Equity Prices and the Dynamics of Corporate Governance*

Arash Fahim† Simon Gervais‡ R. Vijay Krishna§

May 11, 2021

Abstract

Measures of firm performance and corporate governance are well known to be positively correlated, and performance-sensitivity of compensation increases in firm performance. We describe how governance affects stock prices, and in turn, how stock prices affect governance, and how compensation is affected by both of these. We present a model of firm financing structure that permits a unified analysis of corporate governance, pay sensitivity of the agent's compensation, and how these relate to stock prices and other securities issued by the firm. Our main results show why corporate governance and stock prices are positively correlated, and how governance and pay sensitivity are substitutes, which rationalises these empirical observations. Our setting allows us to analyse the impact of policy interventions like the Sarbanes-Oxley Act. We also propose a measure of governance in terms of observables of the securities the firm issues.

*This version is still work in progress, so please do not cite or circulate without the authors' permission. In particular, proofs have not yet been fully incorporated into the appendix.

†217 Love Building, Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, FL 32306, arashfahim@gmail.com, (850) 644-0617.

‡Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708-0120, sgervais@duke.edu, +(919) 660-7683.

§Department of Economics, 239 Bellamy Building, Florida State University, Tallahassee, FL 32306-2180, rvk3570@gmail.com, (850) 645-9040.

1. Introduction

Firms hire managers to run them. They also put in place governance structures, because their managers are capable of malfeasance and misappropriation. As Demsetz and Lehn (1985a) note, if governance structures are chosen optimally in each firm, there should be no reason that performance measures such as stock prices or credit yield spreads should be correlated with the amount of governance (however this is defined). However, a sizeable empirical literature, beginning with Gompers, Ishii, and Metrick (2003), has demonstrated a positive correlation between good corporate governance and stock prices and returns. In this paper, we provide a rationalisation for this correlation.

In this paper, we provide a unified framework that allows for the joint determination of three endogenous quantities: performance-pay sensitivity, corporate governance, and market quantities like stock and bond prices and credit yield spreads. We study a dynamic model of the firm, where levels of corporate governance are chosen optimally, equity prices reflect the true value of returns to shareholders, and the performance sensitivity of agents' compensation varies with both, and we show how all of these are jointly determined in consonance with the evolution of the balance sheet of the firm. Our base model generalises the continuous-time setting of DeMarzo and Sannikov (2006) (henceforth, DS) which, as shown by Biais et al. (2007) (henceforth, BMPR), can be viewed as the continuous-time limit of a discrete-time agency model. We consider a dynamic contracting environment where a risk-neutral entrepreneur (the *agent*) manages a risky technology. The entrepreneur is resource-constrained, and relies on investors (or the *principal*) to provide the initial start-up funds, and to absorb running losses, should they occur. The cash flow stream produced by the technology is noisy over time, where the noise process is modelled as a Brownian motion. The agency problem, ie, the source of discrepancy between the two parties' incentives, is that the entrepreneur can divert cash flow from the firm for his own private benefit.

In our model, the principal can *monitor* the output of the firm, which we view as *governance* because, at its core, governance is about mitigating the costs of the agency problem. The optimal contract specifies the agent's compensation structure, as well as the optimal amount of governance by the principal, over the life-cycle of the firm. Following BMPR, we write the optimal contract in terms of the firm's observable cash reserves, and relate these cash reserves to the firm's stock and bond price, and credit yield spread. The optimal contract induces the following capital structure. The entrepreneur and the investor each hold some fraction of the equity. The firm issues bonds (as debt) to the investors, and makes coupon payments that are proportional to the cash reserves. When cash reserves become sufficiently large, dividends are paid; the payment is such that cash reserves never go above a certain threshold. Moreover, investors can trade their equity holdings in a market, and the resulting stock price is a monotone function of the cash reserves, as in BMPR and consistent with Kaplan and Rauh's (2010) empirical findings, agent compensation is back-loaded and increasing in firm performance.

Besides characterising the optimal level of governance as cash flow changes, our main contribution is our comparative static result which shows that firms that are intrinsically more profitable (ie, provide greater returns under any contract and governance structure). Intuitively, such firms have a greater marginal return for investment in governance, and so profits, stock prices, and governance are correlated.[1] More specifically, we find:

(i) The level of governance is single-peaked in cash reserves and monotone in the volatility aversion of the firm.
(ii) Governance and pay sensitivity are substitutes, and so pay sensitivity is U-shaped as a function of cash reserves, and it is bounded above by the volatility of the stock price.
(iii) Firms that are intrinsically more profitable, or ones where the agent has a lower private benefit from misappropriation adopt higher governance standards and exhibit higher stock prices at every level of cash reserves. They also have a higher cash-reserve threshold for paying dividends, and have lower credit yield spreads.
(iv) A *measure* of governance can be obtained as a product of the following observables: the stock price, the local volatility of stock price, and the Delta of the agent's compensation, ie, the sensitivity of compensation to stock price.

Our findings (i) and (ii) are in line with the empirical results of Fahlenbrach (2009) who shows that the sensitivity of CEO wealth to firm performance is higher when the CEO is chairman and for firms whose boards are less independent, and also shows that wealth-performance sensitivities are decreasing in institutional ownership concentration and in the percentage of equity held by pension funds.[2] Relatedly, Fernandes et al. (2013) show that CEO pay is positively related with institutional ownership, which is often equated to better governance. The upper bound on the pay-performance sensitivity does not seem to be explored in the empirical literature. It is well known that the using publicly available data on stock options, the Dupire formula — Dupire (1994) — tells us how to compute the implied local volatility of the stock process. Our analysis tells us that this local volatility must always be higher than the pay sensitivity of the agent, as a function of cash reserves, or equivalently,

[1]This intuition is also to be found in Hermalin (2010).

[2]See also Dicks (2012) who shows that governance and sensitivity are substitutes, and suggests a role for regulation of corporate governance, once one takes into account the general equilibrium effects of increased corporate governance.

of stock price (which is a deterministic and monotone increasing function of cash reserves).

The result in (iii) is intuitive (though not trivial); intrinsically better firms spend relatively more on governance because the marginal returns are higher. More specifically, monitoring increases the time to (and reduces the probability of) termination, which is the fundamental (and only) inefficiency in the problem. Consequently, this results in higher stock returns, consistent with Gompers, Ishii, and Metrick (2003) and the subsequent literature. When agency problems are exacerbated, the firm becomes less valuable even though its intrinsic profitability is unchanged, as the agent must now be paid larger information rents, which is evidenced by the lower threshold on cash reserves that determine when dividends are paid. This, in turn, means that the principal is less inclined to monitor the agent, and also that stock prices are lower, because the expected dividend payments have decreased (due to the probability of liquidation increasing), once again consistent with the empirical observation that governance and stock prices are positively correlated. The delay in liquidation of the firm also leads to a lower credit yield spread.

Our comparative statics results in (iii) provides an explanation for the findings of Gompers, Ishii, and Metrick (2003) and Bebchuk, Cohen, and Ferrell (2009), for instance, by showing the sense in which corporate governance and stock prices and credit yields are positively correlated. Moreover, we demonstrate two different dimensions along which firms can be ranked, namely their intrinsic profitability and the extent of the agency problem. However, our results also show that ex ante identical firms can have different evolution of stock prices and different levels of corporate governance over time. It is the path of exogenous shocks that dictates the levels of governance and the change in stock prices over the life cycle of the firm. This provides an explanation for firm level heterogeneity in the data.

Our key contribution to the measurement of governance is point (iv) above, which shows that a scaled measure of governance can be written as a simple product of a number of observables. Two of these observables relate to the stock price, while the third – the sensitivity of compensation to stock price – is firmly in the field of corporate finance. This leads to the pleasing conclusion that to measure governance, one needs to use analyses and from both corporate finance as well as asset pricing.

We also make a number of methodological contributions that are critical for the derivation of our results. The solution to the principal's problem, which consists in deriving her value function in equilibrium, can be written as the solution to a free-boundary problem. It is believed that the problem does not permit an explicit solution; see, for instance, Biais et al. (2007, p.347). However, we exhibit a power series solution to the free boundary problem, and an analysis of this solution enables us to characterise the properties of the firm's aversion to volatility, as measured by the negative of the value function's second derivative, for various levels of cash reserves. In particular, we can show that principal' aversion to volatility is single-peaked and completely determines the optimal level of governance, thereby establishing point (i) above. Our comparative statics results amount to demonstrating how the firm's aversion to volatility changes with the model's various parameters. A second methodological contribution that our paper makes is the use of the Comparison Theorem for viscosity solutions of differential equations — see, for example, Crandall, Ishii, and Lions (1992) — as the basis for these comparative statics, thereby complementing the techniques introduced by DeMarzo and Sannikov (2006) and Biais et al. (2007).

We discuss the related literature in section 2. Section 3 lays out the model, while Section 4 discusses incentive-compatible contracts and a recursive formulation for the principal. Section 5 discusses the principal's problem, derives necessary conditions for optimality, and discusses properties of the value function. Our main results are in section 5.4, which characterizes the optimal monitoring strategy, and in section 6 which establishes comparative statics results. All the proofs are in the appendices. Appendix D discusses a discrete-time version of the problem, the basic intuition one can glean from such an exercise, and how the continuous-time model allows a much fuller characterization of the solution.

2. Related Literature

As mentioned in the introduction, our paper contributes to the literature in different ways. One, it complements the existing theoretical literature by providing it with a modeling approach to jointly characterize compensation contracts and corporate governance in a dynamic contracting framework. Two, this joint characterization can explain some of the empirical findings about how corporate governance interact with other, potentially endogenous, variables like compensation, stock price, profitability, volatility, and so on. In this section, therefore, we split our literature review between the theoretical and empirical literature.

Before we embark on this literature review, we note that Murphy (1999), Becht, Bolton, and Röell (2003), Adams, Hermalin, and Weisbach (2010), Kaplan (2012), and Edmans, Gabaix, and Jenter (2017) provide comprehensive reviews of various aspects of the theoretical and empirical literature on executive compensation and corporate governance. Also, Hermalin (2010, 2013) provides an insightful analysis of why governance may be related to performance measures, and indeed foresees many of the results in this paper.

2.1. Theoretical Literature

Our paper builds on the seminal analyses of DeMarzo and Sannikov (2006) and Biais et al. (2007), who initiate continuous-time methods in the dynamic-contracting analysis of the firm and its capital structure, thereby refining and extending the discrete-time framework of Clementi and Hopenhayn (2006), and DeMarzo and Fishman (2007a, 2007b). We add to this literature by allowing the firm to control the volatility of the Brownian noise, by further exploring the properties of the value function, and by exhibiting a power series solution for the defining HJB equation. Our analysis pays special attention to the value function's second derivative, which we show to be continuous and piecewise smooth but not differentiable everywhere, as its evolution is central to the tradeoff faced by the firm when it sets the volatility of the output process. Indeed, because this quantity effectively measures the firm's aversion to contractual contingencies that it commits to at the outset but that are not conditionally optimal, the firm finds it optimal to invest in volatility reduction when this second derivative is large. Finally, our use of the Comparison Theorem for viscosity solutions of differential equations complements the Feynman-Kac approach introduced by DeMarzo and Sannikov (2006) for comparative statics.

The idea that principals might want to gather more or more precise information about the whether their agents refrain from shirking or stealing has been the subject of several theoretical analyses. In static models, Baiman and Demski (1980), Dye (1986), and Demougin and Fluet (2001) investigate the possibility for the principal to add to the information contained in observable output by investing in costly signals about the agent's effort. In a similar vein, and more in line with our approach, Jost (1991), Milgrom and Roberts (1992), Strausz (1997), and Georgiadis and Szentes (2020) give the principal the ability to adjust the precision with which output is observed (ie, reduce the output's variance) and to adjust the agent's compensation contract accordingly. Li and Yang (2020) contribute to this static-model literature by allowing the principal to tailor the information she receives from the output through a flexible (but costly) partitioning of the state space.

In all these models, as in ours, the cost that the principal incurs to increase the quality of the feedback she gets from the output serves to reduce agency costs. That is, even though the principal knows the agent's action in equilibrium, the smaller noise reduces the agent's information rent, a point that Tirole (2006, p.341) also makes. However, in contrast to ours, the static nature of all these models leaves them silent about the joint evolution of contracts, monitoring, and stock prices. In fact, even the equilibrium contemporaneous relationship between compensation and monitoring is different. For example, in their Monitoring Intensity Principle, Milgrom and Roberts (1992, p.227) suggest that monitoring and pay-performance sensitivity are complements, but their intuition depends on the fact that, in their model, the agent's optimal action can vary with the level of monitoring.[3] In contrast, in our model and as in all dynamic-contracting problems with linear costs, the agent's optimal action is always the same, namely, not to steal. This results in monitoring and performance sensitivity being substitutes.[4]

Turning to dynamic analyses, Noe and Rebello (2012) study a discrete-time model in which the firm can jointly set the manager's compensation and the firm's governance. The latter is modelled as the fraction of capital that the firm can protect (at a cost) from diversion by the manager. They show that the firm loosens its governance and increases the expected manager's compensation after a string of positive performance shocks. Besides the fact that governance is modelled differently and that time is discrete, the result that governance is monotonic in the agent's continuation utility differs from ours. The reason for their result is that the firm and agent learn the firm's type (vulnerable or not) over time, and governance is only worthwhile when the firm is likely to be vulnerable (which is not the case after a sting of good results). In contrast, our stationary solution does not ascribe firm types; only the stochastic performance path of firms makes them heterogenous.

Closer to our model and to the stationary solution that we obtain are the dynamic-contracting models of Piskorski and Westerfield (2016), Orlov (2018), Varas, Marinovic, and Skrzypacz (2020), Chen, Sun, and Xiao (2020), and Dai, Wang, and Yang (2021) who all consider continuous-time settings in which the firm can, in addition to observing output, change the extent of the moral hazard problem with its agent. The main difference between these models and ours is the fact their monitoring technology is *retrospective* in that it is about past malfeasance by the agent, as opposed to being *prospective*, ie, set before the agent chooses his action, in our case. This distinction, first made by Holmström and Tirole (1993) and later amplified in Tirole (2006, p.334), means that the firm, instead of reacting to performance shocks by investigating them, implements an ex ante governance system ensuring that it will understand the source of these shocks as it experiences them. In terms of results, our paper is probably closest to Piskorski and Westerfield's (2016) in that the optimal monitoring structure they find is U-shaped in continuation utility. In contrast, they do not reach an explicit solution to their problem (complicated by the fact that it requires them to analyze a sticky Brownian motion process) nor do they characterize the evolution of the firm's aversion to volatility (second derivative

[3] In essence, their intuition is that, when the plan is to make the agent's pay very sensitive to performance, it will pay to measure that performance carefully.

[4] Roughly, Milgrom and Roberts (1992, p.218) consider linear wage contracts, and show that the optimal action is such that the marginal cost of effort is equal to the sensitivity of wages to performance. This relies on the agent having convex costs, and so changes in the intensity of monitoring affect the optimal action.

of the value function), which is central to analyzing monitoring structure. Finally, they do not have counterparts to our comparative statics results relating governance and stock prices.

In continuous time, it is also worth noting that Cadenillas, Cvitanić, and Zapatero (2004), Cvitanić, Possamaï, and Touzi (2017), Leung (2017), and Feng and Westerfield (2020) all study dynamic contracting problems in which the agent not only controls the firm's average productivity but also has some control over the volatility of the firm's output process. In essence, these models seek to capture the idea that, in some scenarios, agents can voluntarily and endogenously exacerbate the agency problems that exist between them and their principals; in other words, they can choose the extent to which their actions are camouflaged. In contrast, we study an environment where it is the principal who controls this volatility, thereby focusing on the idea that the principal can voluntarily and endogenously choose to alleviate such agency problems as part of the firm's optimal corporate governance.

2.2. Empirical Literature

Several empirical analyses have been mindful of the fact that compensation, governance, and firm performance are endogenously correlated and all depend on firm characteristics. For example, Demsetz and Lehn (1985b), Morck, Shleifer, and Vishny (1988), McConnell and Servaes (1990), and Himmelberg, Hubbard, and Palia (1999) all document that the heterogeneity in stock compensation across CEOs largely depends on firm heterogeneity. Palia (2001), in particular, confirms the theoretical idea that executive compensation is an endogenous equilibrium response to the contracting environment faced by the firm which, in addition to the firm's characteristics, might include governance.

Realizing that the relationship between governance and performance is dynamic, Wintoki, Linck, and Netter (2012) propose an empirical methodology that allows current governance to be influenced by past performance and vice versa. They find no statistically significant relation between contemporaneous firm performance and corporate governance (as measured by various aspects of board structure), but they also document that, after accounting for unobserved heterogeneity, simultaneity, and the effect of past board structure on firm characteristics, corporate governance is closely associated with firm size, growth opportunities, firm risk, age, leverage, and past performance. Following them, Balsam, Puthenpurackal, and Upadhyay (2016), Fauver et al. (2017), and Bhagat and Bolton (2018) document that better governance seems to have a positive impact on subsequent firm performance.[5] Finally,

[5]It is worth noting, however, that Cremers, Litov, and Sepe (2017) partially attribute this finding

Hoskisson, Castleton, and Withers (2009) and Conyon (2014) find that monitoring and compensation tend to be complements, and Wright and Kroll (2002) document a positive correlation between CEO compensation and corporate performance but only for firms that have vigilant external monitors.

Pay-performance sensitivity (PPS), as well as how it is affected over time by stock price movements and firm productivity, has also been heavily investigated. For example, Aggarwal and Samwick (1999) find that the PPS of executives decreases with stock price volatility, while Bulan, Sanyal, and Yan (2010) document an inverse U-shaped relationship between productivity and the sensitivity of CEO wealth to share value ("delta"). Particularly relevant to our dynamic contracting environment is the work of Boschen and Smith (1995) who find that the cumulative response to performance is roughly ten times that of the contemporaneous response. Gibbons and Murphy (1992) also document that the sensitivity of pay to performance is significantly greater for CEOs at the end of their careers, a finding that Cook and Burress (2013) attribute to the fact that long-tenured CEOs are subject to less monitoring than their shorter-tenured counterparts. In fact, more generally, Chang, Luo, and Sun (2011) find that monitoring and pay sensitivity tend to be positively correlated.

3. Model

Time is continuous, denoted by $t \in [0, \infty)$. There is a risk-neutral *principal* with deep pockets and a discount rate $r > 0$. A project needs funding, but the risk-neutral entrepreneur (*agent*) with discount rate $\gamma > r$ has limited liability and no wealth.[6] The principal, should she choose to fund the project with initial setup costs $K > 0$, will also cover any operating losses. For simplicity we assume the agent's outside option is 0 and that the project has no liquidation value.

The project produces a cumulative cash flow $Y_t \in \mathbb{R}$, where Y_t is given by

$$Y_t = \mu t + \sigma_t B_t$$

B_t is a standard Brownian motion, and $\sigma_t > 0$ is a process chosen by the principal. We interpret σ_t as the *governance environment* in which the agent operates, with

to low-value firms being more likely to adopt a staggered board, commonly thought to be detrimental to corporate governance.

[6]The assumption that $\gamma > r$ reflects the assumption that the intertemporal marginal rate of substitution for a wealth-constrained agent is greater that r. If $\gamma = r$, it is optimal for the principal to postpone consumption arbitrarily far into the future. An analysis of this case would then require, for instance, a finite horizon or bounds on how much utility can be promised to the agent. See DeMarzo and Sannikov (2006) and BMPR for further discussion of this point.

the understanding that a lower σ_t corresponds to stronger governance (or greater monitoring). This allows us to regard monitoring of the agent as the level of *corporate governance* in the firm. In what follows, we will take (corporate) governance and monitoring to mean the same thing.

The cash flow Y_t is not observable by the principal. Instead, the agent reports the process $(\hat{Y}_t)$ to the principal; $Y_t - \hat{Y}_t$ is the amount of output diverted by the agent for personal consumption. The benefit (to the agent) of diverting $Y_t - \hat{Y}_t$ (up to time t) is $\lambda(Y_t - \hat{Y}_t)$, where $\lambda \in (0, 1]$, ie, there may be some deadweight loss from the diversion. A larger λ naturally reflects a more severe agency problem.

The agent's cumulative compensation is denoted by C_t; limited liability requires this process to be non-decreasing. The agent also cannot save privately.

The principal chooses the monitoring intensity $\sigma_t \in \Sigma := \{\sigma_0, \dots, \sigma_n\}$ as a function of the history of reports $\{\hat{Y}_s : 0 \leq s \leq t\}$, where $\sigma_i > \sigma_{i+1}$ for $i = 0, \dots, n-1$. The principal's choice of monitoring σ_t entails a running cost per unit of time, given by $\rho : \Sigma \to \mathbb{R}$. We denote $\rho(\sigma_i)$ by ρ_i. We assume, without loss of generality, that $\rho(\sigma_0) = 0$, ie, the least amount of monitoring is costless, and that ρ is decreasing in σ, ie, $\rho_j < \rho_{j+1}$ for all $j = 0, \dots, n-1$. To see how the monitoring intensity relates to governance, it is useful to rewrite $\sigma_i = \sigma_0(1 - g_i)$, where $0 \leq g_i < g_{i+1} < 1$ for all $i = 0, \dots, n-1$, and a greater g corresponds to higher levels of governance. To ease notation, we will write $\sigma_t \in \Sigma$ as defined above, but will interpret governance in terms of the g_is.

Our assumption that monitoring amounts to reducing the variance of the output follows Milgrom and Roberts (1992). This is plausible, for instance, when cash flow comes from multiple sources, and monitoring amounts to observing some of these sources, thereby reducing overall uncertainty. We emphasize that our view of monitoring (governance) is different from auditing. Auditing is measurement of past managerial performance. It is what Holmström and Tirole (1993) refer to as *speculative information* and Tirole (2006, p.334) refers to as *retrospective* information. On the other hand, the governance that we have in mind corresponds to the firm investing in internal controls and structure that allow it to more precisely attribute production to its various factors. In a way, this measures the extent to which the firm is able to measure itself and to adjust to various shocks over time. Holmström and Tirole (1993) refer to this as the acquisition of *strategic* information, while Tirole (2006) calls this *active* monitoring.

As an alternative to our model of governance, where we control the noise of output, we could view governance as reducing λ, the agent's benefit from diverting output for personal consumption. That is, governance makes it harder for the agent to benefit from malfeasance or misappropriation. In Appendix E, we show that this view of governance, which does not affect the volatility of the output, is nonetheless isomorphic to the model describe above in terms of the optimal contract and its implementation via securities. We emphasise that regardless of how one models the specifics of governance, it is always concerned with the reduction of agency costs.

3.1. Contracts

The principal conditions his actions on *reports* made by the agent. During the operation of the firm, the agent reports the cash flow $\hat{Y}_t$.

A *contract* is a tuple $\Phi = \big(C = (C_t), \tau, \sigma = (\sigma_t)\big)$ that specifies, contingent on the report process $\hat{Y}$, the *cumulative payment* C_t made to the agent up to time t which is a non-decreasing process, the (stochastic) *termination time* τ, as well as the *monitoring intensity* $\sigma_t \in \Sigma$. The contract is contingent on the entire path of reported cash flows $(\hat{Y}_t)$. Note that any signal observed by the principal is also observed by the agent.

The principal offers the agent a contract at time $t = 0$, and fully commits to this contract. The agent can leave the contract at any time to an outside normalized to 0.

We assume that $\hat{Y}_t$ is continuous[7] and $\hat{Y}_t \leq Y_t$ (ie, the agent can never over-report cumulative output). This is reasonable because discontinuous reports or reports processes whose quadratic variation is different from that of (the unobserved) Y_t is certain evidence that the agent is lying, and will be punished immediately.

A contract Φ is *incentive-compatible* if, given Φ, the agent's optimal reporting strategy is $\hat{Y} = Y$, ie, the agent finds it optimal to report the output process truthfully. It is clear that the inefficiency that arises from the Agency problem is in the termination of the project. Indeed, the first-best, full information solution is to run the project forever, while paying the agent whatever he is owed right away.

3.2. Payoffs and Principal's Problem

As noted above, both principal and agent are risk-neutral. Let $\hat{w}(\hat{Y};\Phi)$ be the agent's utility from choosing the reporting strategy $\hat{Y}$ under the contract Φ. His utility when choosing an *optimal* reporting strategy is

$$w(\Phi) = \max_{\hat{Y}} \hat{w}(\hat{Y};\Phi) = \max_{\hat{Y}} \mathsf{E}^{\hat{Y},\sigma}\left[\int_0^\tau e^{-\gamma t}\big(\mathrm{d}C_t + \lambda(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t)\big)\right]$$

[7] Thus, the process $(\hat{Y}_t)$ has the same quadratic variation as (Y_t), namely $\int_0^t \sigma_s^2\,\mathrm{d}s$, and the drift of $\hat{Y}$, like that of Y, is absolutely continuous with respect to Lebesgue measure. These are consequences of Girsanov's Theorem.

The contract Φ is incentive-compatible if $w(\Phi) = \hat{w}(Y, \Phi)$, ie, if *truth-telling* with $\hat{Y} = Y$ is optimal for the agent. The principal's profit is given by

$$\mathsf{E}^{\hat{Y},\sigma}\left[\int_0^\tau e^{-rt}\big[\,\mathrm{d}\hat{Y}_t - \rho(\sigma_t)\,\mathrm{d}t - \mathrm{d}C_t\big]\right]$$

when the agent chooses $\hat{Y}$. Observe that the choice of monitoring strategy $\sigma = (\sigma_t)$ only affects the volatility of the driving uncertainty, B_t, and hence doesn't affect its expectation. Nonetheless, it matters crucially because it affects the agent's reporting strategy.

Given an initial amount of utility promised to the agent, say W_0, the principal's Contracting Problem is

[3.1] $$F(W_0) := \max_{\Phi} \mathsf{E}\left[\int_0^\tau e^{-rt}\big[\,\mathrm{d}Y_t - \mathrm{d}C_t - \rho(\sigma_t)\,\mathrm{d}t\big]\right]$$

subject to the constraint that Φ is incentive-compatible and to the promise-keeping constraint $w(Y;\Phi) = W_0$.

4. Incentive-Compatible Contracts

Consider a contract $\Phi = (C, \tau, \sigma)$ that conditions on the reporting process $\hat{Y}$ by the agent. The agent maximizes his utility given the contract. As noted above, the contract Φ is incentive-compatible if the agent's optimal reporting strategy is to report truthfully.

It is clear from the Revelation Principle that any contract $\Phi = (C, \tau, \sigma)$ in which the agent decides to use a diversion strategy $\Delta = (\Delta_t)$ (so that $Y_t - \hat{Y}_t = \int_0^t \Delta_s\,\mathrm{d}s$) is payoff equivalent to one where the principal just increases consumption to $C'_t = C_t + \lambda\Delta_t$, and where the agent does not divert any cash flows. Thus, we have the following result.

Lemma 4.1. Given a contract $\Phi = (C, \tau, \sigma)$ where the agent optimally reports $\hat{Y} \neq Y$, there exists another, incentive-compatible contract $\Phi' = (C', \tau', \sigma')$ where the agent reports truthfully, and leaves both principal and agent at least as well off in payoff terms.

Intuitively, because the agent cannot save, and because he discounts the future at rate γ, any diversion can be simulated by the principal and deferred to a later date by compounding at the rate γ.

A formal proof of this assertion, in a slightly more general form can be found in DeMarzo and Sannikov (2006, Lemma 1).[8] With the observation that it is without loss of generality to consider contracts that optimally induce zero cash-flow diversion, we now proceed to a characterization of incentive-compatible contracts.

A contract entails full commitment on the part of the principal. To proceed, as in discrete time principal-agent models, it is useful to understand the evolution of continuation utility. Fix a contract $\Phi = (C, \tau, \sigma)$ and a reporting strategy $\hat{Y}$, and let W_t be the expected utility from time t onwards. Then W_t is given by

$$W_t = \mathsf{E}_t^{\hat{Y},\sigma}\left[\int_t^\tau e^{-\gamma(s-t)}\left[\,\mathrm{d}C_s + \lambda(\,\mathrm{d}Y_s - \mathrm{d}\hat{Y}_s)\right]\right] \tag{4.1}$$

While the promised utility in [4.1] is entirely forward looking, W_t can, in fact, be written as a diffusion process, whereby increments of promised utility depend only on the current report, current output, and the exogenous noise. This is the central insight of Sannikov (2008), and it greatly facilitates further analysis of incentive compatibility and the optimal contract. The next lemma makes this precise.

Lemma 4.2. Let (W_t) be as in [4.1] and fix a reporting strategy $\hat{Y}$. Then there exists a ($\hat{Y}$-measurable) process $Z = (Z_t)$ such that

$$\mathrm{d}W_t = \gamma W_t\,\mathrm{d}t - \mathrm{d}C_t - \lambda(\,\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t) + Z_t\sigma_t^{-1}\underbrace{\left[\,\mathrm{d}Y_t - \mu\,\mathrm{d}t\right]}_{=\sigma_t\,\mathrm{d}B_t} \tag{4.2}$$

It is useful to think of promised utility as a stock that grows at the rate γ. Thus, the increment to promised utility, $\mathrm{d}W_t$ is the interest paid on the stock W_t, net of the payment from the principal, and the amount stolen (which comes from the reported output). The process Z_t is the *sensitivity* of the increment $\mathrm{d}W_t$ to the noise term $\mathrm{d}B_t$, which is output net of the known drift. The proof of the lemma is found in Appendix A. Our more general setting where σ_t is also a $\hat{Y}$-measurable process poses no additional difficulties.

Notice that W_t in [4.1] and [4.2] is how the agent perceives his promised utility. The principal cannot see the true process Y_t, and so requires that the agent report truthfully, ie, that the contract be incentive-compatible.

Given the diffusion representation of promised utility in [4.1], it is now relatively straightforward to characterize incentive-compatible contracts, which is another key insight from Sannikov's (2008) work.

Lemma 4.3. Truth telling, ie $\hat{Y}_t = Y_t$, is incentive-compatible if and only if $Z_t \geq \lambda\sigma_t$ for all $t \leq \tau$.

[8] Their proof does not rely on the fact that σ is constant, and so is also valid in our setting.

The intuition behind this characterization is exactly as in DeMarzo and Sannikov (2006). The benefit to diversion is $\lambda(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t)$, while the cost, as seen from [4.2] is $Z_t\sigma_t^{-1}(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t)$, because $\mathrm{d}Y_t - \mu\,\mathrm{d}t = (\mathrm{d}\hat{Y}_t - \mu\,\mathrm{d}t) + \mathrm{d}(Y_t - \hat{Y}_t)$. Incentive compatibility is therefore the condition that the costs of misreporting are greater than the benefits.[9]

The preceding lemma suggests that as far as the agent is concerned, all that matters is promised utility, and how it evolves, given C, Z, and σ. Thus, it makes sense to consider *recursive* contracts that are contingent on W (which is controlled by the agent via $\hat{Y}$), and contain contractual elements C_t, Z_t, σ_t, and τ that are *deterministic* functions of W_t. The next lemma tells us that this is justified.

Lemma 4.4 (Contract Generation). Let C_t, Z_t, σ_t, and τ be deterministic functions of W_t where $\mathrm{d}W_t$ is as in [4.2]. If $Z_t \geq \lambda\sigma_t$, then the contract (C, τ, σ) is incentive-compatible.

Proof. In the Appendix. □

This allows us to formulate the principal's problem recursively, as we do next.

5. Optimal Contracts

We now use the dynamic programming principle to derive the principal's optimal contract. Lemma 4.4 says that instead of contracts that depend on the entire path of reported output $\hat{Y}$, we may restrict attention to recursive contracts that are Markovian in promised utility W_t, which in turn is controlled by the agent's reports. Clearly, the agent's outside option of 0 dictates that $W_t \geq 0$ for all t. Let $F(w) \mapsto \mathbb{R}_+ \to \mathbb{R}$ denote the principal's value function, which is the largest profit the principal can obtain from all recursive contracts that provide the agent with $w \geq 0$ utiles.

In what follows, we first assume that F is concave and twice differentiable, and that there exists an optimal incentive-compatible contract. This allows us to derive necessary conditions for optimality, and also characterize the contract. In the Appendix, we show in a verification theorem that any function satisfying the necessary conditions (and transversality) is, in fact, the value function, and moreover, there is a concave and smooth (twice-differentiable) function satisfying the necessary conditions.

[9] If the principal cannot control σ_t, ie, his monitoring intensity is constant over time, then we recover the characterization provided by DeMarzo and Sannikov (2006), namely that $\beta_t \geq \lambda$ where $\beta_t := Z_t\sigma^{-1}$, where β_t is the sensitivity in DeMarzo and Sannikov (2006).

5.1. Payment, Termination, and Value Function

We first consider the optimal structure of payment and termination. Consider, first, the case when the agent is promised w utiles. The principal has three instruments to control the agent's utility process $\mathrm{d}W_t$, as seen in [4.2], namely Z_t, C_t, and σ_t. The principal can always pay a lump sum of δC to the agent, and re-start the contract at $w - \delta C$. But concavity of the value function F implies that $F(w) \geq F(w - \delta C) - \delta C$. That is, paying a lump-sum of δC and re-starting the contract can never be strictly preferred. This implies that we must necessarily have $F'(w) \geq -1$ for all $w \geq 0$. Intuitively, the marginal cost of increasing the agent's utility by a dollar can never be greater than 1, which is achieved by simply giving him an extra dollar.

Let $w^\star := \inf\{w : F'(w) = -1\}$ be the smallest level of promised utility such that the principal is indifferent between compensating the agent via payment or promises. Our assumption that F is concave and twice differentiable ensures that $w^\star$ is well defined. Because the agent discounts the future faster than the principal, it is optimal to pay the agent whenever $W_t \geq w^\star$. This is exactly as in DeMarzo and Sannikov (2006), in spite of the additional instruments available to the principal. The intuition behind this result is that because $F'(w) \geq -1$, the principal wants to backload payments (which arise as the agent's information rents) as much as possible.[10] This allows the principal to use promised utility as a stock of carrots which she can add to when performance is good, and deplete when performance is bad, and pay the agent when the stock of carrots is sufficiently high.[11]

It is easy to see that if the agent is promised 0 utiles, it is (strictly) optimal for the principal to liquidate the firm immediately. This is because running the project without the agent diverting all the output entails giving up some information rents, but doing so would give a positive amount, strictly more than 0. Thus, we must have $F(0) = 0$.[12] Thus, it is optimal to terminate the agent when $W_t = 0$, ie, define the random termination time τ as $\tau := \inf\{t : W_t = 0\}$.

In sum, the optimal contract should pay the agent when $W_t \geq w^\star$, and terminate him when $W_t = 0$. All that remains is to characterize the optimal contract on the interval $(0, w^\star)$. Here, the necessary condition for optimality is the HJB equation

[10] This property is also seen in the discrete time literature on dynamic contracting; see, for instance, DeMarzo and Fishman (2007b, 2007a), Clementi and Hopenhayn (2006) for cash-flow diversion models and Krishna, Lopomo, and Taylor (2013) for a dynamic procurement model that features similar backloading of information rents.

[11] Technically, the payment here is more complicated than in discrete time. The payment process is designed to ensure that $W_t \leq w^\star$; formally $C = (C_t)$ is a singular process because it is not absolutely continuous with respect to Lebesgue measure.

[12] For simplicity, we ignore the possibility of the project having a scrap value for both principal and agent, as is the case in DeMarzo and Sannikov (2006).

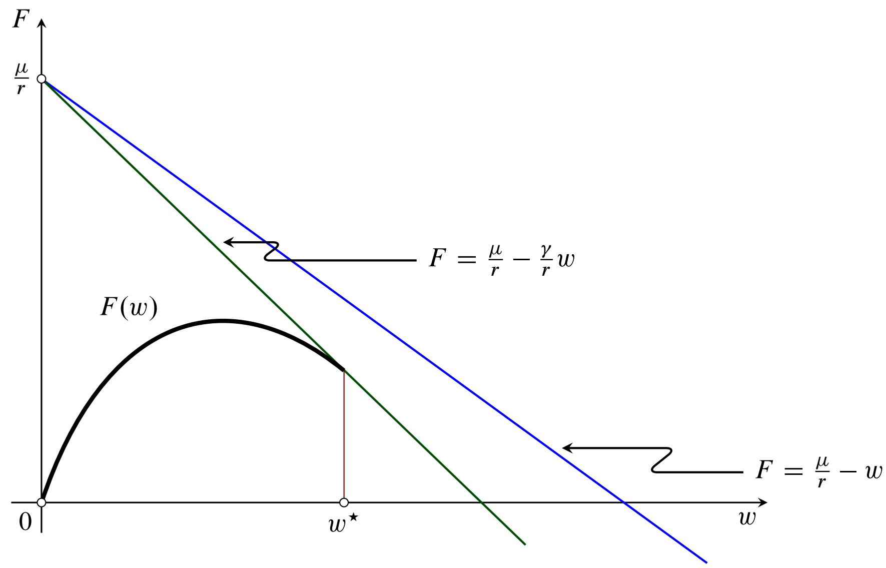

Figure 1: Value function F. The line $F = \mu/r - w$ is the full information payoff for the principal, where B_t is observed at no cost, or equivalently, σ can be set to 0 for free.

which can be written as

$$rF(w) = \mu + \gamma w F'(w) + \max_{z \geq \lambda\sigma, \sigma \in \Sigma} \left[\tfrac{1}{2} z^2 F''(w) - \rho(\sigma)\right] \quad [5.1]$$

The concavity of F implies $F''(w) \leq 0$, so that it is optimal to set $z = \lambda\sigma$ in the HJB equation [5.1]. This results in the reduced HJB equation (with an optimization over the level of monitoring σ):

$$rF(w) = \mu + \gamma w F'(w) + \max_{\sigma \in \Sigma} \left[\tfrac{1}{2} \lambda^2 \sigma^2 F''(w) - \rho(\sigma)\right] \quad [5.2]$$

This is a free-boundary problem because $w^\star$ is yet to be determined. The boundary conditions are $F(0) = 0$, $F'(w^\star) = -1$ which is a smooth pasting condition, and the *super contact* condition $F''(w^\star) = 0$.[13] We summarize our discussion in the following proposition.

Proposition 5.1. The optimal profit-maximizing and incentive-compatible contract

[13] The condition $F''(w^\star) = 0$ ensures that F is maximal among the class of functions satisfying [5.2] and the boundary conditions $F(0) = 0$ and $F'(w^\star) = -1$. See Dumas (1991) for a discussion of such conditions in singular control problems.

that delivers $w_0 \in [0, w^\star]$ to the agent requires that W_t evolve as

$$[5.3] \qquad dW_t = \gamma W_t\, dt - dC_t + \lambda(\underbrace{d\hat{Y}_t - \mu\, dt}_{\sigma_t\, dB_t})$$

$$[5.4] \qquad W_t \le w^\star$$

$$[5.5] \qquad C_t = \int_0^t \mathbf{1}(W_s = w^\star)\, dC_s$$

for all $t \in [0, \tau]$, where $\tau = \inf\{t \ge 0 : W_t = 0\} < \infty$ a.s. is the termination time, and $W_t = 0$ for $t \ge \tau$. The payment process C_t is nondecreasing in time, and payments are made only when W_t hits $w^\star$. If $W_0 > w^\star$, and immediate payment of $W_0 - w^\star$ is made to the agent. The principal's profit is given by the function $F(w)$ defined in [3.1], which is concave, satisfies the HJB equation [5.2] on $[0, w^\star]$, and the boundary conditions

$$[5.6] \qquad F(0) = 0, \qquad F'(w^\star) = -1, \qquad F''(w^\star) = 0$$

determine $w^\star$.

Condition [5.5] is a *flat-off* condition, which requires that C_s increase only when W_s hits the payment boundary $w^\star$, and can equivalently be written as $\int_0^t \mathbf{1}(W_s < w^\star)\, dC_s = 0$ for all $t \ge 0$. The constraint $W_t \le w^\star$ in [5.4] requires that W_t is reflected at $w^\star$, while [5.3] describes the evolution of W_t.

In the appendix, we prove a verification theorem establishing that any smooth and concave function that satisfies the HJB equation [5.2] and the boundary conditions is indeed the value function. We also show that such a smooth and concave function actually exists. Our verification and existence theorems are complicated, relative to DeMarzo and Sannikov (2006), by the fact that we have the additional monitoring controls to contend with, which makes it much more difficult to establish existence and the smoothness of the value function. That $\tau < \infty$ a.s. follows by adapting arguments in Ward and Glynn (2003).

To completely characterize the optimal contract, we need to describe the optimal choice σ_t. We do this in Section 5.4. But first, we follow Biais et al. (2007) and describe how the optimal contract can be implemented via financial securities and claims on cash flows, and how the optimal contract influences security prices.

5.2. Optimal Contract via Securities

Let M_t denote the firm's *observable cash reserves*. These reserves grow at interest rate r and depend on the firm's output process. The agent can divert the firm's cash flow, but at the optimum, because of incentive compatibility, chooses not to.

Because M_t is observable and contractible, financial contracts can be written as a function of firm's cash reserves. Defining $M_t = W_t/\lambda$ gives us the dynamics of M_t. The implementation requires *stocks* and *bonds*. The agent is given a non-tradeable fraction λ of the stocks. Stocks pay out when cash reserves M_t hit $m^\star := w^\star/\lambda$. The total dividend is $\lambda^{-1}\,\mathrm{d}C$: the agent gets $\mathrm{d}C$, while the financiers get $\mathrm{d}P = \frac{1-\lambda}{\lambda}\,\mathrm{d}C$. The other security is bonds. These distribute a continuous coupon flow of $\mu - (\gamma - r)M_t$ at time t that varies with the level of cash reserves.[14] The financiers of the firm (which could be a single principal, or multiple lenders) hold all the bonds.

The contract works as follows. First the agent gets a fraction λ of the firm's outstanding stock, while the financiers get a fraction $1-\lambda$. The agent cannot trade his stocks.[15] When cash reserves hit $m^\star$, dividend payments are made. Continuous coupon payments of $\mu - (\gamma - r)M$ are made to the financier, with the firm being terminated when $M_t = 0$, ie, when it runs out of cash. Notice that the agent can steal output that goes into cash reserves, but chooses not to because the contract is incentive-compatible.

We can write the SDE for M_t as

[5.7]
$$\mathrm{d}M_t = \gamma M_t\,\mathrm{d}t + \big(\underbrace{\mathrm{d}\hat{Y}_t - \mu\,\mathrm{d}t}_{\sigma(W_t)\,\mathrm{d}B_t}\big) - \lambda^{-1}\,\mathrm{d}C_t$$

or alternatively, as

[5.8]
$$\mathrm{d}M_t = (rM_t + \mu)\,\mathrm{d}t + \sigma(\lambda M_t)\,\mathrm{d}B_t - \mathrm{d}C_t - \mathrm{d}P_t$$

where $\mathrm{d}P_t$ is the payment to the financiers. It follows from [5.7] and [5.8] that

$$\mathrm{d}P_t = \big[\mu - (\gamma - r)M_t\big]\,\mathrm{d}t + \frac{1-\lambda}{\lambda}\,\mathrm{d}C_t$$

and the value for investors, net of monitoring costs, is U_t, where

$$\mathrm{d}U_t = \mathrm{d}P_t - \rho\big(\sigma(\lambda M_t)\big)\,\mathrm{d}t$$

[14] Because we insist that this be positive, it must be that $\mu\lambda \geq (r-\gamma)w^\star$. The implementation in DeMarzo and Sannikov (2006) does not suffer from this constraint. They use a *compensating balance*, which is a stock of cash, to account for negative cash flows.

[15] The agent is the holder of *unregistered* or *letter* securities that cannot be publicly traded. This is necessary because $\gamma > r$, which means that the market always values the stock more than the agent does. Allowing the agent to trade his stock will result in the agent trading his stock right away. Notice that if $\gamma = r$, this is no longer an issue, although we would then have to place an upper bound on the payment boundary.

All the processes above are adapted to W_t, and are thus deterministic functions of W_t (or equivalently, M_t). For all $t \in [0, \tau]$, the stock price S_t satisfies

$$S_t = \mathsf{E}_t\left[\int_t^\tau e^{-r(s-t)}\lambda^{-1}\,\mathrm{d}C_s\right] \tag{5.9}$$

Because C_t is a deterministic function of W_t, it follows that we can write $S_t = \mathcal{S}(M_t)$. Furthermore, we have the following result.

Proposition 5.2. The stock price S_t is given by $S = \mathcal{S}(M)$, where $\mathcal{S}$ is a solution to the boundary value problem

$$\begin{aligned} r\mathcal{S}(M) &= \gamma M\mathcal{S}'(M) + \tfrac{1}{2}\sigma^2(\lambda M)\mathcal{S}''(M) \\ \mathcal{S}(0) &= 0 \\ \mathcal{S}'(m^\star) &= 1 \end{aligned} \tag{5.10}$$

and $M \in [0, m^\star]$. Moreover, the stock price is a strictly increasing and strictly concave function of M, the level of cash reserves.

To get some intuition for this result, consider an Itô expansion of $e^{-rt}M_t$, whereby

$$\mathrm{d}(e^{-rt}M_t) = -re^{-rt}M_t\mathrm{d}t + e^{-rt}\big(\underbrace{\gamma M_t\,\mathrm{d}t + \sigma(\lambda M_t)\,\mathrm{d}B_t - \lambda^{-1}\,\mathrm{d}C_t}_{=\,\mathrm{d}M_t \text{ from [5.7]}}\big)$$

Integrating from t to $T \wedge \tau$, taking expectations, and then letting $T \to \infty$, we obtain

$$S_t = \mathcal{S}(M_t) = \mathsf{E}_t\left[\int_t^\tau e^{-rs}\lambda^{-1}\,\mathrm{d}C_s\right] = e^{-rt}M_t + \mathsf{E}_t\left[\int_t^\tau e^{-r(s-t)}(\gamma - r)M_s\,\mathrm{d}s\right]$$

where the first equality is from [5.9], and we use the fact that $M_\tau = 0$.

Observe that increasing M_t increases the time to termination, $M_\tau = 0$. Thus, $\mathcal{S}'(M) > 0$, so that stock prices increases in cash reserves. However, once $M_t = m^\star$, dividends are paid out, so the marginal value of the stock is exactly the value of the revenue generated, which is exactly 1.

Finally, notice that, because $\mathcal{S}$ is a strictly increasing function, we can write

$\mathcal{S}'(M_t) = \mathcal{S}'(\mathcal{S}^{-1}(S_t))$ and so on. Using Itô's Lemma, we obtain

$$\begin{aligned} \mathrm{d}S_t &= \underbrace{\left[\gamma M_t \mathcal{S}'(M_t) + \tfrac{1}{2}\sigma^2(\lambda M_t)\mathcal{S}''(M_t)\right]}_{= r\mathcal{S}(M_t) = rS_t} \mathrm{d}t \\ &\quad + \mathcal{S}'(M_t)\left[\sigma(\lambda M_t)\,\mathrm{d}B_t - \lambda^{-1}\,\mathrm{d}C_t\right] \\ &= rS_t\,\mathrm{d}t + \underbrace{\mathcal{S}'(\mathcal{S}^{-1}(S_t)) \cdot \sigma(\lambda\mathcal{S}^{-1}(S_t))}_{=: \mathsf{V}_t S_t}\,\mathrm{d}B_t - \lambda^{-1}\,\mathrm{d}C_t \end{aligned} \tag{5.11}$$

where we have used the BVP characterization of $\mathcal{S}$ from [5.10] which requires that $\mathrm{d}C_t = 0$ if $M_t < m^\star$, and also noting (from [5.10]) that $\mathcal{S}'(m^\star) = 1$, which is an analog of equation (57) in BMPR.

The volatility of the stock price S_t is $\mathsf{V}_t = \mathcal{S}'(\mathcal{S}^{-1}(S_t))\sigma(\lambda\mathcal{S}^{-1}(S_t))/S_t$. Because $\mathcal{S}$ is concave, it follows that $\mathcal{S}'(\cdot) \geq 1$, with equality only at $S_t = \mathcal{S}(m^\star)$. Therefore, stock prices are always more volatile than the output process, regardless of the amount of monitoring, and are also more volatile than the sensitivity of the contract, ie,

$$\mathsf{V}_t = \mathcal{S}'(\mathcal{S}^{-1}(S_t))\sigma(\lambda\mathcal{S}^{-1}(S_t)) \geq \sigma(\lambda\mathcal{S}^{-1}(S_t)) \geq Z_t$$

for all t. The last inequality holds (strictly) because at the optimum, sensitivity $Z_t = \lambda\sigma(\lambda\mathcal{S}^{-1}(S_t))$ (see [5.1] and [5.2]), and $\lambda \in (0, 1]$ by assumption.

The price of the bond is denoted by D_t and is given by

$$D_t = \mathsf{E}_t\left[\int_t^\tau e^{-r(s-t)}[\mu - (\gamma - r)M_s]\,\mathrm{d}s\right]$$

The price D_t can also be written as a deterministic function of M_t, as $D_t := \mathcal{D}(M_t)$. We can therefore write a boundary value problem that $\mathcal{D}$ must solve, as in BMPR. More importantly, we have the following result.

Proposition 5.3. At any date $t \geq 0$, we have

$$(1-\lambda)S_t + D_t = F(\lambda M_t) + M_t + \mathsf{E}_t\left[\int_t^\tau e^{-r(s-t)}\rho(\sigma_s)\,\mathrm{d}s\right] \tag{5.12}$$

In particular, if there is ever nontrivial monitoring, whereby $\mathsf{E}_t\left[\int_t^\tau e^{-r(s-t)}\rho(\sigma_s)\,\mathrm{d}s\right] \neq 0$, then $(1-\lambda)S_t + D_t > F(\lambda M_t) + M_t$.

The left-hand side of [5.12] is the market value of the securities held by the financiers. The right-hand side corresponds to the value of the assets generating these cash flows, plus the cost of monitoring, which is an input in the production process, that must be borne by the financiers. These financiers can often be blockholders, as suggested by Jensen (1989). Proposition 5.3 says that the market value of asset holdings is greater than the 'true' value, because the former ignores the cost of monitoring the agent. It is a generalisation of Proposition 6 in BMPR, in that the accounting identity now also includes the cost of monitoring the agent, which can be viewed as an additional investment made by some subset of the financiers that is necessary for the optimal evolution of the firm. Put differently, for large shareholders or for bondholders (e.g., blockholders), the market value of the securities they hold is an upper bound on the value they derive by providing capital, because it does not include the cost of monitoring the agent.

Following BMPR, we may also define the credit yield spread ζ on a consol bond that pays \$1 until the firm is dissolved. For each $t \in [0, \tau)$, ζ_t is given by $\int_t^\infty \exp\big(-(r+\Delta_t)(s-t)\big)\,\mathrm{d}s = \mathsf{E}_t\big[\int_t^\tau e^{-r(s-t)}\,\mathrm{d}s\big]$, which implies

$$\zeta_t := \frac{rT_t}{1-T_t}$$

where $T_t := \mathsf{E}_t[e^{-r\tau}]$. It is clear that T_t can be written as $\mathscr{T}(M_t)$ where $\mathscr{T}$ is a deterministic function. Appendix B.1 shows that $\mathscr{T}$ is the solution to a boundary value problem. This leads to the following proposition.

Proposition 5.4. The credit yield spread is a strictly positive, strictly decreasing, and strictly convex function of cash reserves.

5.4. Optimal Governance

To understand the optimal choice of σ, recall the HJB equation [5.2], which is a necessary condition for optimality. The optimal $\sigma(w)$ solves $\max_{\sigma\in\Sigma}\big[\frac{1}{2}\lambda^2\sigma^2 F''(w) - \rho(\sigma)\big]$. Because $F'' \leq 0$, the optimal choice of σ is monotone increasing in F''. In other words, the higher the firm's aversion to volatility, namely $-F''$, the greater the amount of governance (or monitoring).[16]

All of this is for a fixed w, and is a straightforward consequence of the HJB equation [5.2]. Our main methodological contribution in this paper is a description of how $F''(w)$ varies with w. This is a non-trivial exercise because F'' is itself non-differentiable at points where there is a switch in σ.

Theorem 1. *The firm's aversion to volatility, $-F''$, is single peaked in promised utility w (and cash reserves m, as well as in stock price), and optimal governance is monotone in aversion to volatility. Optimal monitoring is monotone in the firm's*

[16]The quantity $-F''(w)$ measures the loss in expected returns from a mean-zero lottery with small variance, relative to the mean w. Clearly, locally risk-neutral firms, with $F''(w) = 0$, do not suffer any such loss, but greater losses come when $-F''$ is larger.

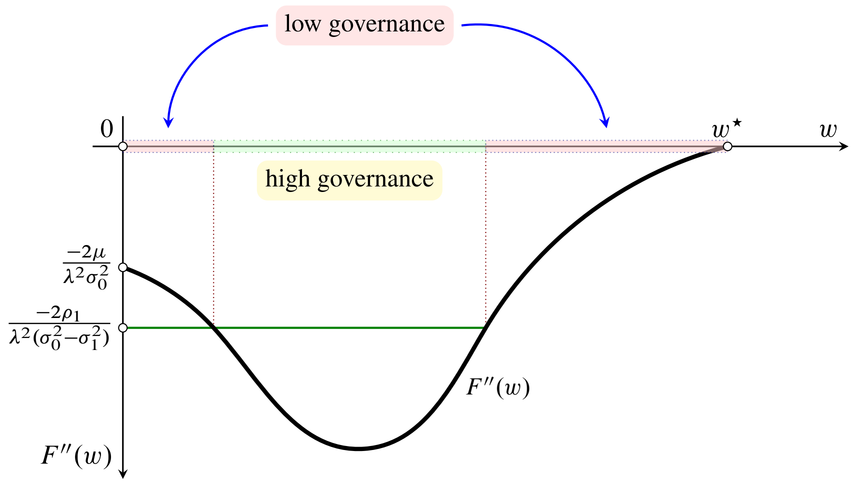

Figure 2: The shape of F''.

aversion to volatility $-F''$, and hence is single peaked in w (and cash reserves m, as well as in stock price).

In other words, optimal monitoring, starting at $w = 0$ promised utiles, initially (weakly) increases in cash reserves, reaches a maximum level of monitoring (which may still entail no monitoring), and then decreases as w increases, until there is minimal monitoring in a neighbourhood of the payment boundary $w^\star$.

Indeed, there exists an interval of promised utility with maximal governance (though not necessarily maximum possible governance). Governance is *decreasing* in promised utility beyond (to the right of) this interval, governance is *increasing* in promised utility up to (ie, to the left of) this interval, and there is minimal governance in a neighbourhood of the payment boundary $w^\star$.

Most importantly, there are regions (intervals) of promised utility wherein a certain level of governance is optimal, thereby defining *regimes* of corporate governance. And regardless of the parameters, in a neighbourhood of the payment boundary $w^\star$, the optimal regime entails the lowest possible amount of governance.

The result depends crucially on the shape of F''. To understand why $-F''$ must be single peaked and decrease to 0, notice that $-F''$ is the marginal value of decreasing the volatility $\lambda\sigma_t$ of the promised utility process. The principal would like a lower volatility because this extends the expected discounted time to liquidation,[17] ie, postpones the inefficient liquidation to a later date, which is valuable because it extends the time that she can benefit from the cash flows of the firm.[18] Where does a small and temporary reduction to σ_t have the greatest effect?

[17] See Proposition C.16 for a precise statement.

[18] Recall that the first best solution is for the principal to never liquidate the firm.

When w is near $w^\star$, the principal knows that the agent is about to get paid, and so a small reduction in volatility is not worth a lot; indeed, at $w = w^\star$, it is worth exactly 0, because it affects neither the expected time to liquidation nor the payment to the agent. Similarly, when w is near 0, the marginal value of reducing volatility is small, though not zero, because reducing volatility increases the time to liquidation, which is valuable, but not by a lot given the proximity to the absorbing boundary $w = 0$. It is for intermediate values of w that the value of reducing volatility is highest, because it is here that such a reduction will have its greatest impact, in terms of delaying liquidation.

These properties of the aversion to volatility $-F''$ hold for any number of levels of governance $\sigma \in \Sigma$, even though Figure 2 only considers two levels of governance.

6. Comparative Statics

We consider two main variations in parameters. First, we consider the impact of a change in μ, the intrinsic profitability of the firm, and next we consider the impact of a change in λ, the severity of the agency problem.

6.1. As Firms get Better …

A firm with a greater μ is *intrinsically* more profitable, in the sense that any policy for a lower level of μ will deliver a higher expected profit for a higher level of μ. Similarly, a firm with a lower λ has *less severe agency problem* because it can adopt the optimal contract of a firm with a higher λ, and still make the same profit, or optimally choose another contract, with a lower sensitivity-to-output (Z_t), which generates higher profit. Thus, firms with a higher λ or lower μ are *better* in that they can generate greater expected profit. Our first result justifies this intuition. Before stating the result, a useful change of variables is in order. Let $\hat{F}(m) = F(\lambda m)$ for all $m \in [0, m^\star]$, where $m^\star = w^\star/\lambda$. This allows us to state results in terms of cash reserves.

Proposition 6.1. Let $\hat{F}(m, \theta)$ be the value function for $\theta = \mu, \lambda$, as a function of cash reserves. Then, for a fixed $m > 0$, $\hat{F}$ is strictly increasing in μ and strictly decreasing in λ. Moreover, m^* is strictly increasing in both μ and λ.

As noted above, the monotonicity of $\hat{F}$ is intuitive, and it is straightforward to prove monotonicity in μ, but somewhat less so to establish monotonicity in λ, which requires the use of a Comparison Theorem for the differential equation for $\hat{F}$. Monotonicity of the of the cash reserve threshold follows from that of $\hat{F}$. These proofs are in Appendix C.

We now state our main comparative statics results.

Theorem 2. *Consider the firm's policies as a function of μ or λ. Then, for any level of cash reserves, the following hold:*

(a) *The level of governance (weakly) increases in μ and decreases in λ.*
(b) *Sensitivity of pay to output (weakly) decreases under μ.*
(c) *Stock prices are increasing in μ and decreasing in λ.*
(d) *Credit yield spread is decreasing in μ and increasing in λ.*
(e) *When μ is sufficiently low, there is no monitoring near the origin or anywhere else. For sufficiently high μ, there is maximal monitoring even at the origin.*

Theorem 2 exhibits one mechanism that rationalizes the findings of Gompers, Ishii, and Metrick (2003), Bebchuk, Cohen, and Ferrell (2009), and others, who find a positive correlation between stock prices and governance. The correlation between credit yield spreads and corporate governance is documented in Bhojraj and Sengupta (2003) and Ashbaugh-Skaife, Collins, and LaFond (2008).

Our analysis also shows that for 'better' firms, corporate governance is naturally higher, because the marginal returns from monitoring are higher. Stock prices are higher, because with increased monitoring, the probability of termination is lower. Finally, observe that stock prices are higher in spite of the fact that the threshold for payment, $w^\star$, is increasing in μ. That is, stock holders have to wait longer to get paid, but in spite of this, stock prices are higher, because the fear of termination is correspondingly lower.

Theorem 2 also suggests two, natural dimensions along which to order or sort firms, namely, the intrinsic profitability of the firm (its μ), or the severity of the agency problem (its λ). Governance is higher when μ is higher, because the marginal returns to monitoring are higher. More precisely, governance delays the termination of the firm, which is the inefficiency in this model, and greater (costly) governance is made worthwhile when μ is higher. On the other hand, governance is lower when the agency problem is more severe, because the agent is now paid greater information rents, which reduces the marginal benefit of monitoring as the firm is now less profitable. This, in turn, reduces the stock prices, because termination is more likely. Finally, observe the seemingly paradoxical property that although dividends are paid out at a lower threshold for cash reserves, stocks are less valuable. This is due to the increased propensity for termination, as noted above.

One may also ask, Would an agent prefer to be at a firm with more, or less, governance? As it turns out, the answer depends on how much bargaining power the agent has, and more crucially, the *source* of the increased governance, which is either greater intrinsic profitability or smaller private benefits for the agent from misappropriation.

It is useful to consider the utility promised to the agent at the time of initialisation, which is determined by the relative bargaining power of principal and agent. At one extreme, if the principal has all the bargaining power, she chooses w_0 such that $F'(w_0) = 0$, ie, an initial promise to maximise her profits. If the agent has all the bargaining power, she chooses the largest w for which $F(w) = K$, where K is the initial capital outlay required for the project. We denote the agent's choice of promised utility by $w_\sharp$.

Proposition 6.2. If the principal has the all the bargaining power, then w_0 increases in μ and λ. On the other hand, if the agent has all the bargaining power, then $w_\sharp$ increases in μ but decreases in λ.

Thus, if the agent has all the bargaining power, he welcomes greater governance because the firm is more profitable, one where he can extract a larger amount of the surplus. On the other hand, if the principal holds all the bargaining power, the reason for increased governance matters. Proposition 6.2 says that agents would prefer to be at a firm with greater governance *if* that increased governance is because the firm is intrinsically more profitable, ie, has a higher μ. This is intuitive, because a higher μ corresponds to greater surplus, and some of that additional surplus goes to the agent, via his information rent. On the other hand, if the increased governance is because the agent's benefits from misappropriation are smaller, then the agent is worse off because his information rents, which is the only reason he gets paid, are lower.

We now briefly discuss why Theorem 2 is true.

6.2. Ideas behind Theorem 2

In spite of the fact that Theorem 2 states that the impact of an increase in μ or decrease in λ is qualitatively the same, establishing these claims requires different approaches, primarily because λ and μ affect the boundary conditions for the HJB equation [5.2] differently. Appendix C.1 contains the proofs for a change in μ, while Appendix C.2 proves Theorem 2 for the cases where λ changes.

First, notice that with a change of variable $m = w/\lambda$, the new value function $\hat{F}(m) = F(\lambda m)$ satisfies the new HJB equation

$$r\hat{F}(m) = \mu + \gamma w \hat{F}'(m) + \max_{\sigma \in \Sigma} \left[\tfrac{1}{2}\sigma^2 \hat{F}''(m) - \rho(\sigma) \right] \tag{6.1}$$

with the boundary conditions $\hat{F}(0) = 0$, $\hat{F}'(w^\star/\lambda) = -\lambda$, and $\hat{F}''(w^\star/\lambda) = 0$. The crucial property of this controlled differential equation is that the points at which there is a regime change, ie, when there is a change in σ, are independent of both λ

and μ. This enables us to show, using an appropriate version of the Comparison Theorem for the boundary value problem in [6.1], that $\hat{F}''(m;\cdot)$ is increasing in λ and decreasing in μ, and so by Theorem 1, governance is decreasing in λ and increasing in μ. An instance of this is in Figure 3 for the case when μ increases and there are only two relevant levels of monitoring.

That stock prices are decreasing in λ and increasing in μ requires us to analyse the boundary value problem [5.10] for stocks, and again use the Comparison Theorem. This is trickier, because as is apparent from [5.10], $\mathcal{S}''(m)$ is discontinuous, and the optimal σ is discontinuous in m, while $\mathcal{S}$ and $\mathcal{S}'$ are continuous in m. Thus, a solution exists, but only as a viscosity solution, and not as a classical, C^2 solution. We adapt the Comparison Theorem in Crandall, Ishii, and Lions (1992, Theorem 3) for our differential equation with a discontinuous coefficient, and use it to establish the desired monotonicity.

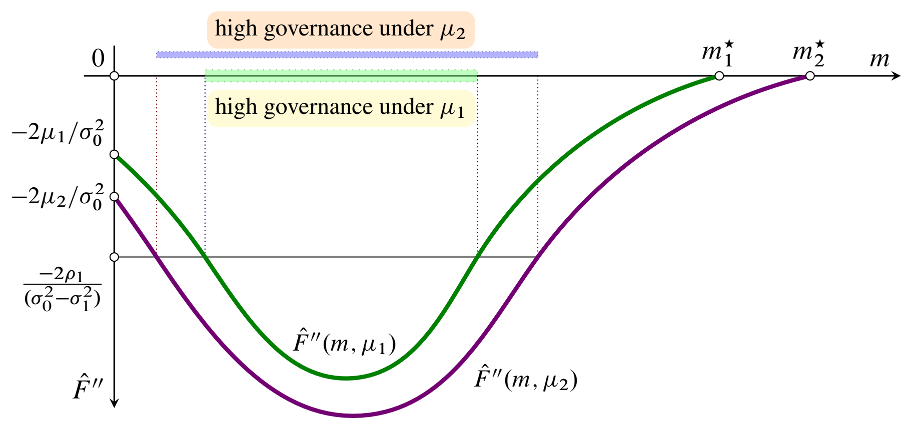

Figure 3: The shape of $\hat{F}''$ as a function of μ where $\mu_1 < \mu_2$. The value $\hat{F}'' = -2\rho_1/(\sigma_0^2 - \sigma_1^2)$ is when the principal is indifferent between high and low monitoring, and this value is independent of μ.

7. Measurement

7.1. A Measure of Governance

Recall from [5.11] that $\mathsf{V}_t = \mathcal{S}'(M_t)\hat{\sigma}(M_t)/S_t$ is the local volatility of the dynamics of the stock price as given in [5.11], where $S_t = \mathcal{S}(M_t)$. Delta (denoted by Δ_t) measures the sensitivity of compensation to stock price. In our setting, we can write $W_t = \lambda\mathcal{S}^{-1}(S_t)$, so that $\Delta_t = \lambda\big(\mathcal{S}^{-1}(S_t)\big)' = \lambda/\mathcal{S}'(M_t)$. Both are well measured empirically. [Fill in details.]

It is easy to see that

$$\mathsf{V}_t \cdot \Delta_t \cdot S_t = \lambda \hat{\sigma}_t \tag{7.1}$$

Thus, by measuring local volatility of stock price and Delta at time t, we can find a measure of governance using [7.1].

7.2. Bounds on Delta

In practice, Delta is often measured at discrete intervals of stock prices. Let s_0, s_1 be two levels of the stock price, with $s_0 < s_1$. Then, define

$$\Delta(s_0, s_1) = \frac{\mathcal{S}^{-1}(s_1) - \mathcal{S}^{-1}(s_0)}{s_1 - s_0}$$

Simple algebra shows that

$$\frac{\hat{\sigma}(\mathcal{S}^{-1}(s_0))}{s_0} \leq \Delta(s_0, s_1) \cdot \tilde{\sigma}(s_0)$$

where $\tilde{\sigma}(s_0)$ is the local volatility of stock price at s_0, as in the SDE [5.11]. Similarly, we can establish an upper bound on the product of Delta and local volatility.

Lemma 7.1. There are upper and lower bounds on the product $\Delta(s_0, s_1) \cdot \tilde{\sigma}(s_0)$ that are decreasing in μ and increasing in λ.

Thus, although we are not able to analytically describe the relation between Delta or stock volatility and governance, we can show that the product of Delta and stock volatility is bounded below and that a reduction in this bound corresponds to greater governance.

7.3. Vintage and Governance

Suppose there are, as in the real world, lots of firms, with many degrees of intrinsic profitability. What can we say about firms that are long-lived? The quantity $\mathsf{E}[\mu \mid T]$ is increasing in T. Thus, older firms are more likely to have higher governance on average.

7.4. The Impact of Public Policy

In the wake of the financial scandals and failures of the late 1990s (Enron, Tyco, Worldcom), the US Congress passed the Sarbanes-Oxley (SOX) Act in 2002. The act requires that top managers *personally* certify the accuracy of financial reports, and also requires, among a host of other things, that firms include an internal control report with their annual audit. The main goal of the act is to improve governance, and to mitigate agency problems, which were seen as the root cause of the corporate failures of the '90s.

In our framework, the SOX Act can be be viewed as imposing lower bounds on the amount of governance in the firm. For concreteness, we assume that the SOX Act requires that firms never have the lowest level of monitoring, namely σ_0. Clearly, this imposes costs on firms that they would not otherwise face.

To understand the effect of such a policy, we provide a useful decomposition. Let $\tilde{\rho}(\sigma) := \max[\rho(\sigma) - \rho(\sigma_1), 0]$ be a hypothetical cost of monitoring, and let $\tilde{\mu} := \mu - \rho(\sigma_1)$ be the implied profitability under the new hypothetical cost of monitoring.

The SOX Act can thus be viewed as lowering the cost of monitoring at all levels while also simultaneously reducing the intrinsic profitability of the firm. For any function $\mathcal{H}(m)$, we denote the change in its value due to the SOX Act by

$$\Delta_{\text{sox}}\mathcal{H}(m) := \underbrace{\Delta_{\text{gov}}\mathcal{H}(m)}_{\text{governance}} + \underbrace{\Delta_{\mu}\mathcal{H}(m)}_{\text{profitability}}$$

Thus, the change $\Delta_{\text{sox}}\mathcal{H}(m)$ in $\mathcal{H}(m)$ due the SOX Act can be decomposed into a *governance* effect denoted by $\Delta_{\text{gov}}\mathcal{H}(m)$, and a *profitability* effect given by $\Delta_{\mu}\mathcal{H}(m)$, where $\Delta_{\text{gov}}\mathcal{H}(m) := \mathcal{H}(m, \tilde{\rho}, \mu) - \mathcal{H}(m, \rho, \mu)$ and $\Delta_{\text{sox}}\mathcal{H}(m) := \mathcal{H}(m, \tilde{\rho}, \tilde{\mu}) - \mathcal{H}(m, \tilde{\rho}, \mu)$.

We now describe the effects of the SOX Act.

Proposition 7.2. Under a policy that renders the monitoring level σ_0 inadmissible, as under the SOX Act, the effects on the firm are as follows.

(a) $\Delta_{\text{sox}}\hat{F}(m) < 0$.
(b) $\Delta_{\text{sox}}m^\star = \Delta_{\text{gov}}m^\star + \Delta_{\mu}m^\star$, where $\Delta_{\text{gov}}m^\star < 0$ and $\Delta_{\mu}m^\star < 0$.
(c) $\Delta_{\text{sox}}\mathcal{S}(m) = \Delta_{\text{gov}}\mathcal{S}(m) + \Delta_{\mu}\mathcal{S}(m)$, where $\Delta_{\text{gov}}\mathcal{S}(m) \lesseqgtr 0$, and $\Delta_{\mu}\mathcal{S}(m) < 0$.
(d) $\Delta_{\text{sox}}\zeta(m) = \Delta_{\text{gov}}\zeta(m) + \Delta_{\mu}\zeta(m)$, where $\Delta_{\text{gov}}\zeta(m) \lesseqgtr 0$, and $\Delta_{\mu}\zeta(m) < 0$.

Moreover, firms that would otherwise be financed by equity would not be any more, because of the increased governance requirements.

Part (a) is clear because the principal is now more constrained, and policies that were once available to him and in use, are no longer so. To see part (b), notice that $\Delta_{\text{gov}}m^\star < 0$ follows immediately from Figure 1 once we observe that a reduction in monitoring costs makes the principal better off. That $\Delta_{\mu}m^\star < 0$ is established in Theorem 2. Notice that the profitability effect on stock price $\mathcal{S}$ and the credit yield spread ζ is always negative. This follows from Theorem 2. However, the governance effect on both these quantities is ambiguous. This is because there are two competing forces at work. The first is that there is likely to be increased monitoring for some values of m, though there could be lower monitoring for others. On average, this effect is ambiguous. The other force is that the payment threshold always decreases which, *ceteris paribus*, results in a faster path to liquidation, but also to more frequent dividend payments. It is not clear what the cumulative effect of these forces is, which renders the governance effect ambiguous.

Our findings are consistent with the work of Core, Guay, and Rusticus (2006), who note that although there is a correlation between higher levels of governance and stock prices (for instance), the causal link is weak at best. That is, increasing the minimum level of governance required need not result in uniformly higher stock prices or lower credit yield spreads.

This is consistent with the evidence on the introduction of new governance regulations ...XXXX. The intuition behind these results is that governance is a substitute for compensation, and so reduces the information rents that the agent gets. More specifically, with increased governance, the expected time to (inefficient) liquidation goes up, so the stock is more valuable as the firm is more likely to pay dividends more often, and credit is more easily available as the firms likelihood of liquidation diminishes, which is reflected in a diminution of the firm's credit yield spread. Firms that may otherwise have been financed are less likely to be so with increased governance requirements, because the principal now has additional monitoring costs to bear. This is also consistent with empirical evidence. Finally, firms are more likely to go private to avoid the costs of governance, also borne out by the data.

Appendices

A. Proofs from Section 4

Recall that $\hat{Y}_t$ is the process that is observed by the principal, and hence the contract is conditioned on.

Proof of Lemma 4.2. Fix a contract $\Phi = (c, \tau, \sigma)$. Let $\hat{Y}$ be a reporting strategy for the agent.[19] His utility from such a strategy, for all $t \in [0, \tau]$, is

[A.1]
$$\begin{aligned} V_t &:= \mathsf{E}_t^{\hat{Y},\sigma}\left[\int_0^\tau e^{-\gamma s}\big(\mathrm{d}C_s + \lambda(\mathrm{d}Y_s - \mathrm{d}\hat{Y}_s)\big)\right] \\ &= \int_0^t e^{-\gamma s}\big(\mathrm{d}C_s + \lambda(\mathrm{d}Y_s - \mathrm{d}\hat{Y}_s)\big) + e^{-\gamma t} W_t \end{aligned}$$

where W_t is the process defined in [4.1]. But for a fixed $\hat{Y}$ and contract, we find that (V_t) is a martingale, and so by the Martingale Representation Theorem there exists a process $Z = (Z_t)$ such that

[A.2]
$$V_t = \int_0^t e^{-\gamma s} Z_s \,\mathrm{d}B_s$$

From [A.1] and [A.2], we find that

$$e^{-\gamma t} W_t = \int_0^t e^{-\gamma s} Z_s \,\mathrm{d}B_s - \int_0^t e^{-\gamma s}\big(\mathrm{d}C_s + \lambda(\mathrm{d}Y_s - \mathrm{d}\hat{Y}_s)\big)$$

Writing this in differential form (and cancelling $e^{-\gamma t}$ throughout), we obtain

[A.3]
$$\mathrm{d}W_t = \gamma W_t \,\mathrm{d}t - \big(\mathrm{d}C_t + \lambda(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t)\big) + Z_t \,\mathrm{d}B_t$$

Noting that $\mathrm{d}B_t = \mathrm{d}Y_t - \mu \mathrm{d}t$ and substituting in [A.3] completes the proof. □

Lemma 4.2 also now lets us characterize incentive compatibility for the agent.

Proof of Lemma 4.3. Suppose the contract is incentive-compatible. By the Comparison Principle for BSDEs (REF???) or equivalently, following Sannikov (2008) and DeMarzo and Sannikov (2006), it is optimal for the agent to minimize the drift of

[19] Recall our assumption that $\hat{Y}$ is absolutely continuous with respect to Y. This implies $\mathrm{d}\hat{Y} = (\mu - a_t\sigma_t)\,\mathrm{d}t + \sigma_t \,\mathrm{d}B_t$ where $a_t\sigma_t$ is the instantaneous diversion of output.

the SDE in [A.3]. Using Girsanov's Theorem, we can rewrite [A.3] as

$$\mathrm{d}W_t = \gamma W_t \mathrm{d}t - \left[\mathrm{d}C_t + \lambda(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t)\right] + Z_t\left[\mathrm{d}\hat{B}_t + \sigma_t^{-1}(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t)\right]$$

For truthtelling (ie, $\mathrm{d}Y_t = \mathrm{d}\hat{Y}_t$) to be optimal, it must be that the contract specifies Z_t and σ_t such that for all t, $-\lambda + Z_t\sigma_t^{-1} \geq 0$, ie, $Z_t \geq \lambda\sigma_t$, as claimed. (This is the content of the Comparison Principle for BSDEs.) The sufficiency of this condition follows from the Comparison Principle for BSDEs. Alternatively, the argument in DeMarzo and Sannikov (2006) may be adapted to our setting. □

B. Proofs from Section 5

B.1. Proofs from Section 5.2

Proof of Proposition 5.2. We have $\mathrm{d}(e^{-rt}\mathcal{S}(M_t)) = -re^{-rt}\mathcal{S}(M_t)\,\mathrm{d}t + e^{-rt}\,\mathrm{d}\mathcal{S}(M_t)$. By Itô's Lemma and from [5.3] which describes the process W_t, we have

$$\begin{aligned}\mathrm{d}(e^{-rt}\mathcal{S}(M_t)) = -re^{-rt}\mathcal{S}(M_t)\,\mathrm{d}t + e^{-rt}\left[\mathcal{S}'(M_t)\gamma M_t + \tfrac{1}{2}\mathcal{S}''(M_t)\sigma^2(\lambda M_t)\right]\mathrm{d}t \\ + e^{-rt}\lambda^{-1}\mathcal{S}'(M_t)\,\mathrm{d}C_t + e^{-rt}\mathcal{S}'(M_t)\sigma(\lambda M_t)\,\mathrm{d}B_t\end{aligned}$$

Integrating from s to $T \wedge \tau$, we obtain

$$\begin{aligned}e^{-rT\wedge\tau}\mathcal{S}(M_{T\wedge\tau}) = e^{-rs}\mathcal{S}(M_s) + \int_s^{T\wedge\tau} e^{-rt}\mathcal{S}'(M_t)\sigma(\lambda M_t)\,\mathrm{d}B_t \\ + \int_s^{T\wedge\tau} e^{-rt}\mathcal{S}'(M_t)\lambda^{-1}\,\mathrm{d}C_t \\ + \int_s^{T\wedge\tau} e^{-rt}\left[-r\mathcal{S}(M_t) + \gamma M_t\mathcal{S}'(M_t) + \tfrac{1}{2}\sigma^2(\lambda M_t)\mathcal{S}''(M_t)\right]\mathrm{d}t\end{aligned}$$

Taking conditional expectations on both sides (relative to $\mathcal{F}_s$), then letting $T \to \infty$, and observing that $M_\tau = 0$, $\mathcal{S}(0) = 0$, and $\mathcal{S}'(w^\star/\lambda) = 1$, we find that[20]

$$\begin{aligned}e^{-rs}\mathcal{S}(M_s) = \mathsf{E}_s\left[\int_s^\tau e^{-rt}\lambda^{-1}\,\mathrm{d}C_t\right] \\ + \mathsf{E}_s\int_s^\tau e^{-rt}\left[-r\mathcal{S}(M_t) + \gamma M_t\mathcal{S}'(M_t) + \tfrac{1}{2}\sigma^2(\lambda M_t)\mathcal{S}''(M_t)\right]\mathrm{d}t\end{aligned}$$

[20]This assumes τ is well-behaved. In particular, need to show it is finite a.s.

By [5.9], it follows that

$$\mathsf{E}\int_s^\tau e^{-rt}\big[-r\mathcal{S}(M_t)+\gamma M_t\mathcal{S}'(M_t)+\tfrac{1}{2}\sigma^2(\lambda M_t)\mathcal{S}''(M_t)\big]\,\mathrm{d}t=0$$

for all $s\in[0,\tau]$. This implies that $r\mathcal{S}(M_t)=\gamma M_t\mathcal{S}'(M_t)+\frac{1}{2}\sigma^2(\lambda M_t)\mathcal{S}''(M_t)$, which establishes the ODE in [5.10]. The boundary conditions easily follow. □

Proof of Proposition 5.3. Recall the SDE for the process (M_t) in [5.7]. Using the Itô expansion of $e^{-rt}M_t$ with $M_0=m$, we find that for each $T\geq 0$

$$e^{-rT\wedge\tau}M_{T\wedge\tau}=m+\int_0^{T\wedge\tau}e^{-rt}\big[(\gamma-r)M_t\,\mathrm{d}t+\lambda\sigma(\lambda M_t)\,\mathrm{d}B_t-\lambda^{-1}\,\mathrm{d}C_t\big]\tag{B.1}$$

Taking expectations in [B.1], then letting $T\to\infty$, and finally observing that $M_\tau=0$ (by definition of τ) and the stochastic integral is a martingale, we find that $m=\mathsf{E}_0^m\big[\int_0^\tau e^{-rt}\big(\lambda^{-1}\,\mathrm{d}C_t-(\gamma-r)M_t\,\mathrm{d}t\big)\big]$. The expectation can be written as

$$\begin{aligned}
&\underbrace{\mathsf{E}_0^m\left[\int_0^\tau e^{-rt}\big(\mu-(\gamma-r)M_t\,\mathrm{d}t\big)\right]}_{=\mathcal{D}(m)}+(1-\lambda)\underbrace{\mathsf{E}_0^m\left[\int_0^\tau e^{-rt}\lambda^{-1}\,\mathrm{d}C_t\right]}_{=\mathcal{S}(m)}\\
&\qquad=m-\lambda\,\mathsf{E}_0^m\left[\int_0^\tau e^{-rt}\lambda^{-1}\,\mathrm{d}C_t\right]+\mathsf{E}_0^m\left[\int_0^\tau e^{-rt}\mu\,\mathrm{d}t\right]\\
&\qquad=m+F(\lambda m)+\mathsf{E}_0^m\left[\int_0^\tau e^{-rt}\rho(\sigma(\lambda M_t))\,\mathrm{d}t\right]
\end{aligned}$$

as claimed. □

Following the arguments above for $\mathcal{S}$, it is not hard to show that $T_s=\mathcal{T}(M_s)$ for all $s\in[0,\tau)$, and that $\mathcal{T}$ is the solution to the following boundary value problem.

$$\begin{aligned}
r\mathcal{T}(m)&=\gamma m\mathcal{T}'(m)+\tfrac{1}{2}\hat{\sigma}^2(m)\mathcal{T}''(m)\\
\mathcal{T}(0)&=1\\
\mathcal{T}'(m^\star)&=0
\end{aligned}\tag{B.2}$$

Lemma B.1. The process T_t is given by $T=\mathcal{T}(m)$, where $\mathcal{T}$ is a solution to the boundary value problem [B.2], and $m\in[0,m^\star]$. Moreover, the expected discounted time to dissolution is a strictly decreasing and strictly convex function of m, the level of cash reserves.

Proof. First part as before ...

To see that $\mathcal{T}$ is strictly decreasing, fix $m_1<m_0$, and define the stopping time

$\xi := \min\{t : M_t = m_1 \mid M_0 = m_0\}$. Then, $\mathcal{T}(m_0) = \mathsf{E}[e^{-r\xi}\mathcal{T}(m_1)] = \mathcal{T}(m_1)\,\mathsf{E}[e^{-r\xi}] < \mathcal{T}(m_1)$ because $\mathsf{E}[e^{-r\xi}] \in (0,1)$, which holds because $\mathsf{E}[e^{-r\xi}] < \infty$ almost surely.

To see that $\mathcal{T}$ is strictly convex, consider the boundary value problem in [B.2]. Thus, $\mathcal{T}''(m) = r\mathcal{T}(m) - \gamma m \mathcal{T}'(m)$. Because $\mathcal{T}(m) > 0$ for all $m \in (0, m^\star]$ and $\mathcal{T}$ is decreasing, so $\mathcal{T}'(m) < 0$, it follows that $\mathcal{T}''(m) > 0$, ie, $\mathcal{T}$ is strictly convex. □

Proof of Proposition 5.4. Let $\Delta = \varphi(T) := rT/(1-T)$. It is easy to see that φ is strictly increasing and strictly convex. By Lemma B.1, $T = \mathcal{T}(m)$. Therefore, we can write $\Delta = \delta(m)$, where $\delta(m) := \varphi(\mathcal{T}(m))$. Differentiation shows that $\delta'(m) = \varphi'(\cdot)\mathcal{T}'(m) < 0$, which follows because φ is increasing while $\mathcal{T}$ is decreasing. The function δ is a composition of two convex functions, and hence is also convex. However, on every interval where $\hat{\sigma}$ is constant, we find $\delta''(m) = \varphi''(\cdot)[\mathcal{T}'(m)]^2 + \varphi'(\cdot)\mathcal{T}''(m) > 0$, where we again use the strict monotonicity and convexity of both φ and $\mathcal{T}$. □

C. Proofs from Section 6

We begin with a principle that finds repeated use below. Consider the nonlinear differential equation

$$\Psi(x, F, DF, D^2F) = 0 \tag{C.1}$$

where $x \in \mathbb{R}$, $F : \mathbb{R}_+ \to \mathbb{R}$, DF and D^2F represent derivatives of F, and Ψ is a nonlinear function. This setting clearly subsumes the HJB equations encountered above, and much else besides. The function F is *proper* if $F(x, r, p, X) \le F(x, s, p, X)$ whenever $r \le s$, and is *degenerate elliptic* if $F(x, r, p, X) \le F(x, r, p, Y)$ whenever $Y \le X$.

A function $g(x)$ is a *subsolution* of [C.1] if $\Psi(x, g, Dg, D^2g) \le 0$, and is a *supersolution* of [C.1] if $\Psi(x, g, Dg, D^2g) \ge 0$. A solution g to [C.1] is both a supersolution, as well as a subsolution.

Theorem 3 (Comparison Principle). *Let I be a bounded open interval of $\mathbb{R}$, and $\Psi : I \times \mathbb{R}^3 \to \mathbb{R}$ in [C.1] be continuous, proper, and degenerate elliptic. Let f (respectively, g) be a subsolution (respectively, supersolution) of $\Psi = 0$ in I, and suppose $f \le g$ on ∂I. Then, $f \le g$ on $\mathrm{cl}\, I$.*

In the above, ∂I is the boundary of I, while $\mathrm{cl}\, I$ is its closure. Theorem 3 is a very special case of the general comparison principle proved as Theorem 3.3 in Crandall, Ishii, and Lions (1992). In particular, they only require f and g to be, respectively, upper and lower semicontinuous. Using the notion of viscosity derivatives, they provide a way to interpret Dg and D^2g, in much that same way that the subdifferential of a convex function generalizes the notion of a derivative of a smooth function. We refer the reader to Crandall, Ishii, and Lions (1992) for an accessible introduction to the theory of viscosity solutions of nonlinear differential equations.

C.1. Proofs from Section 6.1 — Comparative Statics in μ

We analyze here the sensitivity of the value function and other policy variables to the parameter μ.

Lemma C.1. The value function F satisfies

$$\frac{\partial F(w,\mu)}{\partial \mu} = \frac{1 - \mathsf{E}[e^{-r\tau}]}{r} < \frac{1}{r} \tag{C.2}$$

In particular, $F(w,\mu)$ and $\mu - rF(w,\mu)$ are both increasing in μ.

Proof. Let σ^* denote the optimal monitoring strategy and C^* the optimal payment strategy. The value function F is defined in [3.1]. In particular, we have

$$F(w,\mu) = \mathsf{E}^{\sigma^*}\left[\int_0^\tau e^{-rs}\big[(\mu - \rho(\sigma_t^*))\,\mathrm{d}s - \mathrm{d}C_s^*\big] \mid W_0 = w\right]$$

Notice that $F(w,\mu)$ is bounded above by $\mu/r - w$, and below by $-w$. Therefore, we may apply the envelope theorem from Milgrom and Segal (2002), to conclude that [C.2] holds. It follows immediately that $F(w,\mu)$ is increasing in μ. By [C.2], $1 - r\frac{\partial F(w,\mu)}{\partial \mu} = \mathsf{E}[e^{-r\tau}] > 0$, which completes the proof. □

Corollary C.2. The payment boundary $w^\star(\mu)$ is increasing in μ, and hence so is the cash reserve threshold $m^\star = w^\star/\lambda$.

Proof. Recall the smooth pasting condition $F'(w^\star) = -1$, and the supercontact condition $F''(w^\star) = 0$ in [5.6], which imply that at $w = w^\star$, the HJB equation [5.2] becomes

$$rF(w^\star,\mu) = \mu - \gamma w^\star$$

Differentiating with respect to μ, we obtain

$$r\frac{\partial F(w^\star,\mu)}{\partial \mu} + r\underbrace{F'(w^\star,\mu)}_{=-1}\frac{\mathrm{d}w^\star}{\mathrm{d}\mu} = 1 - \gamma\frac{\mathrm{d}w^\star}{\mathrm{d}\mu}$$

which implies

$$(\gamma - r)\frac{\mathrm{d}w^\star}{\mathrm{d}\mu} = 1 - r\frac{\partial F(w^\star, \mu)}{\partial \mu} = \mathsf{E}[e^{-r\tau}] \in (0, 1)$$

where the last equality is from [C.2]. Because $\gamma > r$, it follows that $\frac{\mathrm{d}w^\star}{\mathrm{d}\mu} > 0$, as claimed. □

To understand the effect of a change in μ, one needs to understand how it affects F''. The next lemma takes us in that direction.

Lemma C.3. Given $\mu_1 < \mu_2$, $F'(w, \mu_1) \leq F'(w, \mu_2)$, where the equality holds if, and only if, both sides are -1. Thus, F' is strictly increasing in μ in the relevant part of the domain.

Proof. Let $\mu_2 > \mu_1$ and $w_i^\star$ be the corresponding payment boundary for $i = 1, 2$. By Corollary C.2, $w_2^\star > w_1^\star$, and so $F'(w_1^\star, \mu_2) > F'(w_1^\star, \mu_1) = -1$. Let $w_\circ$ be the largest $w \in [0, w_1^\star)$ such that $F'(w_\circ, \mu_2) = F'(w_\circ, \mu_1)$, so that $F'(w, \mu_2) > F'(w, \mu_1)$ for all $w \in (w_\circ, w_1^\star)$.

In the HJB equation [5.2], let us define $\Phi(\Gamma) := \max_{\sigma \in \Sigma} \left(\frac{1}{2}\sigma^2\lambda^2\Gamma - \rho(\sigma)\right)$. It is easy to see that Φ is increasing in Γ. Notice that the HJB equation [5.2] can now be written as

$$\begin{aligned}
\Phi(F''(w_\circ, \mu_1)) &= rF(w_\circ, \mu_1) - \mu_1 - \gamma w_\circ F'(w_\circ, \mu_1) \\
&> rF(w_\circ, \mu_2) - \mu_2 - \gamma w_\circ F'(w_\circ, \mu_2) \\
&= \Phi(F''(w_\circ, \mu_2))
\end{aligned}$$

where the inequality follows from Lemma C.1 (which says $rF - \mu$ is decreasing in μ) and because $F'(w_\circ, \mu_2) = F'(w_\circ, \mu_1)$ by assumption.

The monotonicity of Φ now implies that $F''(w_\circ, \mu_1) > F''(w_\circ, \mu_2)$. Thus,

$$\begin{aligned}
F''(w_\circ, \mu_1) > F''(w_\circ, \mu_2) &= \lim_{w \downarrow w_\circ} \frac{F'(w, \mu_2) - F'(w_\circ, \mu_2)}{w - w_\circ} \\
&\geq \lim_{w \downarrow w_\circ} \frac{F'(w, \mu_1) - F'(w_\circ, \mu_1)}{w - w_\circ} \\
&= F''(w_\circ, \mu_1)
\end{aligned}$$

which is a contradiction. Thus, there is no such $w_\circ \geq 0$. On the other hand, for all $w \geq w_2^\star$, $F'(w, \mu_2) = F'(w, \mu_1) = -1$, which proves the claim. □

We are now in a position to describe how F'' changes with μ.

Corollary C.4. *$F''(\cdot, \mu)$ is decreasing in μ.*

Proof. Consider the HJB equation [5.2] written as $\Phi(F''(w_\circ, \mu)) = rF(w_\circ, \mu) - \mu - \gamma w F'(w_\circ, \mu)$. By Lemma C.1, it follows that $rF(w_\circ, \mu) - \mu$ is decreasing in μ, while Lemma C.3 says $F'(w_\circ, \mu)$ decreases in μ. The monotonicity of Φ implies $F''(w, \mu)$ is decreasing in μ. □

Let $\sigma^*(w, \mu) = \arg\max_{\sigma \in \Sigma} \left[\frac{1}{2}\sigma^2 \lambda^2 F''(w, \mu) - \rho(\sigma)\right]$ denote the optimal choice of monitoring at w. The behaviour of F'' with respect to μ dictates how optimal monitoring changes with μ.

Proposition C.5. *The optimal level of monitoring, as a function of promised utility or of cash reserves, is increasing in μ.*

Proof. The objective $\frac{1}{2}\sigma^2 \lambda^2 F''(w, \mu) - \rho(\sigma)$ has increasing differences in (σ, μ) if μ is given the reverse order because F'' is monotone decreasing in μ (in the standard order) by Corollary C.4. Therefore, $\sigma^*(w, \mu)$ is decreasing in μ (in the standard order), ie, the level of monitoring increases in μ. It is clear that the same holds as a function of cash reserves, because $m = w/\lambda$ is independent of μ. □

We now show that stock prices are also monotone in μ.

Proposition C.6. *Stock price $\mathcal{S}(m, \mu)$ is increasing in μ.*

Proof. Consider $\mu_1 < \mu_2$, and let σ_i^* be the optimal policy under μ_i, $m_i^\star$ the cash reserve threshold, and $\mathcal{S}_i$ the corresponding stock price. We have already established in Corollary C.2 that $m^\star$ increases in μ, which implies that $\mathcal{S}_2'(m_1^\star) > 1$. By virtue of being a solution to the boundary value problem (at $\mu = \mu_1$), we have

$$0 = r\mathcal{S}_1(m) - \gamma m \mathcal{S}_1'(m) - \tfrac{1}{2}\sigma_1^{*2}\mathcal{S}_1''(m) \geq r\mathcal{S}_1(m) - \gamma m \mathcal{S}_1'(m) - \tfrac{1}{2}\sigma_2^{*2}\mathcal{S}_1''(m)$$

where the inequality is because $\sigma_1^*(\lambda m) \geq \sigma_2^*(\lambda m)$ (by Proposition C.5) and because $S_i'' \leq 0$ for $i = 1, 2$ by Proposition 5.2.

Thus, $\mathcal{S}_1$ is a subsolution to the boundary value problem [5.10]. We have also noted that $\mathcal{S}_2(0) = \mathcal{S}_1(0)$, and $\mathcal{S}_2'(m_1^\star) > \mathcal{S}_1'(m_1^\star) = 1$, and so by the Comparison Theorem XXX, it follows that $\mathcal{S}(m, \mu_2) \geq \mathcal{S}(m, \mu_1)$ for all m (where they are both defined). □

We now show that the expected discounted hitting time is decreasing in μ.

Proposition C.7. *The expected discounted liquidation time $\mathcal{T}(m)$ in [B.2] is decreasing in μ.*

Proof. Let $\mu_1 < \mu_2$, σ_i^* be the optimal policy under μ_i, $m_i^\star$ the cash reserve threshold, and $\mathcal{T}_i$ the corresponding expected discounted liquidation time in [B.2]. We have already established in Corollary C.2 that $m^\star$ increases in μ, which implies that $\mathcal{T}_2'(m_1^\star) > 0$. Because $\mathcal{T}_1$ solves [B.2] when $\mu = \mu_1$, we have

$$0 = r\mathcal{T}_1(m) - \gamma m\mathcal{T}_1'(m) - \tfrac{1}{2}\sigma_1^{*2}\mathcal{T}_1''(m) \geq r\mathcal{T}_1(m) - \gamma m\mathcal{T}_1'(m) - \tfrac{1}{2}\sigma_2^{*2}\mathcal{T}_1''(m)$$

where the inequality is because $\sigma_1^*(\lambda m) \geq \sigma_2^*(\lambda m)$ (by Proposition C.5) and because $T_i'' \leq 0$ for $i = 1, 2$ by Lemma B.1.

Thus, $\mathcal{T}_1$ is a supersolution to the boundary value problem [B.2]. We have also noted that $\mathcal{T}_2(0) = \mathcal{T}_1(0) = 1$, and $\mathcal{T}_2'(m_1^\star) > \mathcal{T}_1'(m_1^\star) = 0$, and so by the Comparison Theorem XXX, it follows that $\mathcal{T}(m, \mu_2) \leq \mathcal{T}(m, \mu_1)$ for all m. □

Corollary C.8. The credit yield spread Δ is decreasing in μ.

Proof. As noted in the proof of Proposition 5.4, $\Delta = \varphi(\mathcal{T}(m, \mu))$, where $\varphi(x) = rx/(1-x)$ is increasing and convex. By Proposition C.7, $\mathcal{T}(m, \mu)$ is decreasing in μ, so it follows that Δ, for a given level of cash reserves, is also decreasing in μ. □

C.2. Proofs from Section 6.1 — Comparative Statics in λ

The HJB equation can be written as a variational inequality as follows:

$$\text{[C.3]} \qquad \begin{aligned} &\Psi(w, F, F', F'', \lambda) \\ &\quad := \min\left[rF(w) - \mu - \gamma wF'(w) - \Phi(F''(w), \lambda), F'(w) + 1\right] \\ &\quad = 0 \end{aligned}$$

with $F(0) = 0$, where $\Phi(\Gamma, \lambda) := \max_{\sigma\in\Sigma}\left(\frac{1}{2}\sigma^2\lambda^2\Gamma - \rho(\sigma)\right)$. We first show that F decreases in λ, and the solution to [C.3] is $F(w, \lambda)$.

Proposition C.9. Let $F(w, \lambda)$ be the solution to the [C.3]. Then, $\lambda_1 < \lambda_2$ implies $F(w, \lambda_1) \geq F(w, \lambda_2)$ for all $w \geq 0$.

Proof. Recall that for any λ, $F''(w, \lambda) \leq 0$ for all $w \geq 0$, with a strict inequality when $w \in [0, w^\star)$. Because $\Gamma \leq 0$, $\Phi(\Gamma, \lambda)$ is decreasing in λ. Therefore, we have

$$0 = \Psi\big(F(w, \lambda_2), \lambda_2\big) \geq \Psi\big(F(w, \lambda_2), \lambda_1\big)$$

Thus, $F(w, \lambda_2)$ is a subsolution to [C.3] $\Psi(F, \lambda_1) = 0$. Because $F(0, \lambda_1) = F(0, \lambda_2)$, we conclude by the Comparison Theorem XXX that $F(w, \lambda_1) \geq F(w, \lambda_2)$ for all $w \geq 0$. □

Corollary C.10. The payment boundary $w^\star$ is increasing in λ.

Proof. The payment boundary is the intersection of $F(w,\lambda)$ and the line $w \mapsto \mu r^{-1} - \gamma r^{-1} w$. Because $F(w,\lambda)$ is decreasing in λ, this point of intersection must be lower, and occur at a higher w, ie, $w^\star$ is increasing in λ. □

To consider the effect of a change in λ on the optimal level of monitoring, it is useful to consider the principal's problem as a function of cash reserves. In particular, consider the change of variable $m = w\lambda^{-1}$, which gives us $\hat{F}(m,\lambda) = F(m\lambda,\lambda)$. Then, we obtain the variational inequality

$$\begin{aligned}&\hat{\Psi}(m,\hat{F},\hat{F}',\hat{F}'',\lambda)\\ &:= \min\left[r\hat{F}(m) - \mu - \gamma m\hat{F}'(m) - \hat{\Phi}(\hat{F}''(m)), \hat{F}'(m) + \lambda\right]\\ &= 0\end{aligned} \tag{C.4}$$

with the boundary condition $\hat{F}(0) = 0$, and where $\hat{\Phi}(\Gamma) := \max_{\sigma\in\Sigma}\left(\frac{1}{2}\sigma^2\Gamma - \rho(\sigma)\right)$. We also let $\hat{\sigma}(m) := \arg\max_{\sigma\in\Sigma}\left(\frac{1}{2}\sigma^2\Gamma - \rho(\sigma)\right)$ denote the optimal choice of monitoring as a function of cash reserves. The advantage of this change in perspective, demonstrated next, is that the nonlinear operator $r\hat{F}(m) - \mu - \gamma m\hat{F}'(m) - \hat{\Phi}(\hat{F}''(m))$ is independent of λ.

In what follows, we suppress the dependence of $\hat{F}$ on λ where this dependence is not emphasized for comparison.

Proposition C.11. Let $\hat{F}(m,\lambda)$ be a solution to [C.4]. Then, $\lambda_1 \le \lambda_2$ implies $\hat{F}(m,\lambda_2) \le \hat{F}(m,\lambda_1)$.

Proof. By assumption, $\hat{\Psi}(m,\hat{F},\hat{F}',\hat{F}'',\lambda_1) = 0$. But we also have

$$\min\left[r\hat{F}(m,\lambda_1) - \mu - \gamma m\hat{F}'(m,\lambda_1) - \hat{\Phi}(\hat{F}''(m,\lambda_1)), \hat{F}'(m,\lambda_1) + \lambda_2\right] \ge 0$$

which implies $\hat{F}(m,\lambda_1)$ is a supersolution to the nonlinear differential equation $\hat{\Psi}(m,\hat{F},\hat{F}',\hat{F}'',\lambda_2) = 0$ in [C.4]. Because $\hat{F}(m,\lambda_2)$ is a solution (and hence a subsolution) to $\hat{\Psi}(m,\hat{F},\hat{F}',\hat{F}'',\lambda_2) = 0$, it follows from Comparison Theorem XXX that $\hat{F}(m,\lambda_1) \ge \hat{F}(m,\lambda_2)$. □

To understand how optimal monitoring $\hat{\sigma}$ and the dividend payment threshold $m^\star = w^\star/\lambda$ vary with λ, we need to understand how $\hat{F}''(m,\lambda)$ changes with λ. The following lemma describes this behaviour.

Lemma C.12. Let $\hat{F}(m,\lambda)$ be a solution to [C.4]. Then, $\lambda_1 \le \lambda_2$ implies $\hat{F}'(m,\lambda_1) \ge \hat{F}'(m,\lambda_2)$ and $\hat{F}''(m,\lambda_1) \le \hat{F}''(m,\lambda_2)$.

Proof. Let $G(m) := r\hat{F}(m) - \gamma m\hat{F}'(m) - \mu = \hat{\Phi}(\hat{F}''(m))$ and notice that $\hat{\Phi}(\Gamma) = \max_\sigma\left(\frac{1}{2}\sigma^2\Gamma - \rho(\sigma)\right)$ is a strictly increasing function of Γ. Therefore, $\hat{\Phi}^{-1}$ is a well defined and strictly increasing function. If $G(m)$ is increasing in λ, then it follows that $\hat{F}''$ is also increasing in λ. In addition, $G(m)$ increasing in λ implies, by Proposition C.11, that $G(m)-rF(m)+\mu = -\gamma m\hat{F}'(m)$ is increasing in λ, ie, $\hat{F}'(m,\lambda)$ is decreasing in λ. Thus, all that remains is to show that $G(m)$ is increasing in λ.

To see that $G(m)$ increases in λ, observe first that

$$
\begin{aligned}
G'(m) &= (r-\gamma)\hat{F}'(m) - \gamma m\hat{F}''(m) \\
&= \underbrace{(r-\gamma)(\gamma m)^{-1}r\hat{F}(m)}_{=:\mathcal{K}(m,\lambda)} + \underbrace{(\gamma - r)(\gamma m)^{-1}G(m) - \gamma m\hat{\Phi}^{-1}(G(m))}_{=:\mathcal{G}(m,G)} \\
&= \mathcal{G}(m, G(m)) + \mathcal{K}(m,\lambda)
\end{aligned}
$$

Thus, G is the solution to the differential equation $G'(m) = \mathcal{G}(m, G(m)) + \mathcal{K}(m,\lambda)$ for each λ. By Proposition C.11, we see that $\mathcal{K}(m,\lambda)$ is increasing in λ. Thus, by the Comparison Theorem for first order differential equations — see, for instance, Birkhoff and Rota (1989, Theorem 8, p.30) — we find that $G(m,\lambda_1) \le G(m,\lambda_2)$, which completes the proof. $\square$

Corollary C.13. The dividend payment threshold $m^\star = w^\star/\lambda$ is decreasing in λ.

Proof. By definition, $m_i^\star$ satisfies $\hat{F}''(m_i^\star,\lambda_i) = 0$ for $i = 1,2$ where $\lambda_1 < \lambda_2$. By Lemma C.12, $\hat{F}''(m,\lambda)$ is increasing in λ. Therefore, $0 = \hat{F}''(m_2^\star,\lambda_2) \ge \hat{F}''(m_2^\star,\lambda_1)$, which implies $m_2^\star \le m_1^\star$. $\square$

We can now describe how optimal monitoring changes with λ.

Proposition C.14. The optimal $\hat{\sigma}(m)$ is increasing in λ. Thus, monitoring is decreasing in λ.

Proof. Notice that $\frac{1}{2}\sigma^2\hat{F}''(m,\lambda) - \rho(\sigma)$ has increasing differences in (σ,λ) because by Proposition C.11, $\hat{F}''(m,\lambda)$ is increasing in λ. Therefore, by Topkis's Theorem, $\hat{\sigma}(m,\lambda) = \arg\max_{\sigma\in\Sigma}\left[\frac{1}{2}\sigma^2\hat{F}''(m,\lambda) - \rho(\sigma)\right]$ is also increasing in λ. $\square$

We now show that stock prices are also monotone in λ.

Proposition C.15. Stock price $\mathcal{S}(m,\mu)$ is decreasing in λ.

Proof. Consider $\lambda_1 < \lambda_2$, and let $\hat{\sigma}_i$ be the optimal policy under λ_i, $m_i^\star$ the cash reserve threshold, and $\mathcal{S}_i$ the corresponding stock price. We have already established in Corollary C.13 that $m^\star$ decreases in λ, which implies that $\mathcal{S}_1'(m_2^\star) > 1$. By virtue of being a solution to the boundary value problem (at $\lambda = \lambda_1$), we have

$$0 = r\mathcal{S}_1(m) - \gamma m\mathcal{S}_1'(m) - \tfrac{1}{2}\hat{\sigma}_1^2\mathcal{S}_1''(m) \leq r\mathcal{S}_1(m) - \gamma m\mathcal{S}_1'(m) - \tfrac{1}{2}\hat{\sigma}_2^2\mathcal{S}_1''(m)$$

where the inequality is because $\hat{\sigma}_1(m) \leq \hat{\sigma}_2(m)$ (by Proposition C.14) and because $\mathcal{S}_i'' < 0$ for $i = 1, 2$ by Proposition 5.2.

Thus, $\mathcal{S}_1$ is a supersolution to the boundary value problem [5.10] when $\lambda = \lambda_2$. We have also noted that $\mathcal{S}_2(0) = \mathcal{S}_1(0) = 0$, and $\mathcal{S}_1'(m_2^\star) > \mathcal{S}_2'(m_2^\star) = 1$, and so by the Comparison Theorem XXX, it follows that $\mathcal{S}(m, \lambda_1) \geq \mathcal{S}(m, \lambda_2)$ for all m. □

We now show that the expected discounted hitting time is increasing in λ.

Proposition C.16. The expected discounted liquidation time $\mathcal{T}(m)$ in [B.2] is increasing in λ.

Proof. Let $\lambda_1 < \lambda_2$, $\hat{\sigma}_i$ be the optimal policy under λ_i, $m_i^\star$ the cash reserve threshold, and $\mathcal{T}_i$ the corresponding expected discounted liquidation time in [B.2] when $\lambda = \lambda_i$. We have already established in Corollary C.13 that $m^\star$ decreases in λ, which implies that $\mathcal{T}_1'(m_2^\star) > 0$. Because $\mathcal{T}_1$ solves [B.2] when $\lambda = \lambda_1$, we have

$$0 = r\mathcal{T}_1(m) - \gamma m\mathcal{T}_1'(m) - \tfrac{1}{2}\hat{\sigma}_1^2\mathcal{T}_1''(m) \geq r\mathcal{T}_1(m) - \gamma m\mathcal{T}_1'(m) - \tfrac{1}{2}\hat{\sigma}_2^2\mathcal{T}_1''(m)$$

where the inequality is because $\hat{\sigma}_1(m) \leq \hat{\sigma}_2(m)$ (by Proposition C.14) and because $\mathcal{T}_i'' > 0$ for $i = 1, 2$ by Lemma B.1.

Thus, $\mathcal{T}_1$ is a subsolution to the boundary value problem [B.2] when $\lambda = \lambda_2$. We have also noted that $\mathcal{T}_2(0) = \mathcal{T}_1(0) = 1$, and $\mathcal{T}_1'(m_2^\star) > \mathcal{T}_2'(m_2^\star) = 0$, and so by the Comparison Theorem XXX, it follows that $\mathcal{T}(m, \lambda_1) \leq \mathcal{T}(m, \lambda_2)$ for all m. □

Corollary C.17. The credit yield spread Δ is increasing in λ.

Proof. As noted in the proof of Proposition 5.4, $\Delta = \varphi(\mathcal{T}(m, \lambda))$, where $\varphi(x) = rx/(1-x)$ is increasing and convex. By Proposition C.16, $\mathcal{T}(m, \lambda)$ is increasing in λ, so it follows that Δ, for a given level of cash reserves, is also increasing in λ. □

D. Intuition in Discrete Time

We now consider a discrete time version of our model. While some of the forces underlying our main results can be seen here, precise statements about optimal contracts are best made in the continuous time model in Section 3. Our formulation is inspired by Milgrom and Roberts (1992, Chapter 7) who also restrict attention to a useful subclass of contracts from which they derive important economic lessons.

Let time be indexed by $t = 0, \triangle, 2\triangle, 3\triangle, \ldots$, where $\triangle > 0$ is the length of a period. Output at time t is given by Y_t. The incremental output $Y_{t+\triangle} = Y_t + \mu \triangle + \sigma_t \varepsilon_t \sqrt{\triangle}$, where $\mu > 0$ is the *intrinsic profitability* of the project, and (ε_t) is an iid process with mean 0 and finite (small) variance. The agent reports output $\hat{Y}_{t+\triangle} = Y_{t+\triangle} - q_t \triangle$, where $q_t \triangle \geq 0$ is the amount of output that the agent diverts for private consumption. To develop intuition, we look at a class of suboptimal contracts, that (as we will see in the sequel) converge to the optimal continuous time contract as $\triangle \to 0$.[21]

A discrete time contract specifies compensation C_{t+1} and sensitivity to output Z_t as a function of the past. It is well known that we may consider recursive contracts that condition on the agent's *promised utility* w_t. Specifically, we let $C_{t+\triangle}$, which is paid at the beginning of period $t + \triangle$, depend only on w_t and not the intervening output. Then, promised utility evolves as

$$w_{t+\triangle} - w_t = \gamma \triangle w_t - \triangle C_{t+\triangle} - \lambda \triangle q_t + Z_t(\hat{Y}_{t+\triangle} - Y_t)/\sigma_t$$

where $\gamma > 0$ is the agent's discount rate. The agent's increment of promised utility in period $t + \triangle$ is the amount of interest he gains from his stock of promised utility over time length $\triangle$, net of his consumption, plus a linear function of his productivity shock via ε_t. Notice that the agent gets extra utility $\lambda q_t \triangle$ by diverting cash flow, where $\lambda \in (0, 1]$. Using our assumptions on output and reported cash flow, we can rewrite the above display as

$$w_{t+\triangle} - w_t = \gamma \triangle w_t - \triangle C_{t+\triangle} + \lambda \triangle q_t + Z_t(\sigma_t \varepsilon_t \sqrt{\triangle} - q_t \Delta)/\sigma_t$$

It is now straightforward to characterize incentive compatibility in such contracts. Requiring the agent to not divert cash, ie $q_t = 0$ for all t, amounts to requiring that

$$q_t \triangle \left(\lambda - \frac{Z_t}{\sigma_t} \right) \leq 0$$

which holds if, and only if,

[D.1] $$Z_t \geq \lambda \sigma_t$$

Lesson 1: Sensitivity and monitoring are substitutes. Monitoring is costly, and if monitoring is high, there is no point in having high-powered incentives.

[21]The characterization of incentive compatibility and the structure of the optimal contract in Theorem 1 resembles the contracts considered in discrete time. While we do not prove formal convergence, ie, we do not consider the formal limit as $\triangle \to 0$, it is clear that, at least in an informal sense, our subclass of contracts form an increasingly good approximation of the optimal discrete time contract when $\triangle$ is sufficiently small.

We now consider the principal's problem. As always, the principal's value function can be written as a Bellman equation, as follows:

$$V(w_t) = \max_{C_{t+\triangle},\sigma_t,Z_t} \left[\frac{(\mu - C_{t+\triangle})\triangle}{1+r\triangle} + \frac{1}{1+r\triangle} \mathsf{E}[V(w_{t+\triangle})] - \rho(\sigma_t) \right] \tag{D.2}$$

subject to incentive compatibility as in [D.1]. Let us assume that V is twice differentiable. Then, we may write a Taylor series expansion of $V(w_t)$ to obtain

$$\mathsf{E}[V(w_{t+\triangle})] = V(w_t) + \gamma \triangle w_t V'(w_t) - \triangle C_{t+\triangle} V'(w_t) + \tfrac{1}{2} Z_t^2 \triangle V''(w_t) + o(\triangle)$$

where $o(\triangle)$ denotes higher order terms in $\triangle$. Notice that in the display above, we have used the fact that $\mathsf{E}[\varepsilon_t] = 0$ and $\mathsf{E}[\varepsilon_t^2] < \infty$. Substituting this in the value function [D.2] and then dividing throughout by $\triangle$, we obtain

$$\begin{aligned} rV(w_t) = \max_{C_{t+\triangle},\sigma_t,Z_t} \Big[&\mu - C_{t+\triangle}(1 + V'(w_t)) + \gamma w_t V'(w_t) \\ &+ \tfrac{1}{2} Z_t^2 V''(w_t) - \rho(\sigma_t) + o(\triangle)/\triangle \Big] \end{aligned} \tag{D.3}$$

It is easy to see that compensation $C_{t+\triangle} > 0$ if, and only if, $V'(w_t) \le -1$. If the value function is concave, then this amounts to deferring compensation. Concavity of V would also imply that Z_t^2 should be as small as possible, ie, we must have $Z_t = \lambda \sigma_t$ for all t, so [D.1] always binds.

All that is left to determine is the optimal level of monitoring σ_t. Recall that, as in Milgrom and Roberts (1992), we interpret monitoring as a reduction in variance of the output. Using the above, we collect all the terms that depend on σ_t, and recall our assumption here that there are only two levels of monitoring, high or low:

$$\max_{\sigma_t} \left[\tfrac{1}{2}\lambda^2 \sigma_\ell^2 V''(w_t) - \rho(\sigma_\ell), \tfrac{1}{2}\lambda^2 \sigma_h^2 V''(w_t) \right]$$

where $\sigma_\ell < \sigma_h$. Concavity of V implies V'' is always nonpositive. It is clear that if V'' is suffficiently close to 0, then there should be low monitoring.

Some lessons from the discrete time analysis:

(i) Sensitivity and monitoring are substitutes. Monitoring is costly, and if monitoring is high, there is no point in having high-powered incentives. This is in contrast with Milgrom and Roberts (1992), and the main reason is that here, costs are linear in effort (cash flow diversion).

(ii) Deferred payments: If V is concave, payments should be deferred. This is a well known idea, and is present in the discrete time literature, for example in DeMarzo and Fishman (2007b), Clementi and Hopenhayn (2006), and Krishna, Lopomo, and Taylor (2013).

(iii) Monitoring should vary over time. In particular, if $V'' \approx 0$, then monitoring should be very low. If V'' is sufficiently small, then monitoring should be high. Thus, $-V''$ is the *marginal value of monitoring*.

Of course, there are many shortcomings with the rudimentary analysis above. First, our contracts are suboptimal, in that w_t evolves linearly with output, and consumption $C_{t+\triangle}$ does not depend on output over the interval $(t, t+\triangle]$. Second, our contract ignores the randomization that would be needed in optimally terminating the contract. This is a well-known property of discrete time contracts — see, Biais et al. (2007).

Shortcomings of discrete time analysis: It is possible to let $\triangle \to 0$, and consider the limit, seeing as how [D.3] resembles an HJB equation. Nonetheless, establishing convergence is not straightforward, and faces technical difficulties of its own. Technically, our analysis is suspect because it relies on the assumption that V is twice differentiable. It is not hard to show that V in [D.2] must be concave. However, it is notoriously difficult to show that value functions in discrete time are twice-continuously differentiable (see Stokey, Lucas, and Prescott (1989)).

The main conceptual problem is that in discrete time, we cannot say much about $V''(w)$. Knowing the structure of $V''(w)$ is crucial in determining the *time structure of the amount of monitoring* to be undertaken; as noted above, $-V''$ is the marginal value of monitoring. As we will see below, our continuous time analysis lays these tradeoffs bare.

Thus, the main contribution of the paper is that the lessons gleaned from our discrete time analysis above hold exactly as properties of the optimal continuous time contract. Moreover, analysis in continuous time also allows us to perform comparative statics in a unified way.

E. Directly Controlling Agency Costs

We consider here a variant of the model where the principal directly controls λ_t, the agency cost, or more precisely, the agent's marginal benefit from diverting cash, while the volatility of output remains fixed at σ_0.

For concreteness, suppose $\lambda_t = a_t \lambda_0$, and $a_t \in A := \{a_0, a_1, \ldots, a_n\}$, where $a_0 = 1$, and $a_i > a_{i+1}$ for all $i = 1, \ldots, n-1$. The instantaneous cost of choosing $a_t = a_i$ is $\kappa(a_i)$.

As in the main model, we can write the agent's promised utility process as

$$\mathrm{d}W_t = \gamma W_t \,\mathrm{d}t - \mathrm{d}C_t + \lambda_t(\mathrm{d}Y_t - \mathrm{d}\hat{Y}_t) + Z_t \,\mathrm{d}B_t$$

where Z_t is a sensitivity process, just as in [4.2]. Incentive compatibility is now characterised as requiring $Z_t \geq \lambda_t \sigma_0 = a_t \lambda_0 \sigma_0$.

Now, consider the change of variables as follows: Let $\sigma_i := a_i \sigma_0$ and $\lambda := \lambda_0$, and define the function $\rho(\sigma_i) := \kappa(a_i)$, so that for all $t \geq 0$ we have $\lambda_t \sigma_0 = \lambda \sigma_t$. But the right hand side is precisely the model studied in the paper, and the cost of controlling λ_t is exactly the cost of changing σ_t.

Thus, the evolution of promised utility in both models is the same, as are the principal's costs, which implies that the principal's value function is identical in both models. It is now easy to show that the optimal contract, as a function of W_t, is also identical, ie, the payment boundary is identical.

It is useful to see how to implement the optimal contract, given that λ_t is changing over time. We define $M_t = W_t / \lambda_0$. Then, we may write the evolution of M as

$$\mathrm{d}M_t = \gamma M_t \,\mathrm{d}t - \lambda_0^{-1} \,\mathrm{d}C_t + \underbrace{a_t \sigma_0}_{=\sigma_t} \,\mathrm{d}B_t$$

which is exactly as in [5.7]. The stock price is $S_t = \mathsf{E}_t \left[\int_t^\tau e^{-r(s-t)} \lambda_0^{-1} \,\mathrm{d}C_s \right]$, and it is easy to see that this is the same stock price as in the main model where σ_t is controlled. Similarly, we consider bonds that have a coupon payment of $\mu - (\gamma - r) M_t$, so that bond price is $D_t = \mathsf{E}_t \left[\int_t^\tau e^{-r(s-t)} [\mu - (\gamma - r) M_s] \,\mathrm{d}s \right]$. Because the stock and bond prices are the same as in the main model, and the evolution of cash reserves is the same, both stock and bond prices are deterministic functions of cash reserves, and these functions satisfy the same boundary value problems as they do for the main model.

In the implementation, let the agent own a fraction λ_0 of the stock, the principal hold a fraction $1 - \lambda_0$ of the stock, and all the debt, ie, the bonds, so that coupon payments on the bond are paid to the principal. Thus, the properties of the implementation and all subsequent results remain the same. In particular, the exact counterpart of Proposition 5.3 holds.

References

Adams, Renée B., Benjamin E. Hermalin, and Michael S. Weisbach. 2010. "The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey." *Journal of Economic Literature* 48 ([1]): 58–107. (Cit. on p. 4).

Aggarwal, Rajesh K., and Andrew A. Samwick. 1999. "The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation." *Journal of Political Economy* 107 ([1]): 65–105. (Cit. on p. 8).

Ashbaugh-Skaife, Hollis, Daniel W. Collins, and Ryan LaFond. 2008. "Corporate Governance, Risk and Cost of Equity." Working Paper. (Cit. on p. 23).

Baiman, Stanley, and Joel S. Demski. 1980. "Economically Optimal Performance Evaluation and Control Systems." *Journal of Accounting Research* 18:184–220. (Cit. on p. 5).

Balsam, Steven, John Puthenpurackal, and Arun Upadhyay. 2016. "The Determinants and Performance Impact of Outside Board Leadership." *Journal of Financial and Quantitative Analysis* 51 ([4]): 1325–1358. (Cit. on p. 7).

Bebchuk, Lucian Arye, Alma Cohen, and Allen Ferrell. 2009. "What Matters in Corporate Governance?" *Review of Financial Studies* 22 ([2]): 783–827. (Cit. on pp. 3, 23).

Becht, Marco, Patrick Bolton, and Ailsa Röell. 2003. "Corporate Governance and Control." In *Handbook of the Economics of Finance*, vol. Volume 1, Part A, 1–109. North-Holland. (Cit. on p. 4).

Berle, Jr., A. A., and Gardiner C. Means. 1930. "Corporations and the Public Investor." *American Economic Review* 20 ([1]): 54–71.

Bhagat, Sanjai, and Brian Bolton. 2018. "Corporate Governance and Firm Performance: The Sequel." *Journal of Corporate Finance* 58:142–168. (Cit. on p. 7).

Bhojraj, Sanjeev, and Partha Sengupta. 2003. "Effect of Corporate Governance on Bond Ratings and Yields: The Role of Institutional Investors and Outside Directors." *Journal of Business* 76 ([3]): 455–475. (Cit. on p. 23).

Biais, Bruno, Thomas Mariotti, Guillaume Plantin, and Jean-Charles Rochet. 2007. "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *Review of Economic Studies* 74 ([2]): 345–390. (Cit. on pp. 1, 3–5, 16, 42).

Birkhoff, Garrett, and Gian-Carlo Rota. 1989. *Ordinary Differential Equations*. Fourth Edition. NJ, USA: John Wiley & Sons. (Cit. on p. 38).

Boschen, John F., and Kimberly J. Smith. 1995. "You Can Pay Me Later: The Dynamic Response of Executive Compensation to Firm Performance." *Journal of Business* 68 ([4]): 577–608. (Cit. on p. 8).

Bulan, Laarni, Paroma Sanyal, and Zhipeng Yan. 2010. "A Few Bad Apples: An Analysis of CEO Performance Pay and Firm Productivity." *Journal of Economics and Business* 62 ([4]): 273–306. (Cit. on p. 8).

Cadenillas, Abel, Jakša Cvitanić, and Fernando Zapatero. 2004. "Leverage Decision and Manager Compensation with Choice of Effort and Volatility." *Journal of Financial Economics* 73 ([1]): 71–92. (Cit. on p. 7).

Chang, Jui-Chin, Mi Luo, and Huey-Lian Sun. 2011. "The Impact of Independent and Overlapping Board Structures on CEO Compensation, Pay-Performance Sensitivity and Accruals Management." *Quarterly Journal of Finance and Accounting* 50 ([2]): 54–84. (Cit. on p. 8).

Chen, Mingliu, Peng Sun, and Yongbo Xiao. 2020. "Optimal Monitoring Schedule in Dynamic Contracts." *Operations Research* 68 ([5]): 1285–1314. (Cit. on p. 6).

Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. "A Theory of Financing Constraints and Firm Dynamics." *Quarterly Journal of Economics* 121:229–265. (Cit. on pp. 5, 14, 41).

Conyon, Martin J. 2014. "Executive Compensation and Board Governance in US Firms." *Economic Journal* 124 ([574]): F60–F89. (Cit. on p. 8).

Cook, Michael L., and Molly J. Burress. 2013. "The Impact of CEO Tenure on Cooperative Governance." *Managerial and Decision Economics* 34 ([3/5]): 218–229. (Cit. on p. 8).

Core, John E., Wayne R. Guay, and Tjomme O. Rusticus. 2006. "Does Weak Governance Cause Weak Stock Returns? An Examination of Firm Operating Performance and Investors' Expectations." *Journal of Finance* 61 ([2]): 655–687.

Crandall, Michael G., Hitoshi Ishii, and Pierre-Louis Lions. 1992. "User's Guide to Viscosity Solutions of Second-Order Partial Differential Equations." *Bulletin of the American Mathematical Society* 27 ([1]): 1–67. (Cit. on pp. 4, 25, 32, 33).

Cremers, K.J. Martijn, Lubomir P. Litov, and Simone M. Sepe. 2017. "Staggered Boards and Long-Term Firm Value, Revisited." *Journal of Financial Economics* 126 ([2]): 422–444. (Cit. on p. 7).

Cvitanić, Jakša, Dylan Possamaï, and Nizar Touzi. 2017. "Moral Hazard in Dynamic Risk Management." *Management Science* 63 ([10]): 3328–3346. (Cit. on p. 7).

Dai, Liang, Yenan Wang, and Ming Yang. 2021. "Dynamic Contracting with Flexible Monitoring." Working Paper. (Cit. on p. 6).

DeMarzo, Peter M., and Michael J. Fishman. 2007a. "Agency and Optimal Investment Dynamics." *Review of Financial Studies* 20 ([1]): 151–188. (Cit. on pp. 5, 14).

— . 2007b. "Optimal Long-Term Financial Contracting." *Review of Financial Studies* 20:2079–2128. (Cit. on pp. 5, 14, 41).

DeMarzo, Peter, and Yuliy Sannikov. 2006. "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model." *Journal of Finance* 61 ([6]): 2681–2724. (Cit. on pp. 1, 4, 5, 8, 12–14, 16, 17, 29, 30).

Demougin, Dominique, and Claude Fluet. 2001. "Monitoring versus Incentives." *European Economic Review* 45 ([9]): 1741–1764. (Cit. on p. 5).

Demsetz, Harold, and Kenneth Lehn. 1985a. "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy* 93 ([6]): 1155–1177. (Cit. on p. 1).

— . 1985b. "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy* 93 ([6]): 1155–1177. (Cit. on p. 7).

Dicks, David L. 2012. "Executive Compensation and the Role for Corporate Governance Regulation." *Review of Financial Studies* 25 ([6]): 1971–2004. (Cit. on p. 2).

Dumas, Bernard. 1991. “Super Contact and Related Optimality Conditions.” *Journal of Economic Dynamics and Control* 15:675–685. (Cit. on p. 15).

Dupire, Bruno. 1994. “Pricing with a Smile.” *Risk* 7 ([1]): 18–20. (Cit. on p. 2).

Dye, Ronald A. 1986. “Optimal Monitoring Policies in Agencies.” *RAND Journal of Economics* 17 ([3]): 339–350. (Cit. on p. 5).

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. 2017. “Executive Compensation: A Survey of Theory and Evidence.” In *The Handbook of the Economics of Corporate Governance*, ed. by Benjamin E. Hermalin and Michael S. Weisbach, 1:383–539. North-Holland. (Cit. on p. 4).

Fahlenbrach, Rüdiger. 2009. “Shareholder Rights, Boards, and CEO Compensation.” *Review of Finance* 13 ([1]): 81–113. (Cit. on p. 2).

Fauver, Larry, Mingyi Hung, Xi Li, and Alvaro G. Taboada. 2017. “Board Reforms and Firm Value: Worldwide Evidence.” *Journal of Financial Economics* 125 ([1]): 120–142. (Cit. on p. 7).

Feng, Felix Zhiyu, and Mark M. Westerfield. 2020. “Dynamic Resource Allocation with Hidden Volatility.” Working Paper. (Cit. on p. 7).

Fernandes, Nuno, Miguel Ferreira, Pedro Matos, and Kevin Murphy. 2013. “Are US CEOs Paid More? New International Evidence.” *Review of Financial Studies* 26:323–367. (Cit. on p. 2).

Georgiadis, George, and Balazs Szentes. 2020. “Optimal Monitoring Design.” *Econometrica* 88 ([5]): 2075–2107. (Cit. on p. 5).

Gibbons, Robert, and Kevin J. Murphy. 1992. “Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence.” *Journal of Political Economy* 100 ([3]): 468–505. (Cit. on p. 8).

Gompers, Paul, Joy Ishii, and Andrew Metrick. 2003. “Corporate Governance and Equity Prices.” *Quarterly Journal of Economics* 118 ([1]): 107–156. (Cit. on pp. 1, 3, 23).

Grossman, Sanford J., and Oliver D. Hart. 1983. “An Analysis of the Principal-Agent Problem.” *Econometrica* 51 ([1]): 7–46.

Hermalin, Benjamin E. 2013. “Corporate Governance.” In *Handbook of Organizational Economics*, ed. by Robert Gibbons and John Roberts, 732–763. Princeton University Press. (Cit. on p. 4).

— . 2010. *Understanding Firm Value and Corporate Governance*. Tech. rep. UC Berkeley. (Cit. on pp. 2, 4).

Himmelberg, Charles P., R. Glenn Hubbard, and Darius Palia. 1999. “Understanding the Determinants of Managerial Ownership and the Link Between Ownership and Performance.” *Journal of Financial Economics* 53 ([3]): 353–384. (Cit. on p. 7).

Holmström, Bengt. 1979. “Moral Hazard and Observability.” *Bell Journal of Economics* 10 ([1]): 74–91.

Holmström, Bengt, and Jean Tirole. 1993. “Market Liquidity and Performance Monitoring.” *Journal of Political Economy* 101 ([4]): 678–709. (Cit. on pp. 6, 9).

Hoskisson, Robert E., Mark W. Castleton, and Michael C. Withers. 2009. “Complementarity in Monitoring and Bonding: More Intense Monitoring Leads to Higher Executive Compensation.” *Academy of Management Perspectives* 23 ([2]): 57–74. (Cit. on p. 8).

Jensen, Michael. 1989. “The Eclipse of the Public Corporation.” *Harvard Business Review*: 61–74. (Cit. on p. 19).

Jost, Peter-Jürgen. 1991. “Monitoring in Principal-Agent Relationships.” *Journal of Institutional and Theoretical Economics* 147 ([3]): 517–538. (Cit. on p. 5).

Kaplan, Steven N. 2012. “Executive Compensation and Corporate Governance in the United States: Perceptions, Facts, and Challenges.” *Cato Papers on Public Policy* 2:99–157. (Cit. on p. 4).

Kaplan, Steven, and Joshua Rauh. 2010. “Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?” *Review of Financial Studies* 23 ([3]): 1004–50. (Cit. on p. 2).

Krishna, R. Vijay, Giuseppe Lopomo, and Curtis R. Taylor. 2013. “Stairway to Heaven or Highway to Hell: Liquidity, Sweat Equity, and the Uncertain Path to Ownership.” *RAND Journal of Economics* 44 ([1]): 104–127. (Cit. on pp. 14, 41).

Leung, Raymond C. W. 2017. “Continuous-Time Principal-Agent Problem with Drift and Stochastic Volatility Control.” Working Paper. (Cit. on p. 7).

Li, Anqi, and Ming Yang. 2020. “Optimal Incentive Contract with Endogenous Monitoring Technology.” *Theoretical Economics* 15 ([3]): 1135–1173. (Cit. on p. 5).

McConnell, John J., and Henri Servaes. 1990. “Additional Evidence on Equity Ownership and Corporate Value.” *Journal of Financial Economics* 27 ([2]): 595–612. (Cit. on p. 7).

Milgrom, Paul R., and John Roberts. 1992. *Economics, Organizations, and Management*. NJ, USA: Prentice Hall. (Cit. on pp. 5, 6, 9, 39, 41).

Milgrom, Paul, and Ilya Segal. 2002. “Envelope Theorems for Arbitrary Choice Sets.” *Econometrica* 70 ([2]): 583–601. (Cit. on p. 33).

Morck, Randall, Andrei Shleifer, and Robert W. Vishny. 1988. “Management Ownership and Market Valuation: An Empirical Analysis.” *Journal of Financial Economics* 20:293–315. (Cit. on p. 7).

Murphy, Kevin J. 1999. “Executive Compensation.” Chap. 38 in *Handbook of Labor Economics*, ed. by Orley C. Ashenfelter and David Card, vol. 3, Part B, 2485–2563. Amsterdam: North-Holland. (Cit. on p. 4).

Noe, Thomas H., and Michael J. Rebello. 2012. “Optimal Corporate Governance and Compensation in a Dynamic World.” *Review of Financial Studies* 25 ([2]): 480–521. (Cit. on p. 6).

Orlov, Dmitry. 2018. "Frequent Monitoring in Dynamic Contracts." Working Paper. (Cit. on p. 6).

Palia, Darius. 2001. "The Endogeneity of Managerial Compensation in Firm Valuation: A Solution." *Review of Financial Studies* 14 ([3]): 735–764. (Cit. on p. 7).

Piskorski, Tomasz, and Mark M. Westerfield. 2016. "Optimal Dynamic Contracts with Moral Hazard and Costly Monitoring." *Journal of Economic Theory* 166:242–281. (Cit. on p. 6).

Ross, Stephen A. 1973. "The Economic Theory of Agency: The Principal's Problem." *American Economic Review* 63 ([2]): 134–139.

Sannikov, Yuliy. 2008. "A Continuous-Time Version of the Principal-Agent Problem." *Review of Economic Studies* 75 ([3]): 957–984. (Cit. on pp. 12, 29).

Stokey, Nancy L., Robert E. Lucas Jr, and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, Ma: Harvard University Press. (Cit. on p. 42).

Strausz, Roland. 1997. "Delegation of Monitoring in a Principal-Agent Relationship." *Review of Economic Studies* 64 ([3]): 337–357. (Cit. on p. 5).

Tirole, Jean. 2006. *The Theory of Corporate Finance*. Princeton, NJ: Princeton University Press. (Cit. on pp. 5, 6, 9).

Varas, Felipe, Iván Marinovic, and Andrzej Skrzypacz. 2020. "Random Inspections and Periodic Reviews: Optimal Dynamic Monitoring." *Review of Economic Studies* 87 ([6]): 2893–2937. (Cit. on p. 6).

Ward, Amy R., and Peter W. Glynn. 2003. "Properties of the Reflected Ornstein-Uhlenbeck Process." *Queueing Systems* 44 ([2]): 109–123. (Cit. on p. 16).

Wintoki, M. Babajide, James S. Linck, and Jeffry M. Netter. 2012. "Endogeneity and the Dynamics of Internal Corporate Governance." *Journal of Financial Economics* 105 ([3]): 581–606. (Cit. on p. 7).

Wright, Peter, and Mark Kroll. 2002. "Executive Discretion and Corporate Performance as Determinants of CEO Compensation, Contingent on External Monitoring Activities." *Journal of Management and Governance* 6 ([3]): 189–214. (Cit. on p. 8).