needed to incentivize effort. In our model with multi-tasking, however, the principal needs to incentivize the manager to exert long-run effort. This generates the distinct prediction that pay-for-luck is only present when the agent's stake in the firm is high enough. Notably, we show that when the continuation payoff of the manager is low, pay-for-luck does not suffice to provide incentives to the manager so that the pay-for-luck effect is shut down. When the continuation payoff is high enough, positive shocks increase pay-for-luck and negative shocks decrease and eventually eliminate it. Our model therefore provides a rationale for the asymmetry of pay-for-luck observed in the executive compensation data. Lastly, we show that the correlation between short- and long-run shocks is a key determinant of incentive provision, effort choice, and firm performance. Notably, correlated short- and long-term shocks to cash flows and firm value lead to externalities in effort choices and incentive provision with higher correlation leading to decreased long-term effort and incentives.

Appendix

Without loss in generality, we consider throughout the whole Appendix that the depreciation rate of capital δ equals zero.

A Proof of Proposition 2

A.1 Auxiliary Results

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

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

with correlated standard Brownian motions $\{\widetilde{W}^X\}, \{\widetilde{W}^K\}$ under this measure, both progressive with respect to \mathbb{F} . The measure \mathcal{P}^0 corresponds to perpetual zero effort. Define $\widetilde{\mathbf{W}}_t \equiv (\widetilde{W}_t^X, \widetilde{W}_t^K)'$ ¹⁸ and let the (unconditional) covariance matrix of $\widetilde{\mathbf{W}}_t$ under \mathcal{P}^0 be

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

Here, $\mathbb{V}^0(\cdot)$ 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

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

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

$$\Gamma_t = \Gamma_t(s, \ell) \equiv \exp \left(\int_0^t \tilde{\Theta}_u \cdot d\widetilde{\mathbf{W}}_u^T - \frac{1}{2} \int_0^t \|\tilde{\Theta}_u\|^2 du \right),$$

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

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

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

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

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

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 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

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

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

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

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.

A.2 Proof of Proposition 2.1

Proof. Consider an incentive compatible contract $\Pi \equiv (\{C\}, \{\hat{s}\}, \{\hat{\ell}\}, \tau)$. 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

$$W_t(\Pi) \equiv \mathbb{E}_t^{\hat{s}, \hat{\ell}} \left[\int_t^\tau e^{-\gamma(z-t)} dC_z - \frac{1}{2} \int_t^\tau e^{-\gamma(z-t)} K_z (\lambda_s \alpha \hat{s}_z^2 + \lambda_\ell \mu \hat{\ell}_z^2) dz \right],$$

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

$$\Gamma_t(\Pi) \equiv \mathbb{E}_t^{\hat{s}, \hat{\ell}} [W_0(\Pi)] = \int_0^t e^{-\gamma z} dC_z - \frac{1}{2} \int_0^t [e^{-\gamma z} K_z (\lambda_s \alpha \hat{s}_z^2 + \lambda_\ell \mu \hat{\ell}_z^2)] dz + e^{-\gamma t} W_t(\Pi). \quad (\text{A1})$$

By construction, $\{\Gamma_t(\Pi) : 0 \leq t \leq \tau\}$ is a square-integrable martingale under $\mathcal{P}^{\hat{s}, \hat{\ell}}$, progressive with respect to \mathbb{F} .

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

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

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

$$dZ_t^1 \equiv \frac{dX_t - K_t \alpha \hat{s}_t dt}{K_t \sigma_X} \quad \text{and} \quad dZ_t^2 \equiv \frac{dK_t - K_t \mu \hat{\ell}_t dt}{K_t \sigma_K}$$

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

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

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 . Differentiating both sides of (A1) with respect to t , one can verify that

$$\begin{aligned} d\Gamma_t(\Pi) &= e^{-\gamma t} K_t (\beta_t^s \sigma_X dZ_t^1 + \beta_t^\ell \sigma_K dZ_t^2) \\ &= e^{-\gamma t} \left[dC_t - \frac{K_t}{2} (\lambda_s \alpha \hat{s}_t^2 + \lambda_\ell \mu \hat{\ell}_t^2) dt \right] - \gamma e^{-\gamma t} W_t(\Pi) dt + e^{-\gamma t} dW_t(\Pi) \end{aligned}$$

and thus equation (5) holds after rearranging. Indeed, since the right hand side of (5) satisfies a Lipschitz-condition under the usual regularity conditions (i.e. square integrability of $\{C\}$ and $\{\hat{s}\}, \{\hat{\ell}\}$ of bounded variation), $\{W\}$ is the unique strong solution to the stochastic differential equation (5).

Next, we provide necessary and sufficient conditions for the contract Π to be indeed incentive compatible. For this purpose, let the 'recommended' effort processes $\{\hat{s}\}$ and $\{\hat{\ell}\}$ and the expected payoff of the agent at time t be W_t , when following the recommended strategy from time t onwards. Further, let $\{s\}$ and $\{\ell\}$ represent the 'actual' effort processes, which may in principle differ from $\{\hat{s}\}$ and $\{\hat{\ell}\}$. We have

$$W_t \equiv \mathbb{E}_t^{\hat{s}, \hat{\ell}} \left[\int_t^\tau e^{-\gamma(z-t)} dC_z - \frac{1}{2} \int_t^\tau e^{-\gamma(z-t)} K_z (\lambda_s \alpha \hat{s}_z^2 + \lambda_\ell \mu \hat{\ell}_z^2) dz \right].$$

Recall that $\mathbb{E}_t^{\hat{s}, \hat{\ell}}$ denotes the expectation, conditional on the public (information) filtration

\mathcal{F}_t , taken under the probability measure $\mathcal{P}^{s, \hat{\ell}}$. As shown above, the process $\{W\}$ solves the stochastic differential equation:

$$dW_t = \gamma W_t dt + \frac{1}{2} K_t (\lambda_s \alpha \hat{s}_t^2 + \lambda_\ell \mu \hat{\ell}_t^2) dt + \beta_t^s (dX_t - K_t \alpha \hat{s}_t dt) + \beta_t^\ell (dK_t - K_t \mu \hat{\ell}_t dt) - dC_t.$$

We can rewrite this stochastic differential equation as

$$dW_t + dC_t = \gamma W_t dt + \frac{1}{2} K_t (\lambda_s \alpha \hat{s}_t^2 + \lambda_\ell \mu \hat{\ell}_t^2) dt + K_t \beta_t^s [\alpha (s_t - \hat{s}_t) dt + \sigma_X dZ_t^X] + K_t \beta_t^\ell [\mu (\ell_t - \hat{\ell}_t) dt + \sigma_K dZ_t^K]$$

with

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

Girsanov's theorem implies now that $dZ_t^X \equiv \frac{dX_t - K_t \alpha s_t dt}{K_t \sigma_X}$ 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 the measure $\mathcal{P}^{s, \ell}$ induced by $(\{s\}, \{\ell\})$.

Next, define the auxiliary gain process

$$g_t = g_t(\{s\}, \{\ell\}) \equiv \int_0^t e^{-\gamma z} dC_z - \frac{1}{2} \int_0^t e^{-\gamma z} K_z (\alpha \lambda_s s_z^2 + \mu \lambda_\ell \ell_z^2) dz + e^{-\gamma t} W_t$$

and note that the agent's 'actual' expected payoff under the strategy $(\{s\}, \{\ell\})$ reads

$$W'_0 \equiv \max_{\{s\}, \{\ell\}} \mathbb{E}^{s, \ell} \left[\int_0^\tau e^{-\gamma z} dC_z - \frac{1}{2} \int_0^\tau e^{-\gamma z} K_z (\alpha \lambda_s s_z^2 + \mu \lambda_\ell \ell_z^2) dz \right] = \max_{\{s\}, \{\ell\}} \mathbb{E}^{s, \ell} [g_\tau(\{s\}, \{\ell\})].$$

Using Ito's Lemma, we obtain

$$\begin{aligned} e^{\gamma t} dg_t &= \frac{K_t}{2} [\alpha \lambda_s (\hat{s}_t^2 - s_t^2) + \mu \lambda_\ell (\hat{\ell}_t^2 - \ell_t^2)] dt \\ &\quad + K_t [\alpha \beta_t^s (s_t - \hat{s}_t) + \mu \beta_t^\ell (\ell_t - \hat{\ell}_t)] dt + K_t [\beta_t^s \sigma_X dZ_t^X + \beta_t^\ell \sigma_K dZ_t^K] \\ &\equiv \mu_t^g dt + K_t [\beta_t^s \sigma_X dZ_t^X + \beta_t^\ell \sigma_K dZ_t^K]. \end{aligned}$$

It is now easy to see that, when choosing $s_t = \hat{s}_t, \ell_t = \hat{\ell}_t$, 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,

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

The inequality is strict if and only if there exist processes $\{s\}, \{\ell\}$ 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 exist $\varepsilon, \delta > 0$ such that

$$\bar{\mathcal{P}} \equiv \mathcal{P}^{s, \ell}(\mu_z(g) > \varepsilon \text{ for all } z \in \mathcal{I} = (\tau', \tau' + \delta)) > 0$$

and therefore

$$W'_0 \geq \mathbb{E}^{s,\ell} \left[\int_{\mathcal{I}} e^{-\gamma z} \mu_z^g dz \right] + W_0 \geq \bar{\mathcal{P}} \mathcal{P}^{s,\ell}(\tau' < \tau) \delta \varepsilon \mathbb{E}^{s,\ell} e^{-\gamma \tau} + W_0 > W_0.$$

Observe that the last inequality utilizes the equivalence of the two probability measures, such that $\mathcal{P}^{\hat{s},\hat{\ell}}(\tau = \infty) = \mathcal{P}^{s,\ell}(\tau = \infty) = 0$ and thus $\mathbb{E}^{s,\ell} e^{-\gamma \tau} > 0$. In case $W'_0 > W_0$, either $s_z \neq \hat{s}_z$ or $\ell_z \neq \hat{\ell}_z$ on the interval \mathcal{I} , 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

$$\max_{s_t, \ell_t} \left\{ \alpha \beta_t^s (s_t - \hat{s}_t) + \frac{1}{2} \alpha \lambda_s (\hat{s}_t^2 - s_t^2) + \mu \beta_t^\ell (\ell_t - \hat{\ell}_t) dt + \frac{1}{2} \mu \lambda_\ell (\hat{\ell}_t^2 - \ell_t^2) dt \right\} = 0$$

or equivalently

$$(\hat{s}_t, \hat{\ell}_t) = \arg \max_{s_t, \ell_t} \left\{ \alpha \beta_t^s (s_t - \hat{s}_t) + \frac{1}{2} \alpha \lambda_s (\hat{s}_t^2 - s_t^2) + \mu \beta_t^\ell (\ell_t - \hat{\ell}_t) dt + \frac{1}{2} \mu \lambda_\ell (\hat{\ell}_t^2 - \ell_t^2) dt \right\}$$

for given β_t^s, β_t^ℓ . After going through the maximization, we obtain that this is satisfied if $\hat{s}_t \lambda_s = \beta_t^s$ and $\hat{\ell}_t \lambda_\ell = \beta_t^\ell$, in case $(\hat{s}_t, \hat{\ell}_t) \in (0, s_{\max}) \times (0, \ell_{\max})$. If $\hat{a}_t \in \{\hat{s}_t, \hat{\ell}_t\}$ is not interior, in that $\hat{a}_t = a_{\max}$ for $a \in \{s, \ell\}$, then $a_t = \hat{a}_t$ solves the above maximization problem if and only if $\beta_t^a \geq \lambda_a \hat{a}_t$. The result follows. \square

A.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-equation (9). 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 effort $(\{s^*\}, \{\ell^*\})$, which we will denote for convenience by just \mathcal{P} , if no confusion is likely to arise. We follow an analogous convention concerning the expectation operator, where we will just write $\mathbb{E}_t(\cdot)$ instead of $\mathbb{E}^{s^*, \ell^*}(\cdot | \mathcal{F}_t)$.

A.3.1 Scaling of the value function

Given the optimal control and stopping problem (3), suppose that the principal's value function $P(W, K)$ satisfies the HJB-equation

$$rP = \max_{s, \ell, \beta^s, \beta^\ell} \left\{ \alpha s K + P_W \left[\gamma W + \frac{K}{2} (\lambda_s \alpha s^2 + \lambda_\ell \mu \ell^2) \right] + P_K \mu \ell K + \frac{1}{2} \left(P_{WW} \left[(\beta^s \sigma_X K)^2 + (\beta^\ell \sigma_K K)^2 + 2\rho \sigma_X \sigma_K K^2 \beta^s \beta^\ell \right] + P_{KK} (\sigma_K K)^2 + 2P_{WK} \left[(\sigma_K K)^2 \beta^\ell + \rho \sigma_X \sigma_K K^2 \beta^s \right] \right) \right\}$$

in some region $\mathcal{S} \subset \mathbb{R}^2$, 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 \mathcal{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

$$P_W = p'(w), P_K = p(w) - wp'(w), P_{WK} = -w/Kp''(w), P_{WW} = p''(w)/K, p_{KK} = w^2/Kp''(w),$$

which implies the HJB-equation (9) and its boundary conditions.

In the following, we will assume that (9) 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 therefore omitted.¹⁹

We first rewrite the principal's problem (3) in a convenient manner. Let

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

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

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

By arguments similar to the ones in Appendix A.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

$$dw_t + dc_t = \left[(\gamma - \mu\ell_t)w_t + \frac{1}{2}(\lambda_s s_t^2 \alpha + \lambda_\ell \ell_t^2 \mu) \right] dt + \beta_t^s \sigma_X d\tilde{Z}_t^X + (\beta_t^\ell - w_t)\sigma_K d\tilde{Z}_t^K$$

under $\tilde{\mathcal{P}}$. Finally, we are able to rewrite the principal's problem (3) as

$$\max_{\{c\}, \{s\}, \{\ell\}, w^*} \tilde{\mathbb{E}} \left[\int_0^\tau e^{-rt + \mu \int_0^t \ell_z dz} \alpha s_t dt - \int_0^\tau e^{-rt + \mu \int_0^t \ell_z dz} dc_t + e^{-r\tau + \mu \int_0^\tau \ell_z dz} R \Big| w_0 = w^* \right],$$

where the expectation $\tilde{\mathbb{E}}[\cdot]$ 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*

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

¹⁹Indeed, 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.

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

$$r_t f(w_t) = g(w_t) + f'(w_t)[\delta_t + \Delta_t w_t] + f''(w_t) \left[\sigma_K^2 (\beta_t^\ell - w_t)^2 + (\beta_t^s \sigma_X)^2 + 2\rho \sigma_X \sigma_K \beta_t^s (\beta_t^\ell - w_t) \right] \quad (\text{A2})$$

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.

Proof. Suppose $f(\cdot)$ solves (A2). Define

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

Applying Itô's Lemma, we obtain

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

The first term in curly brackets equals zero because $f(\cdot)$ solves (A2). 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. □

A.3.2 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-equation (9) 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

²⁰We 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 $[a, b]$.

its auxiliary gain process G as

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

where the agent's continuation payoff evolves according to (5). 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 $t \leq \tau$:

$$\begin{aligned} & e^{rt} \frac{dG_t(\Pi_1)}{K_t} \\ &= \left[- (r - \mu \ell_t) p(w_t) + \alpha s_t + p'(w_t) \left(w_t (\gamma - \mu \ell_t) + \frac{1}{2} (\lambda_s \alpha s_t^2 + \lambda_\ell \mu \ell_t^2) \right) \right. \\ & \quad \left. + \frac{p''(w_t)}{2} [(\beta_t^s \sigma_X)^2 + \sigma_K^2 (\beta_t^\ell - w_t)^2 + 2\rho \sigma_X \sigma_K \beta_t^s (\beta_t^\ell - w_t)] \right] dt - (1 + p'(w_t)) dc_t \\ & \quad + \sigma_K (p(w_t) + p'(w_t) (\beta_t^\ell - w_t)) dZ_t^K + \sigma_X (1 + \beta_t^s p'(w_t)) dZ_t^X. \end{aligned}$$

Under the optimal effort and incentives, the first term in square bracket stays at zero always. Other effort 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 $\mu_G(t) \leq 0$. This implies that $\{G_t\}_{t \geq 0}$ follows a supermartingale. Furthermore, under Π , investors' expected payoff is

$$\tilde{G}(\Pi) \equiv \mathbb{E} \left[\int_0^\tau e^{-ru} dX_u - \int_0^\tau e^{-ru} dC_u + e^{-r\tau} RK_\tau \right],$$

As a result, we have that

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

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.4 Proof of Proposition 2.3

A.4.1 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-equation (9) 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 $\{\hat{s}\}, \{\hat{\ell}\}$ 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 that the policy functions $s(\cdot), \ell(\cdot)$ are continuous in a neighbourhood of w_1 .*

Proof. We start with an important observation. Because the processes $\{\hat{s}\}, \{\hat{\ell}\}$ are by hypothesis of bounded variation, they can be written as the difference of two almost surely increasing processes, such that $\hat{a}_t = \hat{a}_t^1 - \hat{a}_t^2$ for all $t \geq 0$, $a \in \{s, \ell\}$ and $\hat{a}_{(\cdot)}^j$ 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 neighbourhood 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-equation (9) and the fact, that $\ell(z) = \ell^{FB}$ is not necessarily optimal, imply

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

It is now evident that there exists $\varepsilon > 0$ such that $s = s(z) = s_{\max} \geq s^{FB}$ and

$$\alpha s - \varepsilon [(\gamma - \mu\ell^{FB})z + \mathcal{C}(s, \ell^{FB}) + \Sigma(z)] > \alpha s^{FB} - \mathcal{C}(s^{FB}, \ell^{FB}) = (r - \mu\ell^{FB})p^{FB}.$$

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

²¹Froda'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.

Next, let us prove that $\ell(\cdot)$ must be continuous in a neighbourhood of w_1 and assume to the contrary that there is no neighbourhood 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

$$\Gamma_\ell(w) = p(w) - p'(w)w + p'(w)\lambda_\ell \ell_+.$$

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

$$\widehat{\Gamma}_\ell(w) = p(w) - p'(w)w + p'(w)\lambda_\ell \ell_-.$$

Next, observe that

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

Because we have already shown $p'(w_1) < 0$, it follows that $\ell_- \geq \ell_+$, a contradiction.

Finally, assume that there is no neighbourhood 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 $\frac{\partial p(z)}{\partial s} \Big|_{s=s(z)} \geq 0$. Taking the limit $\varepsilon \rightarrow 0$, we obtain $\Gamma_s(w_1) \geq 0$ for $\Gamma_s(w) = \alpha s(w) + p'(w)\lambda_s s_+$. Similarly, $\widehat{\Gamma}_s(w_1) = 0$ for $\widehat{\Gamma}_s(w) = s(w) + p'(w)\lambda_s s_-$. Hence,

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

Due to $p'(w_1) < 0$ it follows that $s_- \geq s_+$, a contradiction. \square

A.4.2 Concavity of the value function

Proof. Since $p''(\cdot)$ is continuous on $[0, \bar{w}]$ and $\{\hat{s}\}, \{\hat{\ell}\}$ 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

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

that $p'''(\cdot)$ exists and is given by

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

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 neighbourhood 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 neighbourhood 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 neighbourhood of w_1 and that $p'(w_1) < 0$. However, this implies $p'''(w_1) > 0$ and therefore $p'''(\cdot) > 0$ in a neighbourhood 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

B Proofs of Propositions 3 and 4

Proof. This follows 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-equation (9). Interior levels $s(w), \ell(w)$ must solve the respective first order conditions of maximization, that is $\frac{\partial p(w)}{\partial s}|_{s=s(w)} = 0$ and $\frac{\partial p(w)}{\partial \ell}|_{\ell=\ell(w)} = 0$. After rearranging the FOC's of the maximization, one arrives at the desired expressions. \square

C Proof of Proposition 5

Proof. At the payout boundary \bar{w} , we have that

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

Taking the first order condition of maximization with respect to s , we obtain that $s(\bar{w}) = \frac{1}{\lambda_s}$, if interior, and $s = s_{\max}$ otherwise, i.e. $s(\bar{w}) = s^{FB}$. Similarly, the first order condition with respect to ℓ reads

$$0 = p(\bar{w}) + \bar{w} - \lambda_\ell \ell < p^{FB} - \lambda_\ell \ell,$$

where $p^{FB} = P^{FB}/K_0$. Note that $\ell(\bar{w})$ solves this first order condition, as we assume it to be interior. Because ℓ^{FB} solves $p^{FB} - \lambda_\ell \ell = 0$, it follows that $\ell(\bar{w}) < \ell^{FB}$. \square

D Proof of Proposition 6

In this section, we prove the natural generalization of Proposition 6. We start with the following Lemma.

Lemma 3. *Let $w \in (0, \bar{w}]$ such that in optimum $\ell(w) = \ell = \ell_{\max}$ and $s(w) = s \in [0, s_{\max}]$. Assume that parameters satisfy $-\rho\sigma_K\lambda_\ell\ell_{\max} < \sigma_X\lambda_s s_{\max}$ for $\rho \in (-1, 1)$. Then*

$$\beta^\ell \equiv \beta^\ell(w) = \max \left\{ \lambda_\ell \ell_{\max}, w - \rho \frac{\sigma_X}{\sigma_K} \lambda_s s \right\} \text{ and } \beta^s \equiv \beta^s(w) = \lambda_s s.$$

Hence, the short-run IC-condition is always tight under the conditions stated.

Proof. Given the optimal choice $\ell(w) = \ell_{\max}$, $s(w) = s$, the tuple $(\beta^s(w), \beta^\ell(w))$ must satisfy

$$\begin{aligned} (\beta^s(w), \beta^\ell(w)) &= \arg \min_{\beta^s, \beta^\ell} \left[(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho\sigma_X\sigma_K\beta^s(\beta^\ell - w) \right] \\ &\text{subject to } \beta^\ell \geq \lambda_\ell \ell_{\max} \text{ and } \beta^s \geq \lambda_s s, \end{aligned}$$

where the last inequality is tight, unless $s = s_{\max}$. Using standard arguments, one obtains:

$$\begin{aligned} \beta^\ell \equiv \beta^\ell(w) &= \max \left\{ \lambda_\ell \ell_{\max}, w - \rho \frac{\sigma_X}{\sigma_K} \beta^s \right\}; \\ \beta^s \equiv \beta^s(w) &= \max \left\{ \lambda_s s_{\max}, \rho \frac{\sigma_K}{\sigma_X} (w - \beta^\ell) \right\} \text{ if } s = s_{\max} \text{ and } \beta^s = \lambda_s s \text{ otherwise.} \end{aligned}$$

The claim is trivial if $s < s_{\max}$ or $\rho = 0$.

Let us suppose $s = s_{\max}$, $\rho \neq 0$ and $\beta^s > \lambda_s s$. Hence, $\beta^s = \rho \frac{\sigma_K}{\sigma_X} (w - \beta^\ell)$. If now $\beta^\ell > \lambda_\ell \ell_{\max}$, then $\beta^\ell = w - \rho\sigma_X/\sigma_K\beta^s$. This implies $\rho\sigma_K/\sigma_X(w - \beta^\ell) = \rho^2\beta^s < \beta^s$ and hence $\beta^s = \lambda_s s_{\max}$, a contradiction.

Next, suppose $\rho < 0$ and $\beta^\ell = \lambda_\ell \ell_{\max}$. Hence, $w > \lambda_\ell \ell_{\max}$. Since $\beta^\ell = \lambda_\ell \ell_{\max}$ it follows that $\lambda_\ell \ell_{\max} > w - \rho\sigma_X/\sigma_K\beta^s$ and - using $\beta^s = \rho \frac{\sigma_K}{\sigma_X} (w - \beta^\ell)$ - one obtains $\lambda_\ell \ell_{\max} > w - \rho^2(w - \lambda_\ell \ell_{\max})$. Hence, $\lambda_\ell \ell_{\max} > w$, a contradiction.

Finally, assume $s = s_{\max}$, $\rho < 0$ and $\beta^\ell = \lambda_\ell \ell_{\max}$. Hence, $\lambda_\ell \ell_{\max} > w$ and $\rho\sigma_K/\sigma_X(w - \lambda_\ell \ell_{\max}) > \lambda_s s_{\max}$, which implies $w - \lambda_\ell \ell_{\max} < \lambda_s s_{\max} \sigma_X / (\sigma_K \rho)$. Therefore, $-\rho\sigma_K\lambda_\ell \ell_{\max} > \sigma_X\lambda_s s_{\max}$, which contradicts the hypothesis. \square

The claim of Proposition 6 follows then already from Lemma 3.

Next, we state Lemma 4, which analyzes the non-binding IC-condition in a more general setting.

Lemma 4. *Assume the usual regularity conditions, $-\rho\sigma_K\lambda_\ell\ell_{\max} < \sigma_X\lambda_s s_{\max}$ and $R, p(\bar{w}) > 0$. Further, assume that $\bar{w} > \lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K}$ and that $\ell(w) = \ell_{\max} = \ell^{FB}$ in a left-neighbourhood of \bar{w} . Then the following holds true:*

- i) *There exists $w' \geq \lambda_\ell \ell_{\max} + \rho \frac{\sigma_X}{\sigma_K} \lambda_s s(w')$, such that for all $w \in [w', \bar{w}]$ long-run effort satisfies $\ell(w) = \ell_{\max}$. Further, the IC-condition $\beta^\ell(w) \geq \lambda_\ell \ell_{\max}$ is not tight on $[w', \bar{w}]$, in that $\beta^\ell(w) = w - \rho \frac{\sigma_X}{\sigma_K} \lambda_s s > \lambda_\ell \ell_{\max}$.*
- ii) *Suppose $s(w) = s_{\max}$ for all $w \geq \lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K} \lambda_s s_{\max}$ as well as $\lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K} \lambda_s s_{\max} \geq w^*$. Then $\beta^\ell(w) = w - \rho \frac{\sigma_X}{\sigma_K} \lambda_s s_{\max} > \lambda_\ell \ell_{\max}$ and therefore $\ell(w) = \ell_{\max}$ for all $w > \lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K} \lambda_s s_{\max}$.*

iii) If $\rho = 0$, then $\beta^\ell(w) = w$ and $\ell(w) = \ell_{\max}$ for all $w \geq \max\{\lambda_\ell \ell_{\max}, w^*\}$.

Proof. i) Note that Lemma 2 implies that $s(\cdot)$ must be continuous in a (left-) neighbourhood of \bar{w} . Hence, the existence of the desired $w' \geq \lambda_\ell \ell_{\max} + \rho \frac{\sigma_X}{\sigma_K} \lambda_s s(w')$ follows from the assumption

$$\bar{w} > \lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K} \lambda_s s(\bar{w}) = \lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K}$$

and Lemma 3.

ii) Let us suppose to the contrary that there exists $w' \geq \lambda_\ell \ell_{\max} + \rho^+ \frac{\sigma_X}{\sigma_K} \lambda_s s_{\max}$ with $\ell(w') < \ell_{\max}$. The optimality of $\ell(w')$ implies

$$p(w') + p'(w')(\lambda_\ell \ell_{\max} - w') \leq 0$$

and therefore $p'(w') > 0$, $w' > \lambda_\ell \ell_{\max}$ due to $p(w) \geq \max\{p(0), p(\bar{w})\} > 0$. This contradicts the concavity of $p(\cdot)$ and the assumption that $w' \geq w^*$. Again Lemma 3 yields the claim.

iii) If there existed $w \geq \max\{\lambda_\ell \ell_{\max}, w^*\}$ such that $\ell(w) < \ell_{\max}$, it would readily follow that $p'(w) > 0$ if $w > w^*$ or $p(w) = 0$ if $w = w^*$, a contradiction. Hence, the result follows by Lemma 3. □

E Solution with One-dimensional Moral Hazard

E.1 Moral Hazard Only over the Short-run

Assume that the process $\{\ell\}$ is observable for the principal and thus contractible, yet there are still both types of cash-flow shocks present. Formally, the (public) information filtration is given by

$$\mathbb{F}^\ell \equiv \{\mathcal{F}_t^\ell : t \geq 0\} \text{ with } \mathcal{F}_t^\ell = \sigma(X_s, K_s, \ell_s : 0 \leq s \leq t) = \mathcal{F}_t \vee \sigma(\ell_s : 0 \leq s \leq t).$$

By standard arguments, one obtains the HJB-equation

$$rp(w) = \max_{s, \ell, \beta^s, \beta^\ell} \left\{ \alpha s + \mu \ell p(w) - \frac{\lambda_\ell}{2} \mu \ell^2 + p'(w) w (\gamma - \mu \ell) + \frac{p'(w)}{2} \lambda_s \alpha s^2 + \frac{p''(w)}{2} \left[(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) \right] \right\},$$

subject to $p(0) = R$, $p'(\bar{w}) = -1$ and $p''(\bar{w}) = 0$. In contrast to the baseline model, the maximization is only subject to the IC-condition $\beta^s \geq \lambda_s s$, where the inequality is tight if $s < s_{\max}$. Without loss of generality, we focus on interior levels, i.e. assume for all w that $s(w) \in (0, s_{\max})$.

The unconstrained maximization over β^ℓ yields now the optimal value $\beta^\ell = \beta^\ell(w) = w - \rho \frac{\sigma_X}{\sigma_K} \beta^s$. Whence,

$$\Sigma(w) \equiv \frac{\mathbb{V}(dw)}{dt} = \frac{[dw, dw]}{dt} = (\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) = (\beta^s \sigma_X)^2 (1 - \rho^2).$$

The optimal values of s, ℓ are now given by

$$s(w) = s = \frac{\alpha}{-p'(w)\lambda_s \alpha - p''(w)\lambda_s \sigma_X^2 (1 - \rho^2)} \text{ and } \ell(w) = \ell = \frac{p(w) - wp'(w)}{\lambda_\ell}.$$

Because $p'(w) \geq -1$, it follows that $p(w) - wp'(w) \leq p(w) + w < p^{FB}$, which implies that $\ell(w) \leq \ell^{FB}$, where the inequality is tight if $\ell^{FB} < \ell_{\max}$.

Because $s^{FB} = 1/\lambda_s$, it follows that $s(\bar{w}) = s^{FB}$. Furthermore, a necessary and sufficient condition for $s(w) < s^{FB}$ for all $w \in [0, \bar{w})$ is given by $p'(w)\alpha + p''(w)\sigma_X^2(1 - \rho^2) \leq -1$ for $w \in [0, \bar{w})$. Since correlation ρ is ‘exploited’ to dampen the magnitude of transitory shocks $\{Z^X\}$, the solution of the model with parameters $(\sigma_X, \sigma_K, \rho)$ and observable $\{\ell\}$ is isomorphic to the solution of the model with parameters $(\sigma_X \sqrt{1 - \rho^2}, 0, 0)$, in that the value functions will be identical.

E.2 Moral Hazard Only over the Long-run

Assume that the process $\{s\}$ is observable by the principal and thus contractible. Formally, the (public) information filtration is given by

$$\mathbb{F}^s \equiv \{\mathcal{F}_t^s : t \geq 0\} \text{ with } \mathcal{F}_t^s = \sigma(X_k, K_k, s_k : 0 \leq k \leq t) = \mathcal{F}_t \vee \sigma(s_k : 0 \leq k \leq t).$$

One obtains the HJB-equation

$$rp(w) = \max_{s, \ell, \beta^s, \beta^\ell} \left\{ \alpha s + \mu \ell p(w) - \frac{\lambda_s}{2} \alpha s^2 + p'(w)w(\gamma - \mu \ell) + \frac{p'(w)}{2} \lambda_\ell \mu \ell^2 + \frac{p''(w)}{2} \left[(\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) \right] \right\},$$

subject to $p(0) = R$, $p'(\bar{w}) = -1$ and $p''(\bar{w}) = 0$. The unconstrained maximization over β^s gives $\beta^s = \rho \frac{\sigma_K}{\sigma_X} (w - \beta^\ell)$. Whence,

$$\Sigma(w) \equiv \frac{\mathbb{V}(dw)}{dt} = \frac{[dw, dw]}{dt} = (\beta^s \sigma_X)^2 + \sigma_K^2 (\beta^\ell - w)^2 + 2\rho \sigma_X \sigma_K \beta^s (\beta^\ell - w) = \sigma_K^2 (\beta^\ell - w)^2 (1 - \rho^2).$$

The optimal values of s, ℓ , if interior, are now given by

$$s(w) = s = \frac{1}{\lambda_s} \text{ and } \ell(w) = \frac{\mu(p(w) - p'(w)w) - p''(w)w\lambda_\ell\sigma_K^2(1 - \rho^2)}{-p'(w)\lambda_\ell\mu - p''(w)(\lambda_\ell\sigma_K)^2(1 - \rho^2)}.$$

Note that the solution in this model is in general well behaved, in that there are no

points w , for which $\beta^\ell(w) = w$ and the second-order ODE collapses to one of the first-order. This is guaranteed by Proposition 7, to be proven in Appendix F. Since correlation ρ is ‘exploited’ to dampen the magnitude of transitory shocks $\{Z^K\}$, the solution of the model with parameters $(\sigma_K, \sigma_X, \rho)$ and observable $\{s\}$ is isomorphic to the solution of the model with parameters $(\sigma_K \sqrt{1 - \rho^2}, 0, 0)$, in that the value functions will be identical.

F Proof of Proposition 7

The claim regarding observable long-run effort $\{\ell\}$ follows directly from Appendix E. We therefore only prove the claim regarding observable short-run effort $\{s\}$. We split the proof in two parts.

F.1 Part 1 - Auxiliary Results

In this section, we establish two auxiliary results, that will be used in the second part of the proof of Proposition 7.

Lemma 5. *Assume the usual regularity conditions and suppose that $\{s\}$ is publicly observable (e.g. $\sigma_X = 0$). Then short-run effort $s(w)$ is contractible and constant over time. If in addition $\ell(w) = \ell_{\max}$ in a left neighbourhood of \bar{w} , then it must be that $\bar{w} < \lambda_\ell \ell_{\max}$.*

Proof. Without loss of generality, we normalize $s(w) = s \equiv 1$ for all w , i.e. set $\alpha = \lambda_s = 1$. Further, in light of Appendix E it is obvious, that it suffices to prove the claim for $\sigma_X = 0$.

- i) Define $\ell(\bar{w}) = \ell = \ell_{\max}$ and by assumption $\ell(w) = \ell$ in an (open) left neighbourhood of \bar{w} .

Let us first show that $\lambda_\ell \ell \neq \bar{w}$. Suppose to the contrary $\lambda_\ell \ell = \bar{w}$. Then

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

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 $\ell_\varepsilon \equiv \ell(\bar{w} - \varepsilon)$ and note that in optimum $\beta^\ell(\bar{w} - \varepsilon) = \lambda_\ell \ell_\varepsilon$. Hence,

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

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

Combining the above yields

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

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

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

However, it follows that for $\varepsilon > 0$ sufficiently close to zero $\ell'(\bar{w} - \varepsilon) = 0$, a contradiction.

- ii) Next, suppose $\bar{w} > \lambda_\ell\ell_{\max}$. Thus, there exists $w' \geq \lambda_\ell\ell_{\max}$ with $\bar{w} > w'$ such that $\ell \equiv \ell(w) = \ell_{\max}$ for all $w \geq w'$. In optimum, $\beta^\ell(w') = w'$. Then,

$$\begin{aligned} p(w') < p(\bar{w}) - (w' - \bar{w}) &= \frac{1}{r - \mu\ell} \left(0.5 - \frac{\lambda_\ell\ell^2\mu}{2} - \bar{w}(\gamma - r) - w'(r - \mu\ell) \right) \\ &< \frac{1}{r - \mu\ell} \left(0.5 - \frac{\lambda_\ell\ell^2\mu}{2} - w'(\gamma - r) - w'(r - \mu\ell) \right) \\ &= \frac{1}{r - \mu\ell} \left(0.5 - \frac{\lambda_\ell\ell^2\mu}{2} - w'(\gamma - \mu\ell) \right), \end{aligned}$$

where the first inequality is due to strict concavity and the second one due to $w' < \bar{w}$. However, as the 'firm' becomes 'riskless' at w' , we get that

$$p(w') \geq \frac{1}{r - \mu\ell} \left(0.5 - \frac{\lambda_\ell\ell^2\mu}{2} - w'(\gamma - \mu\ell) \right).$$

The inequality stems from the fact that constant, scaled payouts at rate $w'(\gamma - \mu\ell) + \frac{\lambda_\ell\ell^2\mu}{2}$ and this way keeping $w_t = w'$ constant for all future times t is always an option but not necessarily optimal. This yields the desired contradiction. □

While the previous result is only valid under the assumption that short-run effort $\{s\}$ is observable, we state now a related claim for interior levels, which does not hinge on the observability of $\{s\}$.

Lemma 6. *Assume the usual regularity conditions and $\min\{R, p(\bar{w})\} > 0$. If in addition $\rho \leq 0$ and $\ell(w) < \ell_{\max}$ for all $w \in [0, \bar{w}]$, it holds that $\beta^\ell(w) > w$ or equivalently $\ell(w) > w/\lambda_\ell$ for all $w \in [0, \bar{w}]$.*

Proof. Let us suppose to the contrary that there exists $w_0 \in [0, \bar{w}]$, such that $\beta^\ell(w_0) \leq w_0$. By assumption, $\ell(w_0) < \ell_{\max}$ and by (14) (and $\rho \leq 0$) it is evident, that $\ell(w_0) > 0$. Hence,

$\ell_0 \equiv \ell(w_0)$ solves $\frac{\partial p(w_0)}{\partial \ell} = 0$, given β^s . We may without loss of generality assume that $\beta^s \geq 0$. Hence,

$$\mu[p(w_0) - w_0 p'(w_0)] + \mu \lambda_\ell \ell_0 p'(w_0) + p''(w_0) [\lambda_\ell (\lambda_\ell \ell_0 - w_0) + \rho \sigma_X \sigma_K \lambda_\ell \beta^s] = 0.$$

Because $\beta^\ell(w_0) = \lambda_\ell \ell_0 \leq w_0$ and $p''(w_0) \leq 0$, it is immediate that

$$p(w_0) + p'(w_0)(\lambda_\ell \ell_0 - w_0) \leq 0,$$

which implies - due to $p(w_0) \geq \min\{R, p(\bar{w})\} > 0$ - that $p'(w_0) > 0$ as well as $\lambda_\ell \ell_0 < w_0$. Hence, $w_0 < w^*$. Next, observe that at $w = 0$, it must be $\ell(0) > 0$ due to $p(0) = R > 0$. Because long-run effort is interior for all $[0, \bar{w}]$, the mapping $\ell(\cdot)$ is continuous on the same interval (due to $p \in C^2$). Hence, the function $\chi : w \mapsto \lambda_\ell \ell(w) - w$, defined on $[0, \bar{w}]$, is also continuous and must therefore have a root w_1 on $(0, w_0)$. However, the optimality of $\ell(w_1) = w_1/\lambda_\ell$ implies $p(w_1) \leq 0$, which contradicts $p(w_1) \geq \min\{R, p(\bar{w})\} > 0$. \square

F.2 Part 2

Proof. Without loss of generality, we normalize in the following σ_X to zero.

- i) As a first step, we show that if there exists $w' \in [0, \bar{w}]$ with $\beta^\ell(w') = w'$, then it must be that $\ell(w) = \ell_{\max}$ for all $w \in [w', \bar{w}]$ and $\bar{w} > \lambda_\ell \ell_{\max}$.

Hence, assume there is $w' \in [0, \bar{w}]$, such that $\beta^\ell(w') = w'$. Then, by Lemma 6 it follows that $\ell(w') = \ell_{\max}$. Hence, $\lambda_\ell \ell_{\max} \leq w' = \beta^\ell(w')$ as well as $\bar{w} > \lambda_\ell \ell_{\max}$. Observe that for $\ell(w') = \ell_{\max}$ to be optimal it must hold that

$$p(w') + (\lambda_\ell \ell_{\max} - w') p'(w') \geq 0.$$

However, by strict concavity of $p(\cdot)$ on $[w', \bar{w}]$, it follows for all $w \in (w', \bar{w}]$ that

$$\frac{\partial p(w)}{\partial \ell} \Big|_{\ell=\ell_{\max}} = p(w) + p'(w)(\lambda_\ell \ell_{\max} - w) > p(w') + (\lambda_\ell \ell_{\max} - w') p'(w') \geq 0.$$

This readily implies that $\ell(w) = \ell_{\max}$ for all $w \in [w', \bar{w}]$.

- ii) Next, let us prove that $\beta^\ell(w') = w'$ cannot be for $w' < \bar{w}$. This can be proven by replicating part ii) of the proof of Lemma 5, because we have already shown that $\beta^\ell(w') = w'$ implies $\ell(\cdot) \equiv \ell_{\max}$ on $[w', \bar{w}]$ and $\bar{w} > \lambda_\ell \ell_{\max}$.
- iii) To complete the proof, we establish that $\bar{w} \neq \beta^\ell(\bar{w})$. Hence, suppose to the contrary that $\bar{w} = \beta^\ell(\bar{w})$. Then, Lemma 6 implies that $\ell(\bar{w}) = \ell_{\max}$ and therefore $\bar{w} \geq \lambda_\ell \ell_{\max}$.

First, let us rule out $\bar{w} = \lambda_\ell \ell_{\max}$. If this were the case, we are able to establish - replicating the argumentation from part i) of the proof of Lemma 5 - that it must be

$$o(\varepsilon) - \mu(p(\bar{w} - \varepsilon) + \bar{w}) \ell'(\bar{w} - \varepsilon) = r - \gamma.$$

If it were $\ell(w) = \ell_{\max}$ in a left neighbourhood of w_{\max} , the claim follows by Lemma 5. Thus, we may assume this is not the case.

Then, there exists for any $\delta > 0$ a value $w' \in (\bar{w} - \delta, \bar{w})$, such that $\ell(w') < \ell_{\max}$. It follows that $\ell(w) < \ell_{\max}$ on an interval $(\bar{w} - \delta, \bar{w})$, because $\ell(w') = \ell_{\max}$ for $w' < \bar{w}$ implies $\ell(w) = \ell_{\max}$ on $[w', \bar{w}]$ by part i) of the proof.

Further, by Lemma 6, it must be that $\ell(w) > w/\lambda_\ell$ in a left neighbourhood of and excluding \bar{w} , i.e. the interval $(\bar{w} - \delta, \bar{w})$. It follows that $p'''(\cdot)$ exists on $(\bar{w} - \delta, \bar{w})$, with $\lim_{w \uparrow \bar{w}} p'''(w) = \infty$. But then also

$$\lim_{w \uparrow \bar{w}} \frac{\partial \ell(x)}{\partial x} \Big|_{x=w} = \lim_{w \uparrow \bar{w}} \ell'(w) = \lim_{w \uparrow \bar{w}} \frac{\sigma_K^2 p(w) p'''(w)}{\mu} = \infty.$$

However, this implies that

$$o(\varepsilon) - \mu(p(\bar{w} - \varepsilon) + \bar{w})\ell'(\bar{w} - \varepsilon) = r - \gamma.$$

cannot hold for all $\varepsilon > 0$ due to $p(\bar{w}) + \bar{w} > 0$, a contradiction.

Second, we show that $\bar{w} > \lambda_\ell \ell_{\max}$ cannot occur. Suppose now to the contrary $\bar{w} > \lambda_\ell \ell_{\max}$. By part ii) of the proof, there cannot be $w' \in [\lambda_\ell \ell_{\max}, \bar{w})$ such that $\ell(w') = \ell_{\max}$. Hence, it follows that $\ell(w) < \ell_{\max}$ for all $w \in [\lambda_\ell \ell_{\max}, \bar{w})$. However, this implies that

$$p(w) + p'(w)(\lambda_\ell \ell_{\max} - w) < 0$$

holds for all $w \in [\lambda_\ell \ell_{\max}, \bar{w})$. Whence, $p'(w) > 0$ for all $w \in [\lambda_\ell \ell_{\max}, \bar{w})$, a contradiction to $p'(\bar{w}) = -1$.

□

Combining Proposition 7 and Lemma 5, we even obtain that $\beta^\ell(w) > w$, whenever $\rho \leq 0$.

G Return Volatility

The principal's instantaneous return dR_t at time t is given by

$$dR_t = dX_t + dP(W_t, K_t),$$

that is by the change in cash-flow and contract value. By Itô's Lemma:

$$\begin{aligned} dX_t + dP(W_t, K_t) &= rP(W_t, K_t)dt + \sigma_X K_t dZ_t^X \\ &\quad + P_W(W_t, K_t) \left(\beta_t^\ell \sigma_K K_t dZ_t^K + \beta_t^s \sigma_X K_t dZ_t^X \right) + P_K(W_t, K_t) \sigma_K K_t dZ_t^K, \end{aligned}$$

where we used that

$$rP(W_t, K_t)dt = \mathbb{E} [dX_t + dP(W_t, K_t)] \iff (r - \mu \ell_t)p(w_t) = \alpha s_t + \frac{\tilde{\mathbb{E}}[dp(w_t)]}{dt}$$

holds in optimum by the HJB-equation (9). Note that the expectation $\widetilde{\mathbb{E}}[\cdot]$ is taken under the auxiliary, equivalent probability measure $\widetilde{\mathcal{P}}$, defined in Appendix A.3.1. Next, observe that $P(W_t, K_t) = K_t p(w_t)$ and therefore $P_W(W_t, K_t) = p'(w_t)$ as well as $P_K(W_t, K_t) = p(w_t) - w_t p'(w_t)$. Consequently, straightforward calculations yield

$$\begin{aligned} \frac{dR_t}{P(W_t, K_t)} &= rdt + \frac{1 + \beta_t^s p'(w_t)}{p(w_t)} \sigma_X dZ_t^X + \frac{p(w_t) + (\beta_t^\ell - w_t) p'(w_t)}{p(w_t)} \sigma_K dZ_t^K \\ &= rdt + \Sigma_X^*(w_t) dZ_t^X + \Sigma_K^*(w) dZ_t^K, \end{aligned}$$

where

$$\Sigma_X^*(w) = \frac{1 + \beta^s p'(w)}{p(w)} \sigma_X \text{ and } \Sigma_K^*(w) = \sigma_K + \frac{p'(w)(\beta^\ell - w)}{p(w)} \sigma_K.$$

The squared volatility term is given by

$$(\Sigma_P^*(w))^2 \equiv \frac{\mathbb{V}(dR_t)}{P(W_t, K_t)^2 dt} = \frac{[dR_t, dR_t]}{P(W_t, K_t)^2 dt}.$$

References

- Abel, A., and J. Eberly, 2011, How q and cash flow affect investment without frictions: An analytic explanation, *Review of Economic Studies* 78, 1179–1200.
- Albuquerque, R., and H. Hopenhayn, 2004, Optimal lending contracts and firm dynamics, *Review of Economic Studies* 71, 285–315.
- Asker, J., J. Farre-Mensa, and A. Ljunqvist, 2015, Corporate investment and stock market listing: A puzzle, *Review of Financial Studies* 28, 342–390.
- Barton, D., J. Manyika, and S. Williamson, 2017, The data: Where long-termism pays off, *Harvard Business Review* 95, 67–68.
- Bernstein, S., 2015, Does going public affect innovation?, *Journal of Finance* 70, 1365–1403.
- Biais, B., T. Mariotti, G. Plantin, and J.-C. Rochet, 2007, Dynamic security design: Convergence to continuous time and asset pricing implications, *Review of Economic Studies* 74, 345–390.
- Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve, 2010, Large risks, limited liability, and dynamic moral hazard, *Econometrica* 78, 73–118.
- Bolton, P., H. Chen, and N. Wang, 2011, A unified theory of tobin’s q , corporate investment, financing, and risk management, *Journal of Finance* 66, 1545–1578.
- Bolton, P., N. Wang, and J. Yang, 2017, Liquidity and risk management: Coordinating investment and compensation policies, forthcoming, *Journal of Finance*.
- Brochet, F., M. Loumiotis, and G. Serafeim, 2012, Short-termism: Don’t blame investors, *Harvard Business Review* 90, 28–29.
- Brochet, F., M. Loumiotis, and G. Serafeim, 2013, Short-termism, investor clientele, and firm risk, Working paper, Harvard Business School.
- Brunnermeier, M., and Y. Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104, 379–421.
- Carlson, M., A. Fisher, and R. Giammarino, 2006, Corporate investment and asset price dynamics: Implications for SEO event-studies and long-run performance, *Journal of Finance* 61, 1009–1034.
- Chang, X., S. Dasgupta, G. Wong, and J. Yao, 2014, Cash-flow sensitivities and the allocation of internal cash flow, *Review of Financial Studies* 27, 3628–3657.

- Cox, J., J. Ingersoll, and S. Ross, 1985, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363–384.
- Décamps, J.-P., S. Gryglewicz, E. Morellec, and S. Villeneuve, 2017, Corporate policies with permanent and transitory shocks, *Review of Financial Studies* 30, 162–210.
- Décamps, J.-P., T. Mariotti, J.-C. Rochet, and S. Villeneuve, 2011, Free cash flow, issuance costs, and stock prices, *Journal of Finance* 66, 1501–1544.
- DeMarzo, P., and M. Fishman, 2007, Optimal long-term financial contracting, *Review of Financial Studies* 20, 2079–2128.
- DeMarzo, P., M. Fishman, Z. He, and N. Wang, 2012, Dynamic agency and the q theory of investment, *Journal of Finance* 67, 2295–2340.
- DeMarzo, P., and Y. Sannikov, 2006, Optimal security design and dynamic capital structure in a continuous-time agency model, *Journal of Finance* 61, 2681–2724.
- Edmans, A., X. Gabaix, and D. Jenter, 2017, Executive compensation: A survey of theory and evidence, Working paper, London Business School.
- Francis, B., H. Iftekhhar, J. Kose, and S. Zenu, 2013, Asymmetric benchmarking of pay in firms, *Journal of Corporate Finance* 23, 39–53.
- Frydman, C., and R. Saks, 2010, Executive compensation: A new view from a long-term perspective, 1936–2005, *Review of Financial Studies* 23, 2099–2138.
- Garvey, G., and T. Milbourn, 2006, Asymmetric benchmarking in compensation: Executives are rewarded for good luck but not penalized for bad, *Journal of Financial Economics* 82, 197–225.
- Gertler, M., and N. Kiyotaki, 2010, Financial intermediation and credit policy in business cycle analysis, In *Friedman, B., and Woodford, M. (Eds.), Handbook of Monetary Economics* Elsevier, Amsterdam, Netherlands.
- Graham, J., M. Leary, and M. Roberts, 2015, A century of corporate capital structure: The leveraging of corporate america, *Journal of Financial Economics* 118, 658–683.
- Gryglewicz, S., and B. Hartman-Glaser, 2017, Dynamic agency and real options, Working paper, Erasmus University Rotterdam.
- Gryglewicz, S., B. Hartman-Glaser, and G. Zheng, 2017, Growth options, incentives, and pay-for-performance: Theory and evidence, Working paper, Erasmus University Rotterdam.

- Gutierrez, German, and Thomas Philippon, 2017, Investment-less growth: An empirical investigation, Working paper, New York University.
- Hackbarth, D., A. Rivera, and T.-Y. Wong, 2017, Optimal short-termism, Working paper, Boston University.
- He, Z., 2009, Optimal executive compensation when firm size follows geometric brownian motion, *Review of Financial Studies* 22, 859–892.
- He, Z., 2011, A model of dynamic compensation and capital structure, *Journal of Financial Economics* 100, 351–366.
- Hoffmann, F., and S. Pfeil, 2010, Reward for luck in a dynamic agency model, *Review of Financial Studies* 23, 3329–3345.
- Holmstrom, B., 1979, Moral hazard and observability, *Bell Journal of Economics* 10, 74–91.
- Holmstrom, B., and P. Milgrom, 1987, Aggregation and linearity in the provision of intertemporal incentives, *Econometrica* 55, 303–328.
- Hugonnier, J., S. Malamud, and E. Morellec, 2015, Capital supply uncertainty, cash holdings, and investment, *Review of Financial Studies* 28, 391–445.
- Jensen, M., and K. Murphy, 1990, Performance pay and top-management incentives, *Journal of Political Economy* 98, 225–264.
- Kaplan, S., 2017, Are U.S. companies too short-term oriented: Some thoughts, Working paper, University of Chicago.
- Kogan, L., 2004, Asset prices and real investment, *Journal of Financial Economics* 73, 411–432.
- Leland, H., 1994, Corporate debt value, bond covenants, and optimal capital structure, *Journal of Finance* 49, 1213–1252.
- Malenko, A., 2016, Optimal dynamic capital budgeting, Working paper, MIT.
- Miao, J., and A. Rivera, 2016, Robust contracts in continuous time, *Econometrica* 84, 1405–1440.
- Morellec, E., B. Nikolov, and N. Schürhoff, 2012, Corporate governance and capital structure dynamics, *Journal of Finance* 67, 803–848.

- Morellec, E., and N. Schürhoff, 2011, Corporate investment and financing under asymmetric information, *Journal of Financial Economics* 99, 262–288.
- Nikolov, B., L. Schmid, and R. Steri, 2017, Testing dynamic agency theory via structural estimation, Working paper, Duke University.
- Rampini, A., and S. Viswanathan, 2013, Collateral and capital structure, *Journal of Financial Economics* 109, 466–492.
- Sannikov, Y., 2008, A continuous-time version of the principal-agent problem, *Review of Economic Studies* 75, 957–984.
- Sannikov, Y., 2014, Moral hazard and long-run incentives, Working paper, Princeton University.
- Shleifer, A., and R. Vishny, 1990, Equilibrium short horizons of investors and firms, *American Economic Review* 80, 148–153.
- Shreve, S., 2004, *Stochastic calculus for finance II: Continuous-time models* (Springer Verlag).
- Spear, S., and S. Srivastava, 1987, On repeated moral hazard with discounting, *Review of Economic Studies* 54, 599–617.
- Stein, J., 1989, Efficient capital markets, inefficient firms: A model of myopic corporate behavior, *The Quarterly Journal of Economics* 104, 655–669.
- Strebulaev, I., 2007, Do tests of capital structure theory mean what they say?, *Journal of Finance* 62, 1747–1787.
- Szydłowski, M., 2015, Incentives, project choice, and dynamic multitasking, Working paper, University of Minnesota.
- Zhu, J., 2012, Optimal contracts with shirking, *Review of Economic Studies* 80, 812–839.