Ambiguity in Dynamic Contracts*

Martin Szydlowski[†] Ji Hee Yoon[‡]

February 23, 2021

Abstract

We study a continuous-time principal-agent model in which the principal is ambiguity averse about the agent's effort cost. The robust contract generates a seemingly excessive pay-performance sensitivity. The worst-case effort cost is high after good performance, but low after bad performance, which leads to *overcompensation* and *undercompensation* respectively and provides a new rationale for performance-sensitive debt. We also characterize the agent's incentives when the contract is misspecified, i.e., he is offered the robust contract, but his true effort cost differs from the worst case. Then, termination can induce shirking, the strength of incentives is hump-shaped, and agents close to firing prefer riskier projects, while those close to getting paid prefer safer ones. This feature resembles careers in organizations, most notably risk-shifting and the quiet life.

Keywords: Principal-Agent model, ambiguity, continuous time, overcompensation

JEL Codes: C61, D86, M12, M53

1 Introduction

Firms need to align the interests of managers and shareholders, but the environment they operate in is prone to change. Events such as the entry of competitors, the arrival of a

*We thank the editor and two anonymous referees for constructive feedback that helped us improve our paper. We also thank Sylvain Chassang, Eddie Dekel, Daniel Garrett, Paolo Fulghieri, Wojchiech Olszewski, and Marciano Siniscalchi for helpful discussions, as well as conference participants at the North American Summer Meeting of the Econometric Society, the SFS Cavalcade, and the World Congress of the Game Theory Society. Szydlowski thanks Jeff Ely, Michael Fishman, and Bruno Strulovici for their guidance.
First Version: January 2012.

†Corresponding author. University of Minnesota, 321 19th Ave S, Minneapolis, MN 55455, USA. Email: szydl002@umn.edu.

‡University College London, Department of Economics, Drayton House, 30 Gordon St, Kings Cross, London WC1H 0AX, United Kingdom. Email: jihee.yoon@ucl.ac.uk.

new technology, or a shift in customer demands are often impossible to anticipate. Their impact on an individual manager's productivity and his daily duties can be intricate, which makes it difficult to quantify and to communicate to others.[1] Consequently, firms write employment contracts with little knowledge about how the relationship with the manager and the contract's ability to provide incentives may change in the future.

In this paper, we model the firm's lack of knowledge as ambiguity and we develop a theory of robust dynamic contracting. The firm (the principal) does not know the arrival times, the impact, or the probability distribution of shocks to the environment. It only knows the range which particular realizations may take, and offers the manager (the agent) a contract which provides incentives and maximizes its payoff under the worst case. We characterize the impact of ambiguity aversion on the shape of the optimal contract, and derive the dynamics of the worst-case process. Our main example is when the ambiguity is about the manager's effort cost. For instance, the difficulty of his work could vary depending on the competitive pressure faced by the firm or his own abilities may change over time.

The optimal contract has several novel features, which are all driven by the dynamics of the worst case effort cost. First, the contract is divided into an over- and an undercompensation region. After sufficiently high performance, the worst case effort cost is high, and, in expectation, the agent receives higher payouts than in the case without ambiguity. After low performance, the worst case is that his effort cost is low, and the agent is undercompensated. This result is driven by the dynamics of the worst case, which changes depending on how close the contract is to termination. Hence, our paper provides a new explanation for why high performing managers receive seemingly excessive compensation, a question which has received significant attention in both the popular press and the academic literature.[2] In our setting, the reason is not managerial power (e.g. Zwiebel (1996)), board capture (e.g. Hermalin and Weisbach (1988)), or the exploitation of information rents, (e.g. Jensen (1986)), but the principal's ambiguity aversion.

Since both over- and undercompensation regions arise in the same optimal contract, our model generates career trajectories. In expectation, a manager with a good track record will continue to receive excessive compensation unless he encounters bad luck in the form of a path of negative outcomes. Although a manager with a bad record can expect to move up to the overcompensation region as long as he exerts effort, the likelihood of advancement is lower than without ambiguity, while the chance of being fired is higher. Similarly, a young manager starts out undercompensated but on average reaches the overcompensation region

[1] The same holds true for changes on a smaller scale, such as the manager's innate ability or the quality of his match with the firm.

[2] See Bebchuk and Fried (2006) and Gabaix and Landier (2008) for two examples.

once enough time has passed. The speed of this advancement depends on his performance record. High performing managers reach the overcompensation region faster, while low performing ones may reach it more slowly and therefore remain undercompensated for long periods of time, or get fired. These features resemble seniority and entrenchment, but they arise because of the interaction of ambiguity aversion and incentive provision.

Ambiguity aversion also introduces a disconnect between the manager's pay-performance sensitivity (PPS) and his current effort cost. In a dynamic contract without ambiguity, the pay-performance sensitivity is proportional to the current effort cost.[3] In our setting, the firm is forced to set the pay-performance sensitivity at the highest level at all times, since it expects the manager to shirk under the worst case otherwise. Essentially, ambiguity aversion generates a precautionary motive for the principal and the manager receives excessive incentives, compared to both the contract without ambiguity and to his realized effort cost. This result offers a new answer to the puzzle raised by Murphy (2003). They find that many managers receive stock options, even when their individual impact on the firm does not seem to be large enough to warrant them.

Finally, the optimal contract can be implemented with performance-sensitive debt. A changing interest rate is necessary because the drift of the agent's continuation value changes under the worst case. The performance sensitive debt is hence used to adjust the firm's cash flows in the over- and undercompensation regions. This interpretation is new and differs from current justifications for performance sensitive debt such as Piskorski and Tchistyi (2011). The implementation consisting of equity and credit lines in DeMarzo and Sannikov (2006) is no longer optimal.

Our notion of ambiguity corresponds to "Type I Ambiguity" in Hansen and Sargent (2012). In that paper, a Ramsey planner does not know the "true model," and believes that the private sector knows the true model. For any policy, the planner evaluates *both* her own payoffs and the private sector's incentives under the worst-case model, which reflects her concerns about robustness and her belief that the private sector knows the true model. The analog holds in our paper. The agent knows the true evolution of the effort cost (represented by a probability measure) and the principal chooses a contract which maximizes her payoff under the worst-case measure. As in Hansen and Sargent (2012), both the principal's and the agent's payoffs are evaluated under the worst-case measure. This setting has an intuitive interpretation as a game between a principal, an agent, and a malevolent nature. Each agent is endowed with an effort cost process, which represents how well his skills are suited to the principal's project. The principal posts a contract, but is uncertain about which agent she is being matched with. A malevolent nature chooses the match between principal and agent,

[3]This result holds in DeMarzo and Sannikov (2006) and Sannikov (2008).

to minimize the principal's value.[4]

Alternatively, we can understand the contract as being optimal under the principal's *subjective* preferences. Then, the contract may fail to provide incentives if the true measure differs from the principal's worst case (see Hansen and Sargent (2012), p. 432 for a discussion). We study this interpretation in Section 5. There, we alter the model and assume that the agent evaluates his payoffs under a reference measure, while the principal remains ambiguity averse. Generally, the reference measure differs from the principal's worst case, so the principal and agent disagree about the evolution of effort costs. This setting corresponds to "Type II Ambiguity" in Hansen and Sargent (2012).[5]

This alternative setting yields novel predictions. Contrary to a long line of literature on dynamic contracts,[6] termination does not motivate the agent to work and instead may induce shirking. Intuitively, the agent's value from exerting effort depends on the payments he is promised to collect in the future. However, when the firing probability is high, the agent is unlikely to collect on the promised rewards, and thus the value of working is low. If the effort cost under the true measure is sufficiently high, the agent shirks. Importantly, shirking occurs even though the pay-performance sensitivity is high.

When the agent shirks, his continuation value drifts downwards. Thus, he expects to be fired as time passes and only a sequence of lucky realizations of the Brownian noise allows him to collect payments. In line with this intuition, agents who shirk are risk-loving. That is, they prefer a project with higher volatility in output, or equivalently a contract with a higher PPS. By contrast, agents who exert effort are risk averse. They can expect to enter and then to stay in the overcompensation region forever if the noise in the output process is sufficiently small, and thus they will always prefer to bear less risk. These results are again driven by the principal's ambiguity aversion. Without ambiguity, by contrast, the agent is always risk-neutral in the PPS. Thus, the misspecified contract provides incentives for risk-shifting at the bottom (Jensen and Meckling (1976)) and conservatism at the top (Bertrand and Mullainathan (2003)). That is, if the agent could select the riskiness of the project, he would choose a risky project when his continuation value is low, and a safe project when his continuation value is high.

Our results crucially depend on the fact that ambiguity is about the effort cost. To show

[4] See Section 3.3 for a more detailed discussion and for more alternatives.

[5] In their paper, the planner is ambiguity averse but the private sector trusts the approximating model under Type II Ambiguity. Similarly, in Section 5, the principal is ambiguity averse, but the agent evaluates his payoffs under the reference measure.

[6] Theoretical works which argue that firing provides incentives to the agent are numerous, and include Spear and Wang (2005), Wang (2011), Garrett and Pavan (2012), Fong and Li (2017), as well as a number of works in inspired by Sannikov (2008), such as Biais et al. (2010), DeMarzo and Sannikov (2006), DeMarzo et al. (2012) and He (2009).

this, we study ambiguity about the firm's average productivity in Section 6.1. The features we have described above do not appear, because ambiguity about the productivity does not interact with providing incentives. As a result, the worst-case productivity is static and at the lowest possible level. The PPS is proportional to the agent's effort cost and there are no over- and undercompensation regions. In the misspecified contract, the agent still exerts effort and the implementation of DeMarzo and Sannikov (2006) is optimal. In both the main model and Section 5, the principal is ambiguity averse while the agent is not. In Section 6.3 we extend our model so that *both* the principal and the agent are ambiguity averse. While agent's worst-case differs from the principal's worst-case, the results are qualitatively similar to the ones in Section 5.

2 Literature

Our paper complements a long line of literature on dynamic contracts.[7] It is, to the best of our knowledge, the first to introduce ambiguity aversion in a continuous-time principal-agent framework. In existing work, the defining features of the relationship between principal and agent, such as effort cost or productivity, are either constant,[8] or the principal knows their objective probability law (see Marinovic and Varas (2019)), which implies that she knows the likelihood of both the timing and the realization of changes ex ante.[9] We model ambiguity via equivalent changes of measures with respect to a reference probability. Uncertainty is represented by Brownian Motions (see Chen and Epstein (2002)). The principal has a set of priors and uses the maxmin criterion to determine the worst case probability and the value of any dynamic contract. The paper thus falls into the class of maxmin expected utility (MMEU) models (see Gilboa and Schmeidler (1989)).[10]

The current paper is an updated version of an earlier draft, Szydlowski (2012). A number of papers which cite this draft have appeared. Miao and Rivera (2016) study a continuous-time contracting problem with ambiguity where nature affects the drift, but is penalized by an entropy cost, Dumav (2017) studies a problem where nature affects the drift without such penalty, but where both principal and agent are ambiguity averse, and Sung (2015)

[7]Seminal papers include Spear and Srivastava (1987), Holmström and Milgrom (1987) and Laffont and Tirole (1988). Continuous-time models have been studied in Sannikov (2008) and DeMarzo and Sannikov (2006).

[8]This assumption is very common, and found for example in Spear and Srivastava (1987), Holmström and Milgrom (1987) and Sannikov (2008).

[9]Often, it is assumed that the realizations are the agent's private information. This is for example the case in Battaglini (2005), Garrett and Pavan (2012) and Garrett and Pavan (2015). See also Giat et al. (2010), Prat and Jovanovic (2014), and He et al. (2017) for the case when principal holds a subjective belief, and disagrees with the agent about the realizations.

[10]See also Zhu (2016), who studies incomplete contracts in a discrete-time setting with maxmin preferences.

studies a problem where nature affects both drift and volatility. Our results on over- and undercompensation, performance sensitive debt, and the agent's incentives under the misspecified contract do not appear in these papers. Miao and Riviera show that ambiguity aversion about the drift may also lead to excessively high PPS. However, their result hinges on having an entropy cost and vanishes without this assumption, as we show in Section 6.1, whereas ours is driven by the changing cost of incentives. Dumav's result about the PPS is opposite - the PPS is lower than without ambiguity. This is because in his model, effort is continuous and the optimal effort level is lower under ambiguity aversion. Finally, Sung finds that if both principal and agent are ambiguity averse about the realized volatility, then the optimal PPS is independent of that volatility. This result mirrors ours, because in our model the PPS is independent of the "realized" effort cost, which changes over time. Also related is Dicks and Fulghieri (2018), who study a Holmström and Milgrom (1987)-type framework with ambiguity aversion and multiple agents. They show that cross-pay is optimal because it allows agents to hedge ambiguity.

The MMEU framework has been used extensively in static contracting. Garrett (2014) studies a procurement contract, and shows that ambiguity about the agent's preferences leads to the optimality of simple incentive schemes. Bergemann and Schlag (2008) and Bergemann and Schlag (2011) study a seller problem in which the principal is ambiguity averse about the buyer's valuation. Di Tillio et al. (2016) demonstrate the optimality of a contract which has ambiguous outcomes when the ambiguity is on the agent's side.

3 Model

We study a dynamic principal agent problem. Time is continuous, infinite, and indexed by $t \geq 0$. The agent operates a single project for the principal, which yields payoffs according to a diffusion process whose drift depends on the agent's effort. Formally, there exists a Brownian motion $B = \{B_t\}_{t\geq 0}$ on the filtered probability space $(\Omega, \mathcal{F}, P)$ with filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions.[11] The process $X = \{X_t\}_{t\geq 0}$ is given by

$$dX_t = \mu dt + \sigma dB_t, \quad X_0 = 0, \tag{1}$$

[11]See e.g. Karatzas and Shreve (1991) for standard concepts in stochastic analysis. Throughout the paper, we use plain letters (e.g. X, a, θ) to denote a stochastic process and subscripts (X_t, a_t, θ_t) to denote the value at time t. All inequalities involving random variables or stochastic processes are understood to hold P-almost surely. All measures used in the paper will be absolutely continuous with respect to P.

which represents the output when the agent exerts effort at all times $t \geq 0$.[12] We denote the filtration generated by X with $\mathcal{F}^X = \left\{\mathcal{F}_t^X\right\}_{t\geq 0}$. Since the drift and diffusion coefficients are constants, the stochastic differential equation (1) has a unique strong solution and the filtration generated by X coincides with the filtration generated by B, i.e. $\mathcal{F}^X = \mathcal{F}^B$.[13]

The agent's effort $a = \{a_t\}_{t\geq 0}$, with $a_t \in \{0,1\}$, generates a flow cost of $a_t d\kappa_t$ to the agent, where $\kappa = \{\kappa_t\}_{t>0}$ follows

$$d\kappa_t = \theta_0 dt + dZ_t, \quad \kappa_0 = 0. \tag{2}$$

Here, $\theta_0 \in \left[\underline{\theta}, \bar{\theta}\right] \subset (0,\mu)$ is constant and common knowledge,[14] $Z = \{Z_t\}_{t\geq 0}$ is a Brownian motion on $(\Omega, \mathcal{F}, P)$, and Z is independent of B. The filtration generated by κ coincides with the filtration generated by Z, i.e., $\mathcal{F}^\kappa = \mathcal{F}^Z$, and Equation (2) has a unique strong solution.

The pair (B, Z) is a two-dimensional Brownian Motion on the probability space $(\Omega, \mathcal{F}, P)$. We denote the filtration generated by (B, Z) as $\mathcal{F}^{B,Z}$. This filtration coincides with the filtration generated by the output process X and the effort cost process κ: $\mathcal{F}^{B,Z} = \mathcal{F}^B \times \mathcal{F}^Z = \mathcal{F}^{X,\kappa} = \mathcal{F}^X \times \mathcal{F}^\kappa$.

3.1 Change of Measures

The principal is ambiguity averse about the agent's effort cost. Following Chen and Epstein (2002), her beliefs consist of a set of probability measures which are mutually absolutely continuous with respect to the reference measure P over any finite time interval. These measures are constructed using density generators and the Girsanov Theorem.[15] To this end, we introduce a stochastic process $\theta = \{\theta_t\}_{t\geq 0}$ which represents the evolution of the agent's *average* flow cost. We call θ *admissible* if it is $\mathcal{F}^X$-progressively measurable, square integrable, and satisfies $\underline{\theta} \leq \theta_t \leq \bar{\theta}$.[16] We denote the set of admissible θ with Θ. Throughout the paper, we use the notation $\theta \in \mathcal{F}^X$ to denote that θ is $\mathcal{F}^X$-progressively measurable, etc. We use the terms *progressively measurable* and *progressive* interchangeably.

[12] We model shirking in Section 3.1 via Girsanov's change of measure.

[13] See e.g. Karatzas and Shreve (1991), Ch. 5.2, p. 284ff. We have $X_t = \mu t + \sigma B_t$, which immediately implies that the filtrations are identical. An equivalent setup is used in Cvitanic and Zhang (2012), p. 47.

[14] If $\bar{\theta} \geq \mu$, then the optimal contract is trivial, since under the worst-case measure, incentivizing effort is too costly even in the first best. The assumption that $\underline{\theta} \geq 0$ guarantees that the agent has a disutility of effort for any θ.

[15] See e.g. Karatzas and Shreve (1991), Th. 5.1, p. 191.

[16] Recall that we understand all relations to hold P-almost surely.

For any $\theta \in \Theta$, a P-martingale $z_0^t(\theta)$ is determined by

$$z_0^t(\theta) = \exp\left(-\int_0^t (\theta_0 - \theta_s)\, dZ_s - \frac{1}{2}\int_0^t (\theta_0 - \theta_s)^2\, ds\right)$$

and we can define a measure P^θ via $dP^\theta/dP\big|_{\mathcal{F}_t} \equiv z_0^t(\theta)$. The effort cost (2) can be written as

$$d\kappa_t = \theta_t dt + dZ_t^\theta, \quad \kappa_0 = 0, \tag{3}$$

where $dZ_t^\theta = (\theta_0 - \theta_t)dt + dZ_t$ is a Brownian Motion under P^θ. The set of probability measures which can be generated using Θ, $\mathcal{P}^\Theta = \left\{P^\theta : \theta \in \Theta\right\}$, is rectangular.[17]

Following Sannikov (2008), we represent different effort choices of the agent via the Girsanov Theorem. An effort process $a = \{a_t\}_{t\geq 0}$, with $a_t \in \{0,1\}$, is admissible if it is $\mathcal{F}^{X,\kappa}$-progressively measurable and square integrable. Any admissible effort process changes the measure P into P^a, which is determined by the P-martingale $\xi_0^t(a)$ with

$$\xi_0^t(a) = \exp\left(-\int_0^t \frac{\mu}{\sigma}(1-a_s)\, dB_s - \frac{1}{2}\int_0^t \left(\frac{\mu}{\sigma}(1-a_s)\right)^2 ds\right),$$

and $dP^a/dP|_{\mathcal{F}_t} = \xi_0^t(a)$. Then, the output process (1) can be written as

$$dX_t = \mu a_t dt + \sigma dB_t^a, \quad X_0 = 0, \tag{4}$$

where $dB_t^a = \frac{\mu}{\sigma}(1-a_t)dt + dB_t$ is a Brownian Motion under P^a.

Given a pair of martingales $\xi_0^t(a)$ and $z_0^t(\theta)$ on $(\Omega, \mathcal{F}, P)$, the process

$$z_0^t(a,\theta) = \xi_0^t(a) z_0^t(\theta)$$

is also a martingale. We define the measure $P^{a,\theta}$ via $dP^{a,\theta}/dP\big|_{\mathcal{F}_t} = z_0^t(a,\theta)$. Lemma 1 below establishes that the pair (B_t^a, Z_t^θ) is a two-dimensional Brownian motion on $(\Omega, \mathcal{F}, P^{a,\theta})$. In particular, B^a and Z^θ are independent under $P^{a,\theta}$. The proof is in Appendix A.1.

Lemma 1 (Brownian Motion under $P^{a,\theta}$)**.** *If (B, Z) is a two-dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$, then (B^a, Z^θ) is a two-dimensional Brownian motion on $(\Omega, \mathcal{F}, P^{a,\theta})$. In particular, B^a and Z^θ are independent under $P^{a,\theta}$.*

Lemma 1 implies that

$$E^{a,\theta}\left[a_t d\kappa_t | \mathcal{F}_t^{X,\kappa}\right] = a_t \theta_t dt,$$

[17] This is analogous to the κ-ignorance case in Chen and Epstein (2002). See Equation 2.11 and Section 3.3 in that paper.

where $E^{a,\theta}[.]$ denotes the expectation under $P^{a,\theta}$. Thus, we can interpret $d\kappa_t$ as the *realized* flow effort cost at time t and θ_t as the *average* flow effort cost. We can interpret P^{θ} as the *distribution* of the agent's effort costs. Since the agent does not observe $\{d\kappa_s\}_{s\geq t}$ when choosing a_t,[18] only θ_t matters for his effort decision. Thus, we refer to θ_t simply as the "effort cost" throughout the paper.

3.2 Principal's and Agent's Problems

Observability Output X is observable to both principal and agent, while effort a and realized flow costs κ are private to the agent. The drift of the effort cost θ in Equation (3)— equivalently, the measure $P^{a,\theta}$— is known to the agent but unknown to the principal. Since effort is $\mathcal{F}^{X,\kappa}$-progressive and since θ is $\mathcal{F}^{X}$-progressive, the agent does not know $\{\theta_s\}_{s>t}$ when choosing effort a_t at time t. Intuitively, the agent knows the process θ as a functional of X, but does not know its realizations in advance. Likewise, the agent does not observe the increment $d\kappa_t$ or future realizations $\{\kappa_s\}_{s>t}$ when choosing effort at time t.[19]

Contracts The principal commits to a contract α, which she offers to the agent. An *admissible* contract consists of a pair of $\mathcal{F}^X$-progressively measurable and square integrable stochastic processes $(\hat{a}, c) = \{\hat{a}_t, c_t\}_{t\geq 0}$ and a $\mathcal{F}^X$-stopping time τ. Additionally, c is right continuous with left limits and satisfies $dc_t \geq 0$. We denote the space of admissible contracts as $\mathcal{A}$. Here, $\hat{a}$ is the *recommended* effort that is determined by the principal, given her observations of X. In equilibrium, where the incentive compatibility of the agent holds, the recommended effort $\hat{a}$ is indeed the agent's optimal effort. The agent's cumulative consumption is given by $c_t = \int_0^t dc_s$ and the restriction $dc_t \geq 0$ reflects the agent's limited liability.[20] The stopping time τ indicates when the agent is fired, in which case the firm is shut down, and the principal receives a scrap value of $L \in [0, (\mu - \bar{\theta})/r)$, while the agent receives an outside value normalized to zero.[21]

It is worth to remark that we rule out screening contracts in which the principal offers the agent a menu to get him to reveal the effort cost. Given the richness of the space $\mathcal{P}^{\theta}$, such screening contracts may be too costly to write or enforce.[22] Additionally, while $\hat{a}$ is

[18]Recall that a is $\mathcal{F}^{X,\kappa}$-progressively measurable.

[19]Intuitively, at time t, the agent does not know the increment $\kappa_{t+h} - \kappa_t$ for any $h > 0$, since that increment is random and driven by a Brownian motion.

[20]This rules out trivial contracts in which the principal can incentivize effort by exacting arbitrarily severe punishments upon observing a history of low output. The qualitative features of the optimal contract remain unchanged if we assume $dc_t \geq -\underline{c}dt$ for some finite $\underline{c} > 0$.

[21]If $L \geq (\mu - \bar{\theta})/r$, then the principal would prefer to liquidate the project immediately in the first best, i.e. when effort is observable.

[22]It is well known that screening contracts are not feasible in the setting of DeMarzo and Sannikov (2006),

$\mathcal{F}^X$-progressive in equilibrium, *off equilibrium* $\mathcal{F}^{X^a}$, the filtration generated by output X^a given the actual effort process a, and $\mathcal{F}^B$ do not necessarily coincide, i.e., $\mathcal{F}^{X^a} \neq \mathcal{F}^B$. This is because the actual effort a is chosen by the agent who observes (X_t, κ_t), or equivalently (B_t, Z_t), and hence, a can be $\mathcal{F}^B \times \mathcal{F}^Z$-progressive but not progressive w.r.t $\mathcal{F}^X$ or equivalently $\mathcal{F}^B$. Lemma 2 below proves that it is without loss of generality to restrict attention to effort $a \in \mathcal{F}^X$. Intuitively, choosing effort that is adapted to a larger filtration than $\mathcal{F}^X$ is never strictly optimal for the agent.

Principal's Problem Both principal and agent are risk neutral. The agent is more impatient than the principal and discounts time at rate $\gamma > r$.[23] He receives consumption payments from the principal and incurs the effort cost when $a_t = 1$. The agent's continuation value at $t \geq 0$ given $\alpha \in \mathcal{A}$ and $\theta \in \Theta$ is

$$W_t(\alpha, \theta) \equiv E^{a,\theta}\left[\int_t^\tau e^{-\gamma(s-t)}(dc_s - a_s d\kappa_s)\big|\mathcal{F}_t^{X,\kappa}\right]. \tag{5}$$

The principal is ambiguity averse about the effort cost θ. In modeling ambiguity aversion, we follow Hansen and Sargent (2012)'s notion of "Type I Ambiguity." Specifically, the principal believes that the agent knows the *true* measure in $\mathcal{P}^\Theta$, which governs the evolution of θ and which is unknown to the principal. The principal then evaluates each contract α under the worst-case measure $P^{\hat{a},\theta(\alpha)}$, which minimizes her value. Importantly, the principal evaluates both her own and the agent's payoffs under the worst case, since she believes that the agent knows the true measure. For any incentive compatible contract, the agent's actual effort a is the same as the recommended effort $\hat{a}$, so that $P^{a,\theta(\alpha)} = P^{\hat{a},\theta(\alpha)}$.

This specification has an intuitive interpretation as a game between the principal, the agent, and a malevolent nature. First, the principal posts the contract α. Then, nature chooses θ, which can be interpreted as the match quality between the agent's skills and the principal's project. Then, given both α and θ, the agent chooses his effort optimally.[24]

Formally, the optimal contracting problem is given by

$$J_0 = \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \Theta} E^{a,\theta}\left[\int_0^\tau e^{-rt}\left(dX_t - dc_t\right)\right] \tag{6}$$

especially given the generality with which we model the process θ.

[23]Given that the utilities of both principal and agent are linear, the principal would optimally choose to defer payments indefinitely into the future if the agent were not impatient.

[24]See the discussion in Section 3.3 for more details.

subject to the incentive compatibility condition

$$W_0(\alpha, \theta(\alpha)) \geq W_0(\tilde{\alpha}, \theta(\alpha))$$

for $\tilde{\alpha} = \left\{\{\tilde{a}_t, c_t\}_{t\geq 0}, \tau\right\}$, where $\tilde{a} = \{\tilde{a}_t\}_{t\geq 0} \in \mathcal{F}^{X,\kappa}$ is the agent's effort when deviating, and the participation constraint

$$W_0(\alpha, \theta(\alpha)) \geq 0.$$

Here, $\theta(\alpha) \in \Theta$ is the worst case for the principal given contract α, which arises from the minimization in Equation (6). To cleanly isolate the effect of ambiguity aversion, we restrict attention to contracts which always induce effort, i.e. $a_t = 1$ for all $t < \tau$. We state a sufficient condition so this is optimal in Section B.1.

3.3 Discussion

Ambiguity Aversion Our definition of ambiguity aversion follows "Type I Ambiguity" in Hansen and Sargent (2012), which considers a Ramsey problem. Under Type I ambiguity, the Ramsey planner believes that "the private sector knows a correct probability specification linked to the planner's approximating model [...] that is unknown to the Ramsey planner but known by the private sector." Then, the planner uses the worst-case model to calculate the robust optimal policy. In particular, she calculates both the private sector's incentives and her own payoffs under the worst-case.

In our setting, the "approximating model" is represented by the measure P and the "worst-case model" for any contract α is represented by the measure $P^{a,\theta(\alpha)}$. As in Hansen and Sargent (2012), the agent knows the true model and both the agent's and the principal's payoffs are evaluated under $P^{a,\theta(\alpha)}$. As in Hansen and Sargent (2012), the agent in our model is not ambiguity averse. Instead, when choosing his effort policy, the agent knows the stochastic process θ, but does not know its realizations in advance. In particular, since θ is adapted to the filtration generated by output, $\mathcal{F}^X$, at time t the agent knows none of the realizations θ_s for any $s > t$. However, for any realized path of output $\{X_s\}_{0\leq s\leq t}$, the agent knows what the realized path $\{\theta_s\}_{0\leq s\leq t}$ is, since θ is $\mathcal{F}^X$-progressive and the agent observes the path of output X.

Under "Type I Ambiguity," the measure $P^{a,\theta(\alpha)}$ can be interpreted as the *physical* (or *true*) measure which governs the evolution of X and κ. Alternatively we can interpret the principal's problem as deriving the robust optimal contract given her *subjective* preferences, which may fail to provide incentives whenever the *true* θ differs from the worst-case $\theta(\alpha)$ (see Hansen and Sargent (2012), p. 432, for a discussion of this interpretation). We consider this alternative in Sections 5 and 6.3. In Section 5, the reference measure P is the true measure which governs the evolution of κ and the agent is not ambiguity averse. While in the main model, both principal and agent evaluate the contract under $P^{a,\theta(\alpha)}$, in Section 5, the agent evaluates the contract under P^{a,θ_0}. That is, the principal and agent disagree and the agent believes that the "average effort cost" at each time t is θ_0. This specification corresponds to "Type II Ambiguity" in Hansen and Sargent (2012). In Section 6.3, the agent is also ambiguity averse and evaluates the contract under her own worst-case process θ^A, with associated measure P^{a,θ^A}. Generally, P^{a,θ^A} differs from $P^{a,\theta(\alpha)}$ and the principal and agent disagree about the evolution of κ.

Games vs. Nature As in Hansen et al. (2006), our setting can be interpreted as a game between the principal, the agent, and a malevolent nature. The principal first chooses the contract, then nature chooses a process θ to minimize the principal's value. Taking as given both the contract α and the process θ, the agent then chooses effort optimally.[25] Taking the agent's continuation value as the state variable, the contracting problem is then a zero-sum game between principal and nature (see Fleming and Souganidis (1989)). This setup allows for two possible interpretations: (1) The firm is endowed with an output process X and a "task difficulty process" κ. Nature can alter the distribution of the task difficulty P^θ to reduce the value to the principal. The agent then learns the process θ after being hired, but does not know its realizations in advance. (2) There are different agents, each endowed with a process θ, which represents the match quality between the agent's skills and the project. The principal posts a contract α and is then matched with an agent. Which agent the principal is matched with is chosen by nature.

Ambiguity via Girsanov's Theorem The role of the second Brownian Motion Z is to make our framework conform to the multiple-priors specification in Gilboa and Schmeidler (1989). We assume that the path of Z is not observed by the principal, so that the contract is $\mathcal{F}^X$-progressive, but not necessarily $\mathcal{F}^{X,\kappa}$-progressive. The increment dZ_t at any time t is independent of the agent's information, so without loss of generality, his optimal effort choice is also progressive w.r.t. $\mathcal{F}^X$ only.[26] It is thus intuitive that Z does not affect the worst case of the principal. We therefore assume that θ is $\mathcal{F}^X$-progressive and, in particular, is independent of Z.

[25] Importantly, the setup is common knowledge, so when choosing whether to accept contract α, the agent knows that nature's best response to α is $\theta(\alpha)$.

[26] We prove this below in Lemma 2.

4 Optimal Contract

Even though the principal is ambiguity averse, the continuation value approach still applies.[27] In particular, for a contract $\alpha \in \mathcal{A}$ and process $\theta \in \Theta$, it is optimal for the agent to choose effort which is $\mathcal{F}^X$-progressive, so that the agent's continuation value is also $\mathcal{F}^X$-progressive. Then, the agent's continuation value can be represented as a diffusion process with respect to the Brownian Motion B^a.

Lemma 2. *For any contract $\alpha \in \mathcal{A}$ and density generator $\theta \in \Theta$, the agent's optimal effort is $\mathcal{F}^X$-progressive without loss of generality. There exists an $\mathcal{F}^X$-progressive and square integrable process $\psi = \{\psi_t\}_{t\geq 0}$ such that the agent's continuation value satisfies*

$$dW_t = (\gamma W_t + \theta_t a_t)\, dt - dc_t + \psi_t dB_t^a. \tag{7}$$

Exerting effort at time t is incentive compatible whenever

$$\psi_t \geq \frac{\sigma}{\mu}\theta_t.$$

We can express the continuation value under measure $P^{a,\theta}$ in terms of the output process X as[28]

$$dW_t = \big(\gamma W_t + a_t\theta_t\big)dt - dc_t + \psi_t \frac{1}{\sigma}(dX_t - \mu a_t dt),$$

using Equation (4). When the agent always exerts effort, i.e. $a_t = 1$ for all $t \leq \tau$, the above equation reduces to

$$dW_t = \big(\gamma W_t + \theta_t\big)dt - dc_t + \psi_t \frac{1}{\sigma} dB_t,$$

which follows from Equation (1).[29] We can interpret the process ψ_t as the agent's pay-performance sensitivity, since it determines how strongly his continuation value reacts to changes in output. The key difference to DeMarzo and Sannikov (2006) is that the principal is ambiguity averse about the agent's effort cost and therefore about whether her incentives are strong enough for the agent to exert effort.

We now characterize the optimal contract. We show that the worst-case effort cost depends on the past performance of the agent and there are two regimes in the contract. For

[27] See e.g. Spear and Srivastava (1987) and Sannikov (2008).

[28] Generally, the agent's continuation value will evolve differently for different θ, since the cost of effort evolves differently under different measures $P^{a,\theta}$. We could highlight this dependency by writing the continuation value process as W_t^θ instead of W_t, but we will omit this for the sake of notation. To characterize the robust optimal contract, we will solve for the contract and the worst-case density generator, which we will call θ_t^*, simultaneously. This will implicitly determine the dynamics of the continuation value process under the worst case.

[29] Note that when the agent always exerts effort, then $B_t = B_t^a$ for all $t < \tau$.

high continuation values, the effort cost is at the highest level $\bar{\theta}$ while for low continuation values it is at $\underline{\theta}$. Ambiguity aversion leads the principal to choose the highest pay-performance sensitivity even when she believes the agent's effort cost is low.

To intuitively derive a solution to the principal's problem, consider the agent's incentives in Equation (7). On a small interval of time, the agent shirks whenever his effort cost θ_t is larger than the loss in continuation value $\frac{\mu}{\sigma}\psi_t$, i.e. $\theta_t > \frac{\mu}{\sigma}\psi_t$. If the contract has pay-performance sensitivity $\psi_t < \frac{\sigma}{\mu}\bar{\theta}$, the agent optimally shirks whenever θ_t is between $\psi_t\frac{\mu}{\sigma}$ and $\bar{\theta}$. For any pay-performance sensitivity ψ_t and effort cost θ_t, his optimal effort is thus

$$a\left(\psi_t, \theta_t\right) = \mathbb{1}\left\{\psi_t \geq \frac{\sigma}{\mu}\theta_t\right\}.$$

Since the principal values effort, whenever $\psi_t < \frac{\mu}{\sigma}\bar{\theta}$ the worst case is $\theta_t = \bar{\theta}$, which leads the agent to shirk. Thus, no contract with $\psi_t < \frac{\sigma}{\mu}\bar{\theta}$ can be incentive compatible. To prevent shirking, the principal must set the pay-performance sensitivity to

$$\bar{\psi} \equiv \frac{\sigma}{\mu}\bar{\theta},$$

independently of the effort cost, i.e. even when $\theta_t < \bar{\theta}$. This way, she ensures the agent works under the worst case.

However, there is another channel for ambiguity to affect the principal's value, because the contract is dynamic. The principal's value is generally hump-shaped in the continuation value W, which is a state variable. When it is increasing in W, a higher value for the agent is preferable, because it entails a lower likelihood of inefficient termination.[30] If this is the case, then the worst-case effort cost is low, i.e. $\theta_t = \underline{\theta}$. Intuitively, if the effort cost is low, the growth in the agent's continuation value (see Equation (7)) is also low and the risk of termination increases, hurting the principal. Conversely, if the principal's value is decreasing in W, the effort cost is high, i.e. $\theta_t = \bar{\theta}$.

To show these results rigorously, we establish that the principal's value function is sufficiently smooth to solve a variant of the HJB equation which accounts for ambiguity,[31] and indeed has the conjectured shape.

[30] This is a standard feature of continuous-time contracts. It appears in DeMarzo and Sannikov (2006), Sannikov (2008), and other related papers.

[31] Precisely, Equation (10) is a HJB-Isaacs equation, which is known to arise in continuous-time zero-sum games. See Fleming and Souganidis (1989) for a seminal reference, and Fleming and Soner (2006), Chapter 11 for a textbook treatment of the deterministic case.

Proposition 1. *The optimal pay-performance sensitivity equals*

$$\psi_t \equiv \bar{\psi} = \frac{\sigma}{\mu}\bar{\theta} \tag{8}$$

for all $0 \le t \le \tau$ *and the worst-case effort cost* $\theta^* = \{\theta_t^*\}_{t\ge 0}$ *satisfies*

$$\theta_t^* = \theta^*(W_t) = \begin{cases} \bar{\theta} & \text{if } W_t \ge W^0 \\ \underline{\theta} & \text{if } W_t < W^0. \end{cases} \tag{9}$$

The principal's value function is strictly concave, twice differentiable, and satisfies

$$rJ(W_t) = \max_{\psi_t \ge \bar{\psi}} \min_{\theta_t \in [\underline{\theta}, \bar{\theta}]} \mu + J'(W_t)(\gamma W_t + \theta_t) + \frac{1}{2}J''(W_t)\psi_t^2 \tag{10}$$

on $[0, \bar{W}]$ *with the boundary conditions* $J(0) = 0$, $J'(\bar{W}) = -1$ *and* $J''(\bar{W}) = 0$. *Further,* $dc_t = 0$ *for all* $W_t < \bar{W}$. W^0 *is the point for which* $J'(W^0) = 0$. *The agent's continuation value follows*

$$dW_t = (\gamma W_t + \theta^*(W_t))\,dt - dc_t + \bar{\psi}dB_t. \tag{11}$$

When $W_t > \bar{W}$, *the principal's value function is* $J(W_t) = J(\bar{W}) - (W_t - \bar{W})$. *The principal pays* $W_t - \bar{W}$ *immediately to the agent and the contract continues with a continuation value of* $\bar{W}$.

The shape of the value function and the worst-case effort cost process are illustrated in Figure 1. We provide a sufficient condition so that implementing effort at all times is optimal in Section B.1 in the Appendix.

The contract has a number of striking implications.

Excessively high pay-performance sensitivity Ambiguity induces a wedge between effort cost and pay-performance sensitivity (PPS). Unlike in DeMarzo and Sannikov (2006), the optimal PPS no longer proportional to the effort cost. This can be seen by inspecting Equation (8) and (9). If the continuation value is below W^0, the worst-case effort cost is $\underline{\theta}$ while the pay-performance sensitivity is $\bar{\psi}$. That is, compared to the effort cost, the principal seems to be using an excessively high PPS. The reason is her ambiguity aversion. If the PPS is below $\bar{\psi}$, the worst case for the principal is an effort cost which induces the agent to shirk. Setting the PPS at $\bar{\psi}$ is the only way to ensure that the agent works for any possible realization of θ_t. This finding adds a new explanation to the long-standing debate about

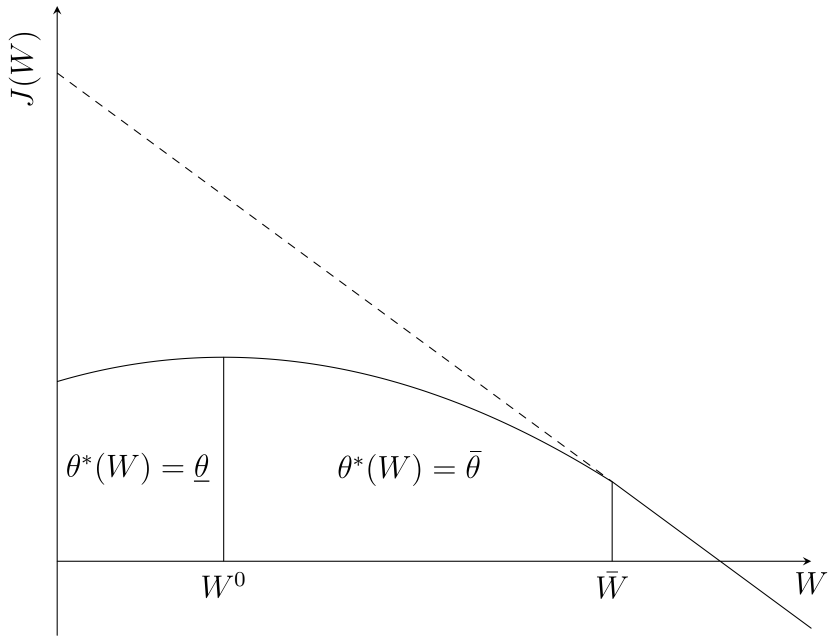

Figure 1: Shape of $J(W)$ and $\theta^*(W)$

whether equity stakes for CEOs are excessive, e.g. Bebchuk and Fried (2006).[32]

Over- and Undercompensation Under ambiguity, agents with good past performance may be overcompensated and those with bad past performance may be undercompensated. To see this, compare the expected payoffs to the agent under the optimal contract to those under the reference probability P, when the effort cost is θ_0. The ambiguity averse principal believes that for high continuation values, the effort cost is high. In that case, she will pay the agent more in expectation. In this sense, the agent is overcompensated relative to the contract without ambiguity. For low continuation values, the principal believes the effort cost to be low and the agent is undercompensated. Thus, ambiguity aversion may not only generate incentives which seem excessively strong, but also lead to higher expected payments to agents. The proposition below formalizes the intuition.

Proposition 2. *For given $\bar{\theta}$ and a $\underline{\theta}$ sufficiently small, there exists a θ_0 such that the agent's expected compensation under the robust contract,*

$$E^{a^*,\theta^*}\left[\int_t^\tau e^{-\gamma(s-t)}dc_s \middle| \mathcal{F}_t^{X,\kappa}\right],$$

[32] As we show below, $\bar{\psi}$ can be understood as the equity share of the CEO in the implementation of the optimal contract.

is larger than his expected compensation without ambiguity for W_t sufficiently large and smaller for W_t sufficiently small.

Career Trajectories The optimal contract can generate a career trajectory in which an agent starts out as being undercompensated[33] and then over time moves up in "seniority" and ends up being overcompensated. This happens because over- and undercompensation arise on the path of the *same* contract. An undercompensated agent's continuation value has low drift, so that on average his expected value form the contract rises slowly. He is also more likely to be fired. If after enough time he moves to the overcompensation region, the drift of his continuation value is faster, that is he accumulates value more quickly, and the likelihood that he is terminated is lower.[34] These features resemble seniority in organizations, see e.g. Farber and Gibbons (1996) and Lazear and Oyer (2012).

Performance-Sensitive Debt With ambiguity, the optimal contract is implemented with performance sensitive debt. Importantly, the equity-and-credit-line implementation of DeMarzo and Sannikov (2006) is no longer optimal.[35] Here is the intuition. With ambiguity, the drift of the agent's continuation value changes depending on the performance history of the firm. Any implementation of the optimal contract must replicate this change in drift. Otherwise, it is either not incentive compatible because it pays the agent too little or not optimal because it pays him too much.

A long-term debt contract with a performance-sensitive interest rates accomplishes this. When the firm is performing poorly (W is low) the debt contract mandates a higher interest rate, which lowers the growth rate of the cash accumulating in the firm. When the firm is performing well, the interest is lower and cash accumulated more quickly.

Proposition 3. *Suppose the firm has long term debt, a credit line, and equity. Let M_t be the draw on the credit line. If the long-term debt has a performance-sensitive interest rate*

$$r_t = \begin{cases} r_1 & \text{if} \quad M_t < M_0 \\ r_2 & \text{if} \quad M_t \geq M_0 \end{cases}$$

for some appropriately chosen values $r_1 > r_2$ and M_0, then the optimal contract is implemented.

[33]This happens whenever the initial continuation value is sufficiently low.

[34]This follows from three facts: (1) the continuation value is higher, (2) the drift of the continuation value is higher, and (3) the volatility is the same as in the undercompensation region. Formally, the probability of hitting $W_{t+\Delta} = 0$ for any $\Delta > 0$ is lower.

[35]See also DeMarzo et al. (2012).

The role of performance sensitive debt is not to incentivize the agent to exert effort, as is the case in Piskorski and Tchistyi (2011), since the managerial equity share already serves this function, or to act as a screening device as in Manso et al. (2010). Instead, the debt is used to adjust the growth of the manager's continuation value. This role for performance sensitive debt is new.

5 Misspecified Contract

We now consider a misspecified contract and show that this contract cannot incentivize the agent to always exert effort. We alter the model as follows. The agent believes that the evolution of κ is given by the reference measure P. That is, he believes that $\theta_t = \theta_0$ for all t and he evaluates any contract under the measure P^{a,θ_0}. The principal is ambiguity averse and evaluates any contract under the worst case $P^{\hat{a},\theta(\alpha)}$ defined in Section 3. This specification corresponds to "Type II Ambiguity" in Hansen and Sargent (2012).

Now, the principal's ambiguity aversion leads to heterogeneous beliefs even when the contract is incentive compatible, i.e. the agent evaluates the contract under P^{a,θ_0} and the principal evaluates it under $P^{a,\theta(\alpha)}$.[36] We distinguish two cases: (1) the principal is sophisticated, i.e. she understands that the agent evaluates the contract under P^{a,θ_0} and that this measure differs from $P^{a,\theta(\alpha)}$; (2) the principal is naive, i.e. she mistakenly believes that the agent evaluates the contract under $P^{a,\theta(\alpha)}$.[37] Since the naive principal uses an incorrect probability measure to evaluate the agent's incentives, we say that the contract is *misspecified*.

With a sophisticated principal, the model collapses to the one in DeMarzo and Sannikov (2006), which we show in Appendix B.2. Intuitively, if the principal knows that the agent uses P^{a,θ_0}, then she knows that the agent's effort cost is θ_0, irrespective of the principal's beliefs. The optimal contract is then the one in DeMarzo and Sannikov (2006). Thus, we focus on the naive principal for the remainder of this section.

The naive principal offers the agent the contract in Proposition 1. The agent understands that the principal (mistakenly) believes that the continuation value follows W_t in Equation (11). Since the agent evaluates the contract under P^{a,θ_0}, his true continuation value differs from W_t. It is given by

[36] This is the crucial difference to Section 3. There, both the principal and the agent evaluate the contract under $P^{a,\theta(\alpha)}$, i.e. there is no disagreement. See Equation (6) for the definition of $\theta(\alpha)$. Recall that $\hat{a} = a$ and thus $P^{\hat{a},\theta(\alpha)} = P^{a,\theta(\alpha)}$ for any incentive compatible contract.

[37] Miao and Rivera (2016) consider a sophisticated principal in a setting with drift ambiguity, see their p. 1416. When studying Type II Ambiguity, Hansen and Sargent (2012) only consider a sophisticated Ramsey planner, i.e. the planner understands that the private sector evaluates payoffs under the reference measure. They do not consider a naive planner as we do.

$$V_t = E^{a,\theta_0}\left[\int_t^\tau e^{-\gamma(s-t)}\left(dc_s - a_s d\kappa_s\right)\middle|\mathcal{F}_t^{X,\kappa}\right],$$

where $a = \{a_t\}_{t\geq 0}$ represents the agent's actual effort.

Since the contract is Markovian in W_t (see Proposition 1), the agent can use the misspecified continuation value W_t to keep track of payments and termination. Generally, the agent may shirk, in which case W_t evolves as

$$dW_t = \left(\gamma W_t + \left(\theta^*\left(W_t\right) - \bar{\theta}\right) + \bar{\theta} a_t\right) dt - dc_t + \bar{\psi} dB_t^a \tag{12}$$

from the agent's perspective.[38]

Thus, the agent's problem is given by

$$V_0 = \sup_a E^{a,\theta_0}\left[\int_0^\tau e^{-\gamma t}\left(dc_t - a_t d\kappa_t\right)\right]$$

subject to Equation (12). Here, the process $\{c_t\}_{t\geq 0}$ and the stopping time τ are obtained from Proposition 1 and the optimization occurs over admissible effort processes. Taking W_t as the state variable, we now characterize the agent's HJB equation and show that the agent may shirk whenever W_t is sufficiently low.

Proposition 4. *The agent's true continuation value V_t is Markovian in the misspecified value W_t, i.e. $V_t = V\left(W_t\right)$, and satisfies almost everywhere the HJB equation*

$$\gamma V\left(W\right) = \max_{a\in\{0,1\}} -\theta_0 a + V'\left(W\right)\left(\gamma W + \left(\theta^*\left(W\right) - \bar{\theta}\right) + \bar{\theta} a\right) + V''\left(W\right)\frac{1}{2}\bar{\psi}^2 \tag{13}$$

with boundary conditions $V\left(0\right) = 0$ and $V'(\bar{W}) = 1$. $V\left(W\right)$ is continuously differentiable for all $W \in [0,\bar{W}]$ and twice continuously differentiable on $[0,W^0)$ and $(W^0,\bar{W}]$.

The agent exerts effort whenever

$$V'\left(W\right) \geq \frac{\theta_0}{\bar{\theta}}.$$

There is a $\hat{\theta}_0$ such that for all $\theta_0 \leq \hat{\theta}_0$, the agent always exerts effort. For $\theta_0 > \hat{\theta}_0$, the agent shirks when $W \in [0, W_s\left(\theta_0\right))$ for some $W_s\left(\theta_0\right) \leq W^0$ and works for all $W \in \left[W_s\left(\theta_0\right), \bar{W}\right]$. Further, there exists a $\tilde{\theta}_0 > \hat{\theta}_0$ so that for $\theta_0 > \tilde{\theta}_0$, $V\left(W\right)$ is concave, while for any $\theta_0 \leq \tilde{\theta}_0$, $V\left(W\right)$ is convex on $[0, W^0)$, and concave on $\left[W^0, \bar{W}\right]$.

The strength of incentives in the misspecified contract can be summarized by $V'\left(W_t\right)$.

[38]This follows by applying Girsanov's Theorem to Equation (11).

Intuitively, shirking leads to a lower drift of W_t (see Equation (12)) and $V'(W_t)$ is the value the agent places on the drift in W_t.

The proposition contains important results about the strength of the agent's incentives and his attitude towards risk. We explain those below for the case when $\hat{\theta}_0 < \theta_0 \leq \tilde{\theta}_0$, which is the most interesting one.

Shirking at the Bottom In the case without ambiguity, the contract's incentives are always sufficient to motivate the agent to work. Because the misspecified contract features an excessively high pay performance sensitivity, it is tempting to conclude that the incentives under this contract are stronger and that the agent always works.[39] This is wrong. Instead, the agent shirks when W_t is low. Intuitively, the worst-case effort cost is low, i.e. $\theta_t = \underline{\theta}$, so the drift of W_t is low as well (see Equation (7)). This increases the likelihood that W_t hits the termination boundary at zero rather than the payout boundary $\bar{W}$. In other words, the agent is more likely to be fired rather than being paid.[40] As the proposition shows, this effect can destroy the agent's incentives despite the higher PPS.

This result is in stark contrast to a large literature on dynamic contracts, which shows that the treat of termination helps provide incentives (e.g. Spear and Wang (2005), DeMarzo and Sannikov (2006), Biais et al. (2010), Wang (2011), Garrett and Pavan (2012), and Fong and Li (2017)). In e.g. DeMarzo and Sannikov (2006), the principal knows the agent's effort cost and under the optimal contract, the PPS is sufficient to incentivize effort at all continuation values. As Proposition 4 shows, this is no longer true if the contract is misspecified, even though the PPS is excessively high.

Importantly, our result on shirking at the bottom is consistent with empirical evidence. Flabbi and Ichino (2001) find that workers with less seniority, who are more likely to be fired, are also more likely to have high absenteeism or to engage in misconduct, which can be interpreted as shirking.[41] In the context of our model, a low seniority corresponds to a low continuation value, since the agent is more likely to be fired.

Hump-Shaped Incentive Strength The agent's marginal value of exerting effort is

$$V'(W_t)\,\bar{\theta} - \theta_0,$$

[39]Precisely, the pay performance sensitivity without ambiguity is $\psi_t = \frac{\sigma}{\mu}\theta_0$, while here it is $\frac{\sigma}{\mu}\bar{\theta}$, which is strictly higher.

[40]Recall that the agent is paid only when $W_t = \bar{W}$, so if the drift of W_t is low, the likelihood of hitting zero before hitting $\bar{W}$ is higher.

[41]See their Table 7, p. 373.

which can be seen from Equation (13). Thus, the strength of the agent's incentives is proportional to $V'(W_t)$. This value first increases for continuation values below W^0 and then decreases for values above W^0.

Here is the intuition. As we argued above, incentives may fail for low continuation values, because the low drift implies a high likelihood of termination. As W moves away from zero, the likelihood of termination decreases and the agent is more likely to reach the payment boundary $\bar{W}$. Because of this, the strength of incentives increases. As W increases beyond W^0, however, the agent is likely to reach the payment boundary and unlikely to be terminated with or without exerting effort. Thus, the strength of incentives declines. Overall, the agent's incentives are the strongest for moderate continuation values.

These results predict that agents with low seniority will shirk, but agents with high seniority will have low incentives to exert effort. This provides an alternative explanation for managerial entrenchment (e.g. Berger et al. (1997)).

Risk-Loving Juniors, Risk-Averse Seniors When the agent shirks, the drift of W_t is negative so in expectation W_t converges to the firing boundary.[42] By staying in the contract and shirking, the agent is essentially betting on luck and his value function is convex.[43] Thus, the agent is risk-loving, in the sense that he would prefer a higher volatility of W_t. For $W > W^0$ however, a higher volatility increases the chance of reaching the undercompensation region, which is bad for the agent. Because of this, his value is concave and the agent is risk-averse. Interpreting W as seniority, this implies that more junior agents are risk-loving while senior ones are risk-averse. This result is consistent with the literature on "the quiet life" (see e.g. Bertrand and Mullainathan (2003)), which documents that entrenched managers become risk-averse.

Figure 2 displays the agent's value function for high, medium and low effort cost values. In the upper panels, the dashed line is the identity, at which $V(W) = W$. Thus, depending on the effort cost as well as the current W, the agent's value may be higher or lower than one the principal believes him to have. The lower panels display the slope of the agent's value. He exerts effort whenever $V'(W)$ is above the solid black line, which corresponds to $\frac{\theta_0}{\theta}$.

[42] See Equation (12).

[43] This is precisely the reason why the agent does not immediately leave the contract, given that he is not getting paid currently, and knows that he eventually will be fired if he continues shirking.

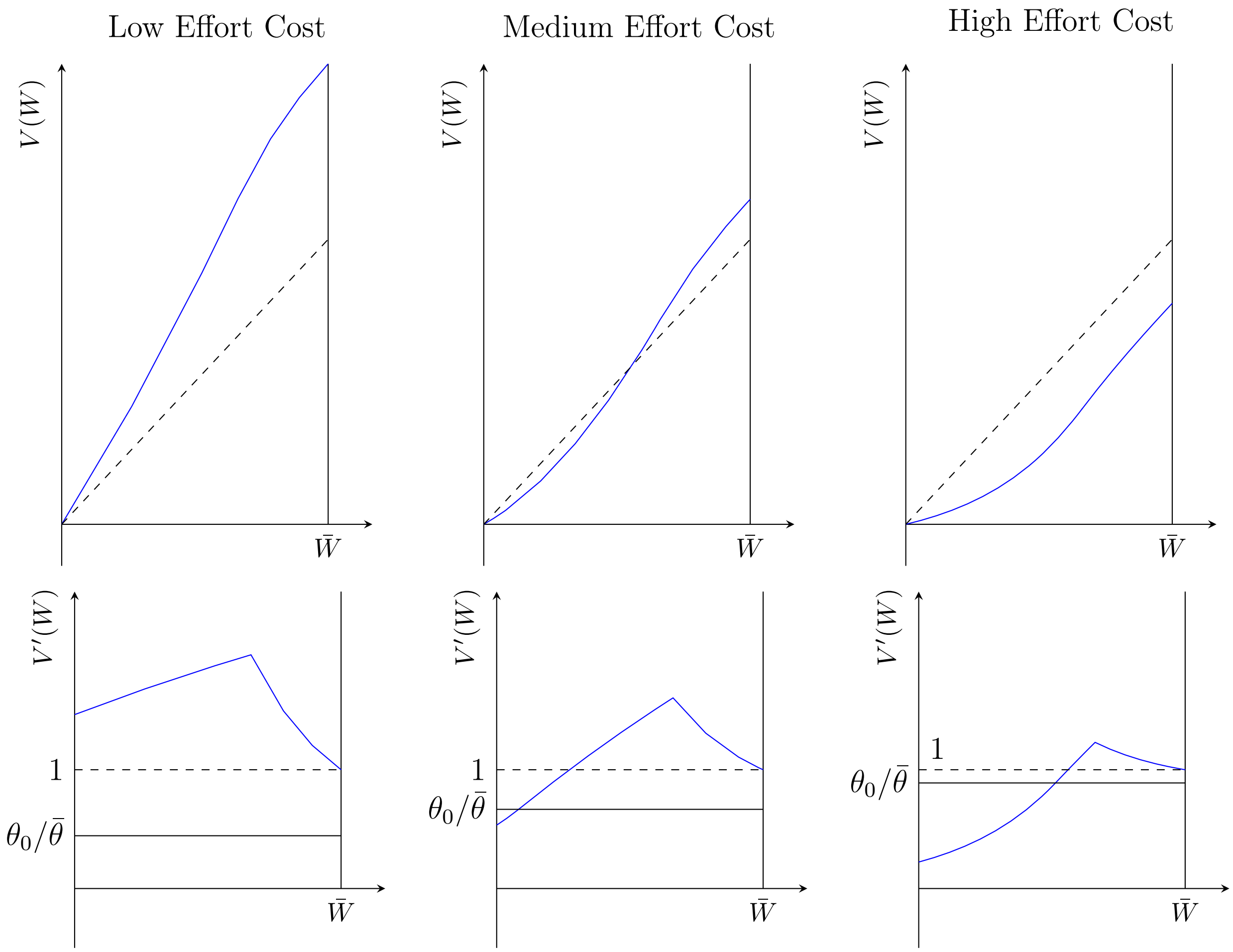

Figure 2: Agent's Value Function and Incentives

6 Extensions

6.1 Drift Ambiguity

When ambiguity aversion is about the productivity of the project, i.e. the drift, there is no interaction between the worst case and incentive provision. This is because under the optimal contract, the agent's incentives are the same independently of the project's productivity. As a result, the worst case is static and given by a low productivity at all points in time. The features we have found with effort cost ambiguity, i.e. excessive pay-performance sensitivity, over- and undercompensation, do not arise, and performance-sensitive debt does not implement the optimal contract.

There is now a single Brownian motion B on the probability space $(\Omega, \mathcal{F}, P)$. Ambiguity is modeled again via Girsanov's theorem and a set of density generators

$$\Theta = \left\{\theta : \underline{\theta} \leq \theta_t \leq \bar{\theta}\right\},$$

so that $-\frac{\mu}{\sigma} < \underline{\theta} < 0 < \bar{\theta}$. The effort cost is now a constant κ with $0 < \kappa < \underline{\theta} + \sigma\mu$, instead of being a stochastic process as in Equation (3).[44] Under measure P, the output follows

$$dX_t = \mu dt + \sigma dB_t,$$

while under measure $P^{a,\theta}$,

$$B_t^{a,\theta} = B_t + \int_0^t \frac{1}{\sigma}\left(\mu\left(1 - a_s\right) - \theta_s\right) ds \tag{14}$$

is a Brownian motion and

$$dX_t = \left(\mu a_t + \theta_t\right) dt + \sigma dB_t^{a,\theta}.$$

Different θ now represent different drifts for the output. The reference measure P corresponds to the case when $\theta_t = 0$ and $a_t = 1$ for all t. That is, we have $\theta_0 = 0$. For any incentive compatible contract, the agent's continuation value is given by

$$W_t = E^{a,\theta}\left[\int_t^\tau e^{-\gamma(s-t)}\left(dc_s - a_s\kappa ds\right)|\mathcal{F}_t^X\right],$$

which is the analog of Equation (5). Under $P^{a,\theta}$, the continuation value admits the repre-

[44] For the same reason as in Section 3, this assumption is needed to prevent the contract form becoming trivial. If $\underline{\theta} + \sigma\mu < \kappa$, then under the worst case implementing effort is not optimal even in the first best.

sentation

$$dW_t = (\gamma W_t + \kappa a_t)\, dt - dc_t + \psi_t dB_t^{a,\theta}$$

and the agent exerts effort whenever

$$\psi_t \geq \psi \equiv \frac{\sigma}{\mu}\kappa.$$

The principal's problem is to solve

$$J_0 = \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \Theta} E^{a,\theta}\left[\int_0^\tau e^{-rt}\,(dX_t - dc_t)\right]$$

subject to incentive compatibility and participation for the agent. Below, we establish the analog of Proposition 1 and characterize the optimal contract.

Proposition 5. *Under drift ambiguity, the principal's value function is twice continuously differentiable, and satisfies*

$$rJ\,(W) = \max_{\psi' \geq \psi} \min_{\theta \in \left[\underline{\theta}, \bar{\theta}\right]} \mu + \theta + J'\,(W)\,(\gamma W + \kappa) + J''\,(W)\,\frac{1}{2}\psi'^2$$

on $\left[0, \bar{W}\right]$ *subject to the boundary conditions* $J\,(0) = 0$, $J'(\bar{W}) = -1$ *and* $J''(\bar{W}) = 0$. *For all* t, *the worst case is* $\theta_t = \underline{\theta}$ *and the pay performance sensitivity* ψ_t *equals*

$$\psi_t = \psi \equiv \frac{\sigma}{\mu}\kappa. \tag{15}$$

A few features are noteworthy.

Static Worst Case The worst case in the dynamic contract is the same as in the static one. In a one-shot contracting model with ambiguity aversion about the average productivity, the worst case is that the productivity is at its lowest possible level. This result extends to a dynamic model, because ambiguity about the drift does not interact with the agent's incentives. Ambiguity about effort cost generates a worst case that changes over time, precisely because of this interaction.

No Excessive Incentives Because ambiguity aversion does not interact with incentive provision, the PPS is the same as in a contract without ambiguity. Similarly, drift ambiguity does not generate regions of over- and undercompensation.

Equity and Credit Lines Since the drift of W_t is constant, the standard equity-and-credit-line implementation of DeMarzo and Sannikov (2006) remains optimal.

It is worthwhile to compare the result of Proposition 5 to the ones obtained by Miao and Rivera (2016) and Dumav (2017). In Miao and Riviera's paper, the PPS can be above the level required to incentivize effort (here $\frac{\sigma}{\mu}\kappa$). This is because their paper features an entropy cost.[45] A lower drift increases the principal's payoff via the entropy cost, so the worst case depends continuously on the choice of PPS. Setting the PPS higher than necessary may be beneficial for the principal because of this interaction. As our Proposition shows, Miao and Riviera's result disappears if there is no entropy cost, i.e. when we are in a standard maxmin framework. In Dumav's paper, effort is continuous as in Sannikov (2008). Ambiguity aversion then lowers the optimal effort and therefore the optimal PPS compared to the case without ambiguity. His result does not rely on an entropy cost.

6.2 Misspecified Contract under Drift Ambiguity

Just as in the baseline model, we can study the agent's value function when the contract is misspecified. Because the interaction with incentive provision is missing, drift ambiguity fails to generate the effects found in Section 5.

As in the setup in Section 5, the agent evaluates the contract under measure P^{a,θ_0} and knows that the output process follows[46]

$$dX_t = \mu a_t dt + \sigma dB_t^{a,\theta_0}.$$

His true continuation value is given by

$$V_t = E^{a,\theta_0}\left[\int_t^{\tau} e^{-\gamma(s-t)}\left(dc_s - a_s\kappa ds\right)|\mathcal{F}_t^X\right],$$

which is the analog of Equation (12). The principal is naive, i.e. she believes that the agent evaluates the contract under the same measure as her, and she offers the contract of Proposition (5). Thus, the agent knows that the principal evaluates the contract under $P^{a^*,\underline{\theta}}$,

[45]See their Eq. (25) on p. 1417.

[46]Consistent with the notation in Section 6.1, $P^{a,0}$ is the measure under which

$$B_t^{a,0} = B_t + \int_0^t \frac{\mu}{\sigma}\left(1 - a_s\right) ds$$

is a Brownian motion, i.e., when $\theta_t = 0$ for all t. See Equation (14).

i.e. that she believes that the drift is $\mu + \underline{\theta}$.[47] As in Section 5, the dynamics of the contract are determined by W_t, which is now the misspecified continuation value from the principal's viewpoint and which follows

$$\begin{aligned} dW_t &= (\gamma W_t + \kappa)\, dt - dc_t + \psi dB_t^{a^*,\underline{\theta}} \\ &= \left(\gamma W_t + \kappa a_t - \frac{\psi}{\sigma}\underline{\theta}\right) dt - dc_t + \psi dB_t^{a,\theta_0}, \end{aligned}$$

using Girsanov's theorem. Here, ψ is defined in Equation (15). As in Section 5, the true continuation value is Markovian in W_t, i.e. $V_t = V(W_t)$. Characterizing the agent's HJB equation shows that it is always optimal to exert effort.

Proposition 6. *The agent's value function $V(W_t)$ has the unique solution $V(W_t) = W_t - \frac{\kappa}{\gamma\mu}\underline{\theta}$ on $\left[0, \bar{W}\right]$. The misspecified contract is incentive compatible, i.e., $a_t = 1$ for all $t < \tau$, and the agent is risk-neutral in W.*

There is no shirking for low W, the strength of incentives, $V'(W)$, is constant, and the agent is indifferent to the volatility of W. Ambiguity about the drift thus fails to replicate the results of Section 5.[48]

6.3 Ambiguity Averse Agent

We now consider the case when the principal and the agent are both ambiguity averse. Then, the contract either collapses to the one in DeMarzo and Sannikov (2006) (if the principal is sophisticated) or has the same qualitative features as the misspecified contract in Section 5 (if the principal is naive).

The setting is the same as in Section 3, except that the agent is also ambiguity averse. For any given contract α, the agent's value is given by

$$W_0(\alpha) = \inf_{\theta^A \in \Theta} E^{a,\theta^A}\left[\int_0^\tau e^{-\gamma t}\left(dc_t - a_t d\kappa_t\right)\right], \tag{16}$$

That is, the agent evaluates the contract under P^{a,θ^A} and his continuation value is given by

$$W_t = E^{a,\theta^A}\left[\int_t^\tau e^{-\gamma(s-t)}\left(dc_s - a_s d\kappa_s\right) \Big| \mathcal{F}_t^{X,\kappa}\right].$$

[47] In Proposition 5, the worst-case is $\theta_t^* = \underline{\theta}$ for $t < \tau$ and the optimal effort is $a_t^* = 1$ for all $t < \tau$. Thus, $P^{a^*,\underline{\theta}}$ denotes the probability measure under a^* and θ^*, from the principal's perspective.

[48] Since $\underline{\theta} < 0$, we have $V(W_t) > W_t$ for all $t < \tau$.

The contract is incentive compatible whenever

$$W_0(\alpha) \geq \inf_{\theta^A \in \Theta} E^{\tilde{a},\theta^A} \left[\int_0^\tau e^{-\gamma t} \left(dc_t - \tilde{a}_t d\kappa_t \right) \right] \tag{17}$$

for any admissible effort process $\tilde{a}$.

The principal's problem is now given by

$$J_0 = \sup_{\alpha \in \mathcal{A}} \inf_{\theta^P \in \Theta} E^{a,\theta^P} \left[\int_0^\tau e^{-rt} \left(\mu a_t dt - dc_t \right) \right], \tag{18}$$

subject to the incentive compatibility condition (17), and the participation constraint $W_0(\alpha) \geq 0$. As in Section 3, we focus on contracts which implement effort at all times. Generally, the agent's worst-case θ^A and the principal's worst-case θ^P may differ.[49] Thus, as in Section 5, ambiguity potentially leads to disagreement, and we have to distinguish between a naive and a sophisticated principal.

Proposition 7. *If the principal is sophisticated, the optimal contract is the same one as in DeMarzo and Sannikov (2006), with constant effort cost $\bar{\theta}$. If the principal is naive, the agent may shirk. The agent's effort is given by Proposition 4, with θ_0 replaced by $\bar{\theta}$.*

The intuition for this result is simple. For any contract which requires effort, the agent's worst case is simply that the effort cost is high, i.e. $\theta_t^A = \bar{\theta}$ for all $t < \tau$. If the principal is sophisticated, she understands that the agent effectively has a constant effort cost of $\bar{\theta}$ and her own ambiguity aversion does not affect her optimal contract. She then optimally offers the contract of DeMarzo and Sannikov (2006). If the principal is naive, she offers the contract of Proposition 1, but the agent, essentially, has a constant effort cost $\bar{\theta}$. The characterization in Section 5 applies and the agent shirks whenever W_t is sufficiently low.

6.4 Ambiguity about Constant Effort Costs

In this section, we restrict the set of density generators so that the effort cost θ_t is constant across time. In this setting, the worst-case effort cost is either always high or always low and the features highlighted in Section 4 do not arise. Thus, having a sufficiently rich set of density generators which allows for θ_t to be time-varying is crucial for our results.

Consider the setup of Section 3 and suppose that the set of admissible density generators is given by

$$\Theta = \left\{ \{\theta_t\}_{t \geq 0} : \theta_t = \hat{\theta} \text{ for all } t \geq 0, \hat{\theta} \in \{\underline{\theta}, \bar{\theta}\} \right\}.$$

[49] Hence, even if the contract is incentive compatible, the measures P^{a,θ^P} and P^{a,θ^A} may differ.

Intuitively, after the principal chooses the contract, nature chooses the effort cost, but is restricted to choosing one which is constant in time. Alternatively, the principal is ambiguity averse about the effort cost, but only considers constant effort costs to be possible. For simplicity, we restrict attention to contracts which implement a constant pay-performance sensitivity.

Proposition 8. *The principal's value function satisfies*

$$rJ(W) = \sup_{\psi \geq \bar{\psi}} \mu + J'(W)(\gamma W + \theta^*) + J''(W) \frac{1}{2}\bar{\psi}^2 \tag{19}$$

on $\left[0, \bar{W}\right]$ *with boundary conditions* $J(0) = L$, $J'(\bar{W}) = -1$ *and* $J''(\bar{W}) = 0$, *where*

$$\bar{\psi} = \frac{\sigma}{\mu}\bar{\theta}$$

and where θ^* *is the worst-case effort cost. The optimal PPS is given by* $\psi_t = \bar{\psi}$.

Since in general, $J(W)$ is non-monotone in θ,[50] it is difficult to characterize the worst-case θ^* analytically. However, we can numerically solve the principal's HJB equation and determine the worst-case for given parameter values. The results are given in Figure 3. For sufficiently low initial continuation value W_0, the worst-case is the constant process $\underline{\theta}$, while for sufficiently high values, the worst-case is $\bar{\theta}$.

7 Conclusion

This paper aims towards understanding contracts when firms have limited information, which we interpret as ambiguity aversion. In a dynamic setting, this is natural. Firms may not even know the distribution of shocks to the agent's willingness to work or his productivity. Since the cost of incentivizing the agent changes over time, so does the worst case. These dynamics generate important predictions for the agent's compensation and provide a new role for performance-sensitive debt. To insure against ambiguity, the principal must provide incentives which seem excessively strong, both compared to the agent's realized effort cost and the contract without ambiguity. Our model thus can generate overcompensation, i.e.

[50]We have

$$\frac{\partial}{\partial\theta} J(W_t) = E\left[\int_t^\tau e^{-r(s-t)} J'(W_s) \,|\mathcal{F}_t^X\right],$$

which follows from differentiating the principal's HJB equation with respect to θ and using the Feynman-Kac formula (see DeMarzo and Sannikov (2006), p. 2699). Since $J'(W)$ can be both positive and negative, $\frac{\partial}{\partial\theta} J(W_t)$ can also be positive or negative, depending on parameters.

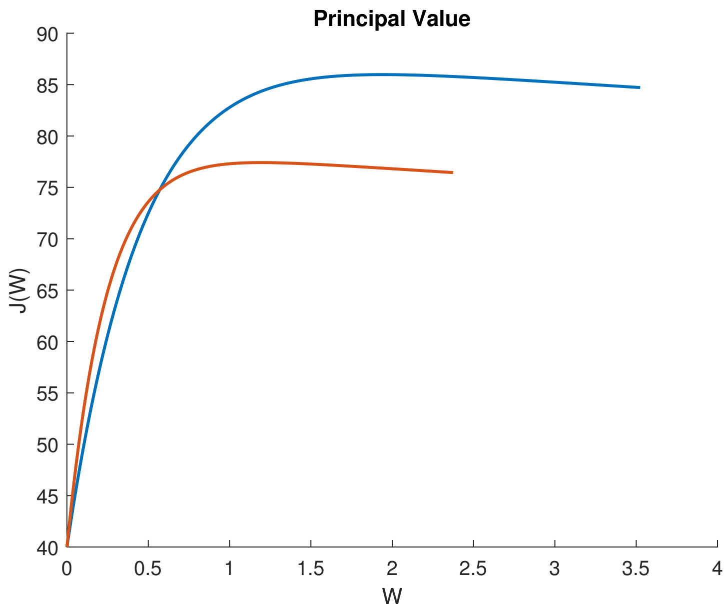

Figure 3: Principal's value function for different effort costs in Equation (19). Parameters: $\gamma = 0.15$, $r = 0.1$, $\sigma = 5$, $\mu = 10$, $L = 40$, $\underline{\theta} = 1$ and $\bar{\theta} = 2$. The blue line corresponds with effort cost $\underline{\theta}$ and the red line corresponds with $\bar{\theta}$.

higher expected payments, for successful managers together with a seemingly generous equity package, which induces a high PPS. These features have long puzzled researchers interested in CEO compensation. To our knowledge, our explanation based on robustness is new.

The second contribution of our paper is to study the agent's incentives under a misspecified contract. Firing, which in many contracts acts as an incentive device, now diminishes the agent's incentives, because the misspecified contract does not accurately compensate the agent for the likelihood of being laid off. As a result, this contract generates shirking at the bottom and an incentive to risk-shift. That is, if the agent were allowed to pick the project, he would prefer a riskier one. By staying in the contract, he is gambling on a sequence of good performance, which brings him into a region where exerting effort is valuable. Once that region is reached, the agent's preferences reverse. He seeks to minimize the risk of the project for fear of being terminated. These features, which resemble preferences for risk-shifting and the quiet life, do not occur in the canonical model of DeMarzo and Sannikov (2006). They are solely the result of introducing ambiguity.

A Proofs

A.1 Proof of Lemma 1

This result is a standard consequence of the Girsanov Theorem (see Jeanblanc et al. (2009), Sec. 1.7.4, p. 72). It is proven here for the reader's convenience.

Suppose that (B_t, Z_t) is a two-dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$. By the Lévy characterization of Brownian motion (Øksendal (2003), Theorem 8.6.1), (B_t^a, Z_t^θ) is a Brownian motion with respect to $P^{a,\theta}$ if and only if

(i) (B_t^a, Z_t^θ) is a martingale w.r.t. $P^{a,\theta}$; and

(ii) $(B_t^a)^2 - t, (Z_t^\theta)^2 - t$, and $B_t^a Z_t^\theta$ are martingales w.r.t. $P^{a,\theta}$.

Given the definition of $P^{a,\theta}$ (i.e., $dP^\theta/dP^{a,\theta}|_{\mathcal{F}_t} \equiv z_0^t(a,\theta)$ for all t), a process A_t is a martingale w.r.t. $P^{a,\theta}$ if and only if $A_t z_0^t(a,\theta)$ is a martingale w.r.t. P. Thus, $\big(B_t^a, Z_t^\theta\big)$ is a martingale under $P^{a,\theta}$ if and only if $(B_t^a z_0^t(a,\theta), Z_t^\theta z_0^t(a,\theta))$ is a martingale under P. Respectively, $(B_t^a)^2 - t, (Z_t^\theta)^2 - t$, and $B_t^a Z_t^\theta$ are martingales under $P^{a,\theta}$ if and only if $((B_t^a)^2 - t)\, z_0^t\,(a,\theta)$, $\big((Z_t^\theta)^2 - t\big)\, z_0^t\,(a,\theta)$, and $B_t^a Z_t^\theta z_0^t\,(a,\theta)$ are martingales under P.

In what follows, we prove the equivalent to conditions (i) and (ii) under measure P. We first characterize the SDE for the density process

$$z_0^t(a,\theta) = \exp\Big(-\int_0^t (\theta_0 - \theta_s) dZ_s - \frac{1}{2}\int_0^t (\theta_0-\theta_s)^2 ds\Big) \exp\Big(-\int_0^t \frac{\mu}{\sigma}(1-a_s) dB_s - \frac{1}{2}\int_0^t \big(\frac{\mu}{\sigma}(1-a_s)\big)^2 ds\Big).$$

Since the exponential map $x \mapsto e^x$ is a $\mathcal{C}^2$ function, we apply the Ito formula to $z_0^t(a,\theta) = exp(M_t)$, for the process M_t which satisfies

$$dM_t = -(\theta_0 - \theta_t) dZ_t - \frac{1}{2}(\theta_0 - \theta_t)^2 dt - \frac{\mu}{\sigma}(1-a_t) dB_t - \frac{1}{2}\big(\frac{\mu}{\sigma}(1-a_t)\big)^2 dt,$$

to get the differential equation for $z_0^t(a,\theta)$ under P:

$$dz_0^t(a,\theta) = z_0^t(a,\theta) dM_t + \frac{1}{2} z_0^t(a,\theta) dM_t \cdot dM_t = -z_0^t(a,\theta)((\theta_0 - \theta_t) dZ_t + \frac{\mu}{\sigma}(1-a_t) dB_t). \quad (20)$$

Equation (20) provides the differential equations of $(B_t^a z_0^t(a,\theta), Z_t^\theta z_0^t(a,\theta))$ under P, by the

Ito formula:

$$
\begin{aligned}
d(B_t^a z_0^t(a,\theta)) &= B_t^a dz_0^t(a,\theta) + z_0^t(a,\theta)dB_t^a + dB_t^a \cdot dz_0^t(a,\theta) \\
&= -B_t^a z_0^t(a,\theta)(\theta_0-\theta_t)dZ_t + z_0^t(a,\theta)\big(1-\frac{\mu}{\sigma}(1-a_t)B_t^a\big)dB_t; \\
d(Z_t^\theta z_0^t(a,\theta)) &= Z_t^\theta dz_0^t(a,\theta) + z_0^t(a,\theta)dZ_t^\theta + dZ_t^\theta \cdot dz_0^t(a,\theta) \\
&= z_0^t(a,\theta)\big(1-(\theta_0-\theta_t)Z_t^\theta\big)dZ_t - Z_t^\theta z_0^t(a,\theta)\frac{\mu}{\sigma}(1-a_t)dB_t.
\end{aligned}
$$

These two equations show that $B_t^a z_0^t(a,\theta)$ and $Z_t^\theta z_0^t(a,\theta)$ are Ito integrals:

$$
B_t^a z_0^t(a,\theta) = B_0^a - \int_0^t B_s^a z_s^0(a,\theta)(\theta_0-\theta_s)dZ_s + \int_0^t z_s^0(a,\theta)\big(1-\frac{\mu}{\sigma}(1-a_s)B_s^a\big)dB_s, \tag{21}
$$

$$
Z_t^\theta z_0^t(a,\theta) = Z_0^\theta + \int_0^t z_s^0(a,\theta)\big(1-(\theta_0-\theta_s)Z_s^\theta\big)dZ_s - \int_0^t Z_s^\theta z_s^0(a,\theta)\frac{\mu}{\sigma}(1-a_s)dB_s. \tag{22}
$$

Without loss of generality, we normalize $B_0^a = 0$ and $Z_0^\theta = 0$. Next, by using Equations (21)-(22), we prove conditions (i) and (ii).

(Part (i)) The pair (B_t^a, Z_t^θ) is integrable under $P^{a,\theta}$, since $(B_t^a z_0^t(a,\theta), Z_t^\theta z_0^t(a,\theta))$ is integrable under P, because $a = \{a_t\}_{t\geq 0}$ and $\theta = \{\theta_t\}_{t\geq 0}$ are bounded. Then, from Equations (21)-(22) we see that $(B_t^a z_0^t(a,\theta), Z_t^\theta z_0^t(a,\theta))$ are Ito integrals under P and hence P-martingales.[51] Thus, $E^{a,\theta}[(B_T^a, Z_T^\theta)|\mathcal{F}_t^{B,Z}] = (B_t^a, Z_t^\theta)$ for all $T \geq t$.[52]

(Part(ii)) A similar argument as in part (i) shows that $(B_t^a)^2 - t, (Z_t^\theta)^2 - t$, and $B_t^a Z_t^\theta$ are martingales w.r.t. $P^{a,\theta}$.[53] Specifically, we characterize the differential equations for $((B_t^a)^2 - t)z_0^t(a,\theta), ((Z_t^\theta)^2 - t)z_0^t(a,\theta)$, and $B_t^a Z_t^\theta z_0^t(a,\theta)$, in terms of (B_t, Z_t) under P:

$$
\begin{aligned}
d(((B_t^a)^2 - t)z_0^t(a,\theta)) &= -z_0^t(a,\theta)dt + ((B_t^a)^2 - t)dz_0^t(a,\theta) + 2B_t^a z_0^t(a,\theta)dB_t^a \\
&\quad + z_0^t(a,\theta)dB_t^a \cdot dB_t^a + 2B_t^a dB_t^a \cdot dz_0^t(a,\theta) \\
&= -z_0^t(a,\theta)((B_t^a)^2 - t)(\theta_0-\theta_t)dZ_t \\
&\quad + z_0^t(a,\theta)\left(2B_t^a - ((B_t^a)^2 - t)\frac{\mu}{\sigma}(1-a_t)\right)dB_t.
\end{aligned}
$$

The first equality holds by the Ito formula and the second equality holds by Equation (20). The differential equation for $d(((B_t^a)^2 - t)z_0^t(a,\theta))$ shows that the process $((B_t^a)^2 - t)z_0^t(a,\theta)$

[51] All Ito integrals are martingales, see e.g. Karatzas and Shreve (1991), Prop. 2.10, p. 139.

[52] In particular, note that the Girsanov Theorem implies that $\big(B^a, Z^\theta\big) \in \mathcal{F}^{B,Z}$ and $\mathcal{F}^{B^a,Z^\theta} \subset \mathcal{F}^{B,Z}$ (see e.g. Jeanblanc et al. (2009), p. 77). It is sufficient to prove that $\big(B^a, Z^\theta\big)$ is a martingale under $\mathcal{F}^{B,Z}$. Then, by the tower property of conditional expectations, $\big(B^a, Z^\theta\big)$ is also a martingale under its own filtration, $\mathcal{F}^{B^a,Z^\theta}$, i.e. $E^{a,\theta}[(B_T^a, Z_T^\theta)|\mathcal{F}_t^{B^a,Z^\theta}] = (B_t^a, Z_t^\theta)$.

[53] Equivalently, $((B_t^a)^2 - t)z_0^t(a,\theta)$, $((Z_t^\theta)^2 - t)z_0^t(a,\theta)$, and $B_t^a Z_t^\theta z_0^t(a,\theta)$ are martingales w.r.t. P.

is characterized as an Ito integral given the Brownian motion (B_t, Z_t) under P. Given the integrability of $(B_t^a)^2 - t$ and the boundedness of a_t and θ_t, we have

$$\begin{aligned}((B_t^a)^2 - t)z_0^t(a,\theta) &= -\int_0^t z_s^0(a,\theta)((B_s^a)^2 - s)(\theta_0 - \theta_s)dZ_s \\ &\quad + \int_0^t z_s^0(a,\theta)\left(2B_s^a - ((B_s^a)^2 - s)\frac{\mu}{\sigma}(1-a_s)\right)dB_s.\end{aligned}$$

Thus, $((B_t^a)^2 - t)z_0^t(a,\theta)$ is an Ito integral and therefore is a P-martingale:

$$E\left[((B_T^a)^2 - T)z_0^t(a,\theta)\middle|\mathcal{F}_t^{B,Z}\right] = ((B_t^a)^2 - t)z_0^t(a,\theta) \quad \forall T \geq t.$$

Equivalently, $(B_t^a)^2 - t$ is a martingale under $P^{a,\theta}$, using $dP^{a,\theta} = z_0^t(a,\theta)dP$:

$$E^{a,\theta}\left[(B_T^a)^2 - T\middle|\mathcal{F}_t^{B,Z}\right] = E\left[((B_T^a)^2 - T)z_0^t(a,\theta)\middle|\mathcal{F}_t^{B,Z}\right] = (B_t^a)^2 - t \quad \forall T \geq t.$$

Note that because $\mathcal{F}_t^{B^a,Z^\theta} \subseteq \mathcal{F}_t^{B,Z}$,[54] we also have $E^{a,\theta}[(B_T^a)^2 - T|\mathcal{F}_t^{B^a,Z^\theta}] = (B_t^a)^2 - t$ for all t. Similarly, we can show that $((Z_t^\theta)^2 - t)z_0^t(a,\theta)$ and $B_t^a Z_t^\theta z_0^t(a,\theta)$ are martingales under P, and hence, $((Z_t^\theta)^2 - t)$ and $B_t^a Z_t^\theta$ are martingales under $P^{a,\theta}$:

$$\begin{aligned}d(((Z_t^\theta)^2 - t)z_0^t(a,\theta)) &= -z_0^t(a,\theta)dt + ((Z_t^\theta)^2 - t)dz_0^t(a,\theta) + 2Z_t^\theta z_0^t(a,\theta)dZ_t^\theta \\ &\quad + z_0^t(a,\theta)dZ_t^\theta \cdot dZ_t^\theta + 2Z_t^\theta dZ_t^\theta \cdot dz_0^t(a,\theta) \\ &= -z_0^t(a,\theta)((Z_t^\theta)^2 - t)\frac{\mu}{\sigma}(1-a_t)dB_t \\ &\quad + z_0^t(a,\theta)\left(2Z_t^\theta - ((Z_t^\theta)^2 - t)(\theta_0 - \theta_t)\right)dZ_t\end{aligned}$$

and

$$\begin{aligned}d(B_t^a Z_t^\theta z_0^t(a,\theta)) &= B_t^a Z_t^\theta dz_0^t(a,\theta) + Z_t^\theta z_0^t(a,\theta)dB_t^a + B_t^a z_0^t(a,\theta)dZ_t^\theta + B_t^a dZ_t^\theta \cdot dz_0^t(a,\theta) \\ &\quad + Z_t^\theta dB_t^a \cdot dz_0^t(a,\theta) + z_0^t(a,\theta)dB_t^a \cdot dZ_t^\theta \\ &= z_0^t(a,\theta)B_t^a\left(1 - Z_t^\theta(\theta_0 - \theta_t)\right)dZ_t + z_0^t(a,\theta)Z_t^\theta\left(1 - B_t^a\frac{\mu}{\sigma}(1-a_t)\right)dB_t.\end{aligned}$$

From parts (i) and (ii), we conclude that (B_t^a, Z_t^θ) is a two-dimensional Brownian motion under $P^{a,\theta}$. Furthermore, B_t^a and Z_t^θ are independent:

$$E^{a,\theta}\left[B_T^a Z_T^\theta|\mathcal{F}_t^{B,Z}\right] = B_t^a Z_t^\theta = E^{a,\theta}\left[B_T^a\middle|\mathcal{F}_t^{B,Z}\right]E^{a,\theta}\left[Z_T^\theta\middle|\mathcal{F}_t^{B,Z}\right] \quad \forall T \geq t,$$

[54] See again Jeanblanc et al. (2009), p. 77.

when the first equality holds because $B_t^a Z_t^\theta$ is a martingale under $P^{a,\theta}$ (condition (ii)) and the second equality holds because (B_t^a, Z_t^θ) is a martingale under $P^{a,\theta}$ (condition (i)).

A.2 Proof of Lemma 2

The agent optimizes over admissible effort $\tilde{a} \in \mathcal{F}^{X,\kappa}$ taking the contract α and the process θ as given. We prove the result using the martingale characterization of non-Markovian control problems in Elliott (1977), together with a guess and verify approach which shows that $\tilde{a} \in \mathcal{F}^X$ without loss of generality.

The agent maximizes his continuation value, at each t, under $P^{\tilde{a},\theta}$:

$$\sup_{\widetilde{a} \in \mathcal{F}^{X,\kappa}} E^{\widetilde{a},\theta}\left[\int_t^\tau e^{-\gamma(s-t)}\left(dc_s - \widetilde{a}_s d\kappa_s\right) \big| \mathcal{F}_t^{X,\kappa}\right].$$

By Lemma 1, Z^θ is a Brownian Motion under $P^{\widetilde{a},\theta}$ and therefore we have for any admissible $\tilde{a} \in \mathcal{F}^{X,\kappa}$,

$$E^{\tilde{a},\theta}\left[\tilde{a}_s d\kappa_s \big| \mathcal{F}_t^{X,\kappa}\right] = E^{\tilde{a},\theta}\left[\tilde{a}_s\left(\theta_s ds + dZ_s^\theta\right) \big| \mathcal{F}_t^{X,\kappa}\right] = E^{\tilde{a},\theta}\left[\tilde{a}_s \theta_s ds \big| \mathcal{F}_t^{X,\kappa}\right],$$

since $E^{\tilde{a},\theta}[\tilde{a}_s dZ_s^\theta \big| \mathcal{F}_t^{X,\kappa}] = 0$. Thus, the agent's problem simplifies to

$$\sup_{\widetilde{a} \in \mathcal{F}^{X,\kappa}} E^{\widetilde{a},\theta}\left[\int_t^\tau e^{-\gamma(s-t)}\left(dc_s - \widetilde{a}_s \theta_s ds\right) \big| \mathcal{F}_t^{X,\kappa}\right].$$

Elliott (1977), Corollary 3.2, implies that the effort process $a^* \in \mathcal{F}^{X,\kappa}$ is optimal if and only if

$$\int_0^t e^{-\gamma s}\left(dc_s - a_s^* \theta_s ds\right) + e^{-\gamma t} W_t(\alpha, \theta)$$

is a martingale on the probability space $(\Omega, \mathcal{F}, P^{a^*,\theta})$ with respect to the filtration $\mathcal{F}^{X,\kappa}$. Elliott (1977), Theorem 4.3, implies that $a^* \in \mathcal{F}^{X,\kappa}$ is optimal if and only if there exists a $\mathcal{F}^{X,\kappa}$-progressive and square integrable process $\psi^* = \left\{\psi_{t,B}^*, \psi_{t,Z}^*\right\}_{t \geq 0}$ such that

$$\int_0^t e^{-\gamma s}\left(dc_s - a_s^* \theta_s ds\right) + e^{-\gamma t} W_t(\alpha, \theta) = \int_0^t e^{-\gamma s} \psi_{s,B} dB_s^{a^*} + \int_0^t e^{-\gamma s} \psi_{s,Z} dZ_s^\theta \tag{23}$$

Then, the optimal effort $a^* \in \mathcal{F}^{X,\kappa}$ satisfies

$$\psi_{t,B}^* \frac{\mu}{\sigma} a_t^* + \psi_{t,Z}^* \theta_t + \dot{c}_t - \theta_t a_t^* = \sup_{\tilde{a}_t} \psi_{t,B}^* \frac{\mu}{\sigma} \tilde{a}_t + \psi_{t,Z}^* \theta_t + \dot{c}_t - \theta_t \tilde{a}_t. \tag{24}$$

where $\dot{c}_t = \frac{dc_t}{dt}$ at each t.[55] It is worth to remark that a part of optimal contract (c, τ) are treated independent of agent's effort choice $\widetilde{a}$ in Equation (24). This is because, following a change in effort $\widetilde{a}_t$, the future consumption $\int_t^\tau dc_s$ and hence the continuation value W_t change only through a change in X_t—or equivalently, a change in measure $P^{\widetilde{a},\theta}$. This simplifies Equation (25) to

$$\psi^*_{t,B}\frac{\mu}{\sigma}a^*_t + \psi^*_{t,Z}\theta_t - \theta_t a^*_t = \sup_{\widetilde{a}_t} \psi^*_{t,B}\frac{\mu}{\sigma}\tilde{a}_t + \psi^*_{t,Z}\theta_t - \theta_t\tilde{a}_t. \tag{25}$$

Rewriting the RHS of Equation (25) shows that the Hamiltonian is linear in the effort $\widetilde{a}_t$, given $(\psi^*_{t,B}, \psi^*_{t,Z})$ and θ_t:

$$H_t(\widetilde{a}_t) = \big(\psi^*_{t,B}\frac{\mu}{\sigma} - \theta_t\big)\widetilde{a}_t + \psi^*_{t,Z}\theta_t.$$

Thus, the agent exerts effort $a^*_t = 1$ at t if and only if

$$\psi^*_{t,B}\frac{\mu}{\sigma} \geq \theta_t. \tag{26}$$

To show that $a^* \in \mathcal{F}^X$ without loss of generality, we use a guess-and-verify approach. We conjecture that $\{\psi^*_{t,B}, \psi^*_{t,Z}\}_{t\geq 0} = \{\psi_t, 0\}_{t\geq 0}$ for a $\mathcal{F}^X$-progressively measurable and square-integrable process $\psi = \{\psi_t\}_{t\geq 0}$. Such a process $\{\psi^*_{t,B}, \psi^*_{t,Z}\}_{t\geq 0}$ is progressively measurable and square integrable in $\mathcal{F}^{X,\kappa}$. We now show that the conjectured process $\{\psi^*_{t,B}, \psi^*_{t,Z}\}_{t\geq 0}$ satisfies (23) with the optimal effort process $a^* \in \mathcal{F}^X$ characterized by (26).

First, since $\psi \in \mathcal{F}^X$ and $\theta \in \mathcal{F}^X$, the Hamiltonian (25) implies that the agent's optimal effort satisfies $a^* \in \mathcal{F}^X$ without loss of generality. It remains to show that for such an $a^* \in \mathcal{F}^X$, the continuation value is $\mathcal{F}^X$-progressive as well, so that our conjecture of $\{\psi^*_{t,B}, \psi^*_{t,Z}\}_{t\geq 0} = \{\psi_t, 0\}_{t\geq 0}$ with $\psi \in \mathcal{F}^X$ is indeed correct.

The LHS of Equation (23) with effort $a^* \in \mathcal{F}^X$ implies that

$$E^{a^*,\theta}\left[\int_0^\tau e^{-\gamma s}\left(dc_s - \theta_s a^*_s ds\right)\big|\mathcal{F}^{X,\kappa}_t\right]$$

is the conditional expectation of the random variable $\int_0^\tau e^{-\gamma s}\left(dc_s - \theta_s a^*_s ds\right)$, whose integrand

[55]We allow $\dot{c}_t$ to be $\pm\infty$. Any admissible consumption process $\{c_t\}_{t\geq 0}$ is right continuous. If c_t is not left-continuous at t, set $\dot{c}_t = \infty$ when $\lim_{s\downarrow t} c_s > \lim_{s\uparrow t} c_s$ and $\dot{c}_t = -\infty$ when $\lim_{s\downarrow t} c_s < \lim_{s\uparrow t} c_s$.When c_t is not continuous, Equation (24) holds in the limit sense, with a truncated process $\dot{c}^M_t \equiv \max\{\min\{\frac{dc_t}{dt}, M\}, -M\}$ at each t. The optimal effort $a \in \mathcal{F}^{X,\kappa}$ is the limit of $a^M \in \mathcal{F}^{X,\kappa}$ as $M \to \infty$, when $a^M \in \mathcal{F}^{X,\kappa}$ satisfies

$$\psi^*_{t,B}\frac{\mu}{\sigma}a^*_t + \psi^*_{t,Z}\theta_t + \dot{c}^M_t - \theta_t a^*_t = \sup_{\widetilde{a}_t} \psi^*_{t,B}\frac{\mu}{\sigma}\tilde{a}_t + \psi^*_{t,Z}\theta_t + \dot{c}^M_t - \theta_t\tilde{a}_t.$$

is $\mathcal{F}^X$-progressive. Hence,

$$E^{a^*,\theta}\left[\int_0^\tau e^{-\gamma s}\left(dc_s - \theta_s a_s^* ds\right)\big|\mathcal{F}_t^{X,\kappa}\right] = E^{a^*,\theta}\left[\int_0^\tau e^{-\gamma s}\left(dc_s - \theta_s a_s^* ds\right)\big|\mathcal{F}_t^{X}\right],$$

which is a martingale with respect to $\mathcal{F}^X$. The proof is immediate from the Tower property of conditional expectation, given the integrability of the integrand $\{e^{-\gamma s}(dc_s - \theta_s a_s^* ds)\}_{s\geq 0}$. Finally, the martingale representation theorem[56] implies that there exists a process $\psi \in \mathcal{F}^X$ such that

$$\int_0^t e^{-\gamma s}\left(dc_s - \theta_s a_s^* ds\right) + e^{-\gamma t} W_t(\alpha,\theta) = E^{a^*,\theta}\left[\int_0^\tau e^{-\gamma s}\left(dc_s - \theta_s a_s^* ds\right)\big|\mathcal{F}_t^{X}\right] = \int_0^t e^{-\gamma s}\psi_s dB_s^{a^*}. \tag{27}$$

This verifies that the conjecture $\{\psi_{t,B}^*, \psi_{t,Z}^*\}_{t\geq 0} = \{\psi_t, 0\}_{t\geq 0}$ with $\psi \in \mathcal{F}^X$ is indeed correct, i.e. the conjecture satisfies Equation (23) at the agent's optimal effort a^*. Hence, a^* solves the agent's problem, follows from again applying Elliott (1977) Theorem 4.3.

Moreover, the agent's HJB equation (25) implies that $a_t^* = 1$ whenever

$$\psi_t \geq \frac{\sigma}{\mu}\theta_t.$$

Lastly, Equation (27) derives the differential equation for $W_t(\alpha,\theta)$:

$$e^{-\gamma t}(dc_t - \theta_t a_t^* dt) - \gamma e^{-\gamma t} W_t + e^{-\gamma t} dW_t = e^{-\gamma t}\psi_t dB_t^{a^*},$$

which can be rewritten as

$$dW_t = \big(\gamma W_t + a_t^*\theta_t\big)dt - dc_t + \psi_t dB_t^{a^*} = \big(\gamma W_t + a_t^*\theta_t\big)dt - dc_t + \psi_t \frac{1}{\sigma}(dX_t - \mu a_t^* dt).$$

Writing the agent's optimal effort a^* simply as a then yields Equation (7). This complete the proof.

A.3 Proof of Proposition 1

The proof consists of three steps. We show that the principal's HJB-Isaacs equation (10) has a unique solution with the properties described in the proposition, using a shooting method (e.g. Bailey et al. (1968)). Then, in Lemma 7 we show that the optimal contract and the worst-case effort cost solve the maxmin problem in Equation (10). Using the shape of the

[56]Here, we are using the Martingale Representation Theorem under $P^{a^*,\theta}$, which is obtained form P via Girsanov's change of measure. Doing so is standard in mathematical finance, see Shreve (2004), Corollary 5.3.2, p. 222, or Jeanblanc et al. (2009), Prop. 1.7.7.1 , p. 78.

value function, we then verify in Lemma 8 that the optimal contract and worst-case process solve the principal's problem in Equation (6).

A.3.1 Solution of the HJB Equation via the Shooting Method

Define the function $H : [0, W_{max}] \times \mathbb{R}^2 \to \mathbb{R}$ for W_{max} and ψ_{max} sufficiently large as

$$H(W, u, p) = -\min_{\psi \in [\bar{\psi}, \psi_{max}]} \max_{\theta \in [\underline{\theta}, \bar{\theta}]} \frac{ru - \mu - p(\gamma W + \theta)}{\frac{1}{2}\psi^2}. \tag{28}$$

The principal's HJB equation (10) can now be written as

$$J''(W) + H(W, J(W), J'(W)) = 0. \tag{29}$$

We define an initial value problem (IVP) consisting of Equation (29) and initial conditions $J(0) = L$ and $J'(0) = s$ for some $s \in \mathbb{R}$ instead of the boundary conditions provided in Proposition 1. Lemmas 3 and 4 present properties of the solution of the IVP (29), which is denoted by $J_s(.) : [0, W_{max}] \to \mathbb{R}$, with W_{max} chosen sufficiently large. Lemmas 5 and 6 establish that there exists a unique s^* such that the solution to the IVP (29) is equivalent to the solution to the boundary value problem (10) in Proposition 1, i.e. $J_{s*}(\cdot) = J(\cdot)$. This shooting method concludes that the principal's HJB equation (10) has a unique solution which satisfies the properties in Proposition 1. The proof is analogous to Szydlowski (2019), Section A.2.[57]

Lemma 3. *Fix a domain $[0, W_{max}]$ with W_{max} sufficiently large. Then, the initial value problem (IVP) in Equation (29) has a unique twice continuously differentiable solution on the domain $[0, W_{max}]$ for any starting slope s. Moreover, the solution is uniformly continuous with respect to s.*

Proof. The proof consists of showing that the function $H(W, u, p)$ satisfies the conditions of Hartman (2002), Th. 1.1, p. 8, which establishes the existence of a unique solution via the Picard-Lindelöf Theorem, and applying Hartman (2002), Th. 2.1, p. 94, which establishes continuity with respect to initial conditions.

Hartman (2002), Th. 1.1, p. 8 requires that $H(W, u, p)$ is continuous in (W, u, p) and Lipschitz continuous in (u, p), uniformly on the domain $[0, W_{max}]$.

That $H(W, u, p)$ is jointly continuous in (W, u, p) follows from successively applying Berge's maximum theorem (see Aliprantis and Border (2006), Th. 17.31, p. 570), first to the inner maximization and then the outer minimization.

[57] In particular, the Lemmas below and their proofs are nearly identical to Szydlowski (2019), Prop. 7, Lem. 2, and Lem. 3, p. 835-837. They are provided here as a convenience to the reader.

Moreover, $H(W,u,p)$ is Lipschitz continuous in (u,p), uniformly on $[0,W_{max}]$. Here is the argument. First, consider the inner maximization and write

$$\begin{aligned} f(W,u,p) &= \max_{\theta\in[\underline{\theta},\bar{\theta}]} ru-\mu-p(\gamma W+\theta) \\ &= ru-\mu-p\gamma W-p\underline{\theta}-\mathbb{1}\{p<0\}p\left(\bar{\theta}-\underline{\theta}\right). \end{aligned}$$

Clearly, we have

$$|f(W,u,p)-f(W,\tilde{u},\tilde{p})|\le K(|u-\tilde{u}|+|p-\tilde{p}|)$$

for some K sufficiently large whenever $p,\tilde{p}\ge 0$ or $p,\tilde{p}<0$.[58] When $p\ge 0>\tilde{p}$,

$$f(W,u,p)=ru-\mu-p\gamma W-p\underline{\theta}$$

and

$$f(W,\tilde{u},\tilde{p})=r\tilde{u}-\mu-p\gamma W-\tilde{p}\bar{\theta}.$$

Thus

$$|f(W,u,p)-f(W,\tilde{u},p)|\le\left(r|u-\tilde{u}|+|\gamma W||p-\tilde{p}|+\left|\tilde{p}\bar{\theta}-p\underline{\theta}\right|\right).$$

Since $p\ge 0>\tilde{p}$, we have

$$\left|\tilde{p}\bar{\theta}-p\underline{\theta}\right| \le p\underline{\theta}-\tilde{p}\bar{\theta}\le(p-\tilde{p})\,\bar{\theta}=|p-\tilde{p}|\,\bar{\theta}.$$

Thus,

$$|f(W,u,p)-f(W,\tilde{u},p)|\le K(|u-\tilde{u}|+|p-\tilde{p}|)$$

for some sufficiently large K. The case $p<0\le\tilde{p}$ is analogous. The above arguments establish that $f(W,u,p)$ is Lipschitz continuous in (u,p), uniformly on $[0,W_{max}]$.

Now, it remains to apply the same argument to

$$g(W,u,p)=\min_{\psi\in[\bar{\psi},\psi_{max}]}\frac{f(W,u,p)}{\frac{1}{2}\psi^2}.$$

Whenever $f(W,u,p)$ and $f(W,\tilde{u},\tilde{p})$ have the same sign, the minimizer ψ is the same for (u,p) and $(\tilde{u},\tilde{p})$. Therefore, the continuity and uniform-Lipschitz continuity of f implies the same properties of g. Specifically, we again have[59]

$$|g(W,u,p)-g(W,\tilde{u},\tilde{p})|\le K(|u-\tilde{u}|+|p-\tilde{p}|),$$

[58] Here, notice that the domain $[0,W_{max}]$ is finite.

[59] We write K as the Lipschitz constant again, with slight abuse of notation.

and g is Lipschitz on that region. We next consider the case when $f(W,u,p) \geq 0 > f(W,\tilde{u},\tilde{p})$. The same argument as before applies. In particular, we have

$$g(W,u,p) = \frac{f(W,u,p)}{\frac{1}{2}\psi_{max}^2}$$

and

$$g(W,\tilde{u},\tilde{p}) = \frac{f(W,\tilde{u},\tilde{p})}{\frac{1}{2}\bar{\psi}^2},$$

so that

$$\begin{aligned}
|g(W,u,p) - g(W,\tilde{u},\tilde{p})| &= \left| \frac{f(W,u,p)}{\frac{1}{2}\psi_{max}^2} - \frac{f(W,\tilde{u},\tilde{p})}{\frac{1}{2}\bar{\psi}^2} \right| \\
&= \frac{f(W,u,p)}{\frac{1}{2}\psi_{max}^2} - \frac{f(W,\tilde{u},\tilde{p})}{\frac{1}{2}\bar{\psi}^2} \\
&\leq \frac{f(W,u,p) - f(W,\tilde{u},\tilde{p})}{\frac{1}{2}\bar{\psi}^2} \\
&= \frac{|f(W,u,p) - f(W,\tilde{u},\tilde{p})|}{\frac{1}{2}\bar{\psi}^2} \\
&\leq \frac{K}{\frac{1}{2}\bar{\psi}^2}\left(|u-\tilde{u}| + |p-\tilde{p}|\right).
\end{aligned}$$

Thus, $g(W,u,p)$ is Lipschitz continuous for such values. The argument for the case $f(W,u,p) < 0 \leq f(W,\tilde{u},\tilde{p})$ is analogous.

Thus, Hartman (2002), Th. 1.1, p. 8 applies and guarantees the existence and uniqueness of a twice differentiable solution. Hartman (2002), Th. 2.1, p. 94, requires that $H(W,u,p)$ is continuous and that for any s, the IVP (29) has a unique solution. We have already established both properties. □

Now, we can use the shooting method. The boundary conditions $J'(\bar{W}) = -1$ and $J''(\bar{W}) = 0$ in the statement of Proposition 1 imply that $J(\bar{W}) = J_*(\bar{W})$, where $J_*(W)$ is given by

$$J_*(W) = \frac{\mu - \bar{\theta}}{r} - \frac{\gamma}{r}W.$$

We will characterize $\bar{W}$ as the first point at which $J(W)$ hits $J_*(W)$.

Pick a large negative number b. There exists a unique W_b such that $J_*(W_b) = b$. Let us define the set $\mathcal{B} \subset \mathbb{R}^2$ as

$$\mathcal{B} = \{(x,y) : x \in [0,W_b], y = b\} \cup \{(x,y) : x \in [0,W_b], y = J_*(x)\}.$$

Here, the first set on the RHS is the graph of a horizontal line with value b on the domain $[0, W_b]$ and the second set is the graph of $J_* (W)$ on the domain $[0, W_b]$.

Let us denote the solution to IVP (29) for any given s as $J_s (W)$ and define $\bar{W} (s) = \inf \{W : J_s (W) = J_* (W)\}$. The function $\bar{W} (s)$ is well defined whenever s is sufficiently large. We will show below Lemma 6 that there indeed exists an s at which $\bar{W} (s)$ is well defined.[60]

Lemma 4. *Any solution to the IVP (29) such that $0 > J_s' \left(\bar{W} (s)\right) \geq -1$ holds is strictly concave on $\left(0, \bar{W} (s)\right)$.*

Proof. Since s is kept constant throughout the proof, we simplify the notation $\bar{W} (s)$ to $\bar{W}$ and $J_s (\cdot)$ to $J (\cdot)$. First, suppose that $0 > J' \left(\bar{W}\right) > -1$. Then, the HJB Equation (10) implies that $J''(\bar{W}) < 0$, since otherwise $J(\bar{W}) > J_*(\bar{W})$. By continuity of $J'' (W)$, there exists an interval $(\widetilde{W}, \bar{W})$ such that $J'' (W) < 0$ for $W \in (\widetilde{W}, \bar{W})$. If we can pick $\widetilde{W} = 0$, we have established the result. By way of contradiction, define $\hat{W} = \sup \left\{W < \bar{W} : J'' (W) \geq 0\right\}$ and suppose that such a $\hat{W}$ exists. By continuity of $J'' (W)$, we have $J''(\hat{W}) = 0$. If $J'(\hat{W}) \geq 0$, the HJB equation implies that

$$rJ \left(\hat{W}\right) = \mu + J' \left(\hat{W}\right) \left(\gamma \hat{W} + \bar{\theta}\right) \geq \mu > rJ_* \left(\hat{W}\right),$$

which is a contradiction, since we defined $\bar{W}$ as the first point at which $J (\cdot)$ crosses $J_* (\cdot)$. If $J'(\hat{W}) < 0$, then we have

$$J_+''' \left(\hat{W}\right) = \frac{-\left(\gamma - r\right) J' \left(\hat{W}\right)}{\frac{1}{2}\bar{\psi}^2} > 0,$$

where $J_+''' (W)$ is the right derivative of $J'' (W)$, which is obtained by differentiating the HJB equation. But this implies that the function $J'' (W)$ cannot cross zero at $\hat{W}$ from above. In particular, we cannot have $J''(\hat{W}) = 0$ and $J''(\hat{W} + \varepsilon) < 0$ for any small $\varepsilon > 0$. Thus, we have another contradiction. In sum, the point $\hat{W}$ cannot exist and we have $J'' (W) < 0$ for all $W < \bar{W}$.

Now, consider the case $J' \left(\bar{W}\right) = -1$. Then, the HJB equation implies that $J'' \left(\bar{W}\right) = 0$, since otherwise $J(\bar{W}) \neq J_*(\bar{W})$. Calculating the left derivative of the HJB equation at $\bar{W}$ yields

$$\begin{aligned} \frac{1}{2}\bar{\psi}^2 J_-''' \left(\bar{W}\right) &= -\left(\gamma - r\right) J_-' \left(\bar{W}\right) - J_-'' \left(\bar{W}\right) \left(\gamma \bar{W} + \bar{\theta}\right) \\ &= -\left(\gamma - r\right) J_-' \left(\bar{W}\right) = \gamma - r > 0. \end{aligned}$$

[60] Intuitively, whenever s is very large, $J_s (W)$ is very steep and then is guaranteed to hit $J_* (W)$.

Thus, there exists an interval $(\widetilde{W}, \bar{W})$ such that $J''(W) < 0$ for $W \in (\widetilde{W}, \bar{W})$. The remainder of the proof is the same as in the previous case. $\square$

The remainder of Appendix A.3.1 shows that there exists a unique starting slope s^* that satisfies the boundary condition in Proposition 1: i.e., $J'_{s*}(\overline{W}(s^*)) = -1$, which concludes that the solution $J_{s*}(\cdot)$ of IVP (29) is a solution of the boundary value problem (10) in Proposition 1. To this end, we define a mapping $\mathcal{S}(s) = J'_s(\overline{W}(s)) : \mathbb{R} \to \mathbb{R}$. The argument is two-fold: Lemma 5 shows that $\mathcal{S}(s)$ is strictly decreasing in s, and Lemma 6 shows the existence of s^* such that $\mathcal{S}(s^*) = -1$, using the continuous mapping theorem.

Lemma 5. *There exists at most one s^* such that $J'_{s^*}(\bar{W}(s^*)) = -1$.*

Proof. Consider two initial slopes $s' > s$. We first show that $J_{s'}(W) > J_s(W)$ on $(0, W_{max})$. To see this, let $\hat{W} = \inf \{W : J'_{s'}(W) \le J'_s(W)\}$ and suppose by way of contradiction that such a $\hat{W}$ exists. By construction, we have $J'_{s'}(W) > J'_s(W)$ for all $W < \hat{W}$ and therefore $J_{s'}(\hat{W}) > J_s(\hat{W})$. Further, we have

$$H\left(\hat{W}, J_s(\hat{W}), J'_s(\hat{W})\right) = H\left(\hat{W}, J_s(\hat{W}), J'_{s'}(\hat{W})\right) > H\left(\hat{W}, J_{s'}(\hat{W}), J'_{s'}(\hat{W})\right),$$

since $H(W, u, p)$ is decreasing in u (see Equation (28)). Then, Equation (29) implies that $J''_{s'}(\hat{W}) > J''_s(\hat{W})$. This is a contradiction to the definition of $\hat{W}$: $J'_{s'}(\hat{W}) = J_s(\hat{W})$ and $J''_{s'}(\hat{W}) > J''_s(\hat{W})$ imply the existence of $\varepsilon > 0$ such that $J'_{s'}(\hat{W} + \varepsilon) < J'_s(\hat{W} + \varepsilon)$, but then $\hat{W} \neq \inf \{W : J'_{s'}(W) \le J'_s(W)\}$. Thus, we have a contradiction and $J'_{s'}(W) > J'_s(W)$ for all $W \in (0, W_{max})$. This, in turn, implies that $J_{s'}(W) > J_s(W)$, because $J_{s'}(0) = J_s(0) = L$. Since $J_{s'}(W) > J_s(W)$, we also have $\bar{W}(s') < \bar{W}(s)$, i.e. $\bar{W}(s)$ is strictly decreasing in s.

To establish the result in the statement of the lemma, suppose by way of contradiction that $s' > s$ and that $J'_{s'}(\bar{W}(s')) = J'_s(\bar{W}(s)) = -1$. Then, we have

$$-1 = J'_{s'}\left(\bar{W}(s')\right) > J'_s\left(\bar{W}(s')\right).$$

Since $\bar{W}(s') < \bar{W}(s)$ and since $J_s(W)$ is strictly concave on $(0, \bar{W}(s))$ by Lemma 4, it must be the case that $J'_s(\bar{W}(s')) > J'_s(\bar{W}(s))$, which together with the previous inequality implies that $-1 > J'_s(\bar{W}(s))$, a contradiction. $\square$

Lemma 6. *There exist two values $\bar{s} > \underline{s}$, such that for all $s \ge \bar{s}$, $\mathcal{S}(s) \ge 0$, and for all $s < \underline{s}$, $\mathcal{S}(s) \le -1$.*

Proof. Consider the mapping

$$\mathcal{T}(s) = \inf \{W : (W, J_s(W)) \in \mathcal{B}\}.$$

Intuitively, $\mathcal{T}(s)$ is the lowest value of W for which $J_s(W)$ hits the boundary $\mathcal{B}$. For $s \leq \underline{s}$ and $\underline{s}$ sufficiently small, $J_s(W)$ hits the constant line $\{(W, y) : W \in [0, W_b], y = b\}$. By choosing $\underline{s}$ sufficiently small, the hitting point can be made arbitrarily close to 0, i.e. $\mathcal{T}(\underline{s}) \downarrow 0$ as $\underline{s} \to -\infty$. Similarly, for $s > \bar{s}$ and $\bar{s}$ sufficiently large, $J_s(W)$ hits the function $J_*(W)$ and the hitting point can be made arbitrarily close to zero by picking $\bar{s}$ sufficiently large.

We now prove this, using Gronwall's Lemma. Fix $s > 0$ and suppose that $\widetilde{W} = \inf\{W : J_s'(W) \leq 0\} \in (0, W_b)$. On the interval $[0, \widetilde{W}]$, we have $J_s'(W) > 0$ and $J_s(W) \geq L > 0$. Then, Equation (28) implies that for any $W \in [0, \widetilde{W}]$,

$$0 = J_s''(W) + H(W, J_s(W), J_s'(W)) \leq J_s''(W) + H(W, 0, J_s'(W))$$

so that

$$J_s''(W) - \min_{\psi \in [\bar{\psi}, \psi_{max}]} \max_{\theta \in [\underline{\theta}, \bar{\theta}]} \frac{-\mu - J_s'(W)(\gamma W + \theta)}{\frac{1}{2}\psi^2} = J_s''(W) + \frac{\mu + J_s'(W)(\gamma W + \underline{\theta})}{\frac{1}{2}\bar{\psi}^2} \geq 0,$$

which implies that

$$J_s''(W) + \frac{\mu + J_s'(W)(\gamma W_b + \underline{\theta})}{\frac{1}{2}\bar{\psi}^2} \geq 0$$

for all $W \in [0, \widetilde{W}]$. Defining the auxiliary function $g(W) = -\left(J_s'(W) + \frac{\mu}{\gamma W_b + \underline{\theta}}\right)$, the above inequality is equivalent to

$$g'(W) \leq -\frac{\gamma W_b + \underline{\theta}}{\frac{1}{2}\bar{\psi}^2} g(W).$$

Then, Gronwall's Lemma implies that

$$g(W) \leq g(0) \exp\left(-\frac{\gamma W_b + \underline{\theta}}{\frac{1}{2}\bar{\psi}^2} \cdot W\right)$$

or equivalently

$$J_s'(W) \geq \left(s + \frac{\mu}{\gamma W_b + \underline{\theta}}\right) \exp\left(-\frac{\gamma W_b + \underline{\theta}}{\frac{1}{2}\bar{\psi}^2} \cdot W\right) - \frac{\mu}{\gamma W_b + \underline{\theta}}$$

for all $W \in [0, \widetilde{W}]$. Thus, by picking s arbitrarily large, we can guarantee that $J_s(W)$ is arbitrarily steep on $[0, W_b]$. Because of this, we can pick a sufficiently large $\bar{s}$ such that $\widetilde{W} > W_b$ and $J_{\bar{s}}'(W) \geq 0$ for $W \in (0, W_b)$. For any $s \geq \bar{s}$, $J_s(W)$ hits $J_*(W)$ arbitrarily close to 0. The argument establishing $\underline{s}$ is analogous.

Now, we can define for some small $\varepsilon > 0$,

$$\mathcal{B}_\varepsilon = \{(W, y) \in \mathcal{B} : W \geq \varepsilon\}.$$

By the preceding arguments, there exists a sufficiently small $\underline{s}$ such that $J_{\underline{s}}(W)$ hits the line $\{(y, W) : y = b\}$ at $W = \varepsilon$ and there exists a sufficiently large $\bar{s}$ such that $J_{\bar{s}}(\bar{W}(\bar{s})) = J_*(\bar{W}(\bar{s}))$ and $\bar{W}(\bar{s}) = \varepsilon$. Since the solution to the IVP (29) is uniformly continuous with respect to s, $\mathcal{T}(s)$ is continuous for $s \in [\underline{s}, \bar{s}]$. Since the set $\mathcal{B}_\varepsilon$ is compact, the continuous mapping theorem implies that $\mathcal{T}(s)$ is onto, i.e. $\mathcal{T}([\underline{s}, \bar{s}]) = \mathcal{B}_\varepsilon$. Thus, there exists a subinterval of $[\underline{s}, \bar{s}]$ for which $J_s(W)$ hits $J_*(W)$.

Now, we establish that there exists an $\underline{s}$ such that $\mathcal{S}(\underline{s}) = J_{\underline{s}}(\bar{W}(\underline{s})) \leq -1$. Pick $\hat{W}$ sufficiently close to W_b, so that $J_*(\hat{W}) = -M > b$ for some sufficiently large $M > 0$.[61] By the preceding argument, there exists a slope s such that $\bar{W}(s) = \hat{W}$, i.e. $J_s(\bar{W}(s)) = J_*(\hat{W})$. Suppose by way of contradiction that $0 > J_s'(\hat{W}) > -1$. Then, by Lemma 4, $J_s(W)$ is strictly concave on $(0, \hat{W})$ and we have $J_s'(W) > -1$ for $W < \hat{W}$. This implies that $J_s(\hat{W}) > J_s(0) - \hat{W}$. Since $J_*'(\hat{W}) = -\frac{\gamma}{r} < -1$ and since $\hat{W}$ can be chosen so that $J_*(\hat{W}) = -M$ is negative and large, we have $J_s(0) - \hat{W} > J_*(\hat{W})$.[62] But then $J_s(W)$ cannot hit $J_*(W)$ at $\hat{W}$, because $J_s(\hat{W}) > J_s(0) - \hat{W} > J_*(\hat{W})$. Thus, we must have either $J_s'(\hat{W}) \geq 0$ or $J_s'(\hat{W}) \leq -1$.

Again by way of contradiction, suppose that $J_s'(\hat{W}) \geq 0$. Then, it must be the case that $J_s''(\hat{W}) < 0$, otherwise Equation (29) implies that $J_s(\hat{W}) \geq \frac{\mu}{r} > J_*(\hat{W})$, a contradiction. If $J_s''(W) < 0$ for all $W < \hat{W}$, then $J_s'(\hat{W}) \geq 0$ implies that $J_s'(W) > 0$ for all $W < \hat{W}$, so that $J_s(W) \geq L > 0$ for all $W \in (0, \hat{W})$. But then, we cannot have $J_s(\hat{W}) = J_*(\hat{W})$, since we have chosen $\hat{W}$ so that $J_*(\hat{W}) < 0$. Thus, there must exist some $W < \hat{W}$ such that $J_s''(W) > 0$ and $J_s'(W) > 0$. By continuity of $J_s'(\cdot)$ and $J_s''(\cdot)$, there exists an interval of W such that $J_s''(W) > 0$ and $J_s'(W) > 0$ on that interval. But then, Equation (29) implies that $J_s(W) > \frac{\mu}{r} > J_*(W)$ for any such W, which is a contradiction. Thus, we cannot have $J_s'(\hat{W}) \geq 0$. Together with the previous argument, this establishes that $J_s'(\hat{W}) \leq -1$ and therefore that there exists an $\underline{s}$ such that $\mathcal{S}(\underline{s}) \leq -1$. An analogous argument establishes that $\mathcal{S}(\bar{s}) \geq 0$ for $\bar{s}$ sufficiently large. □

We can now use the continuous mapping theorem on $\mathcal{S}(s)$ for $s \in [\underline{s}, \bar{s}]$. Since $\mathcal{S}(\underline{s}) \leq -1$, $\mathcal{S}(\bar{s}) > -1$, and $\mathcal{S}(s)$ is continuous, there exists an s^* such that $\mathcal{S}(s^*) = -1$. By Lemma 5, this s^* is unique.

Together, these arguments establish that there exists a unique starting slope s^*, such that $J_{s^*}(W)$ satisfies the boundary conditions $J_{s^*}(0) = 0$, $J_{s^*}'(\bar{W}) = -1$, $J_{s^*}''(\bar{W}) = 0$ and

[61] Recall that throughout the argument, we are picking b and W_b to be sufficiently large.

[62] Here, notice that $J_*(W) + W$ is strictly decreasing in W.

that $J_{s^*}(W)$ is strictly concave on $(0,\bar{W})$. Thus, $J_{s^*}(W)$ solves the HJB equation (10) with the above boundary conditions.

A.3.2 Verification

We first characterize necessary conditions for optimality.

Lemma 7. *For any incentive compatible contract which implements effort at time $t \leq \tau$, we have $\psi_t \geq \bar{\psi}$. The worst-case effort cost is given by*

$$\theta(\psi_t, W_t) = \begin{cases} \bar{\theta} & \text{if} \quad J'(W_t) \leq 0 \quad \text{and} \quad \psi_t \geq \bar{\psi} \\ \underline{\theta} & \text{if} \quad J'(W_t) > 0 \quad \text{and} \quad \psi_t \geq \bar{\psi} \\ \bar{\theta} & \text{if} \quad \psi_t < \bar{\psi}. \end{cases}$$

Any contract with $\psi_t > \bar{\psi}$ is suboptimal.

Proof. First, a contract that features $\psi_t < \bar{\psi}$ and implements effort (i.e., $a_t = 1$) at the same time cannot be incentive compatible. This is because the worst-case effort cost at time t is

$$\theta(\psi_t, W_t) \in \arg\min_{\theta \in [\underline{\theta}, \bar{\theta}]} (\mu + J'(W_t)\theta)\, a(\psi_t, \theta) + J'(W_t)\gamma W_t + J''(W_t)\frac{1}{2}\psi_t^2, \tag{30}$$

where $a(\psi_t, \theta_t) = \mathbb{1}\{\psi_t \geq \frac{\sigma}{\mu}\theta_t\}$ is the agent's optimal choice of effort. Because $J'(W_t) \geq -1$ and $\mu > \bar{\theta}$, the term multiplying $a(\psi_t, \theta)$ in the above equation is positive, i.e.

$$\mu + J'(W_t)\theta \geq 0$$

for all $\theta \in [\underline{\theta}, \bar{\theta}]$. Thus, the RHS of Equation (30) is minimized by picking any $\theta(\psi_t, W_t) \in (\frac{\sigma}{\mu}\psi_t, \bar{\theta}]$, which induces effort $a_t = 0$ whenever $\psi_t < \bar{\psi}$. Thus, no contract with $\psi_t < \bar{\psi}$ is incentive compatible. Hence, we have $\psi_t \geq \bar{\psi}$ for all $t \leq \tau$ for any incentive compatible contract.

If $\psi_t \geq \bar{\psi}$, we have $a(\psi_t, \theta_t) = 1$ for any $\theta_t \in [\underline{\theta}, \bar{\theta}]$. Then, the minimizer $\theta(\psi_t, W_t)$ satisfies

$$\theta(\bar{\psi}, W_t) \equiv \theta^*(W_t) = \begin{cases} \bar{\theta} & \text{if} \quad J'(W_t) \leq 0 \\ \underline{\theta} & \text{if} \quad J'(W_t) > 0 \end{cases}$$

for any $\psi_t \geq \bar{\psi}$.

Finally, we show that any contract with $\psi_t > \bar{\psi}$ is suboptimal. Because $J(W)$ is strictly concave on $[0, \bar{W})$, we have

$$\bar{\psi} = \arg\max_{\psi_t \geq \bar{\psi}} (\mu + J'(W_t)\theta(\psi_t, W_t)) + J'(W_t)\gamma W_t + J''(W_t)\frac{1}{2}\psi_t^2$$

given that the contract implements $a_t = 1$. Thus, at the optimal contract, we have $\psi_t = \bar{\psi}$ and any $\psi_t > \bar{\psi}$ is suboptimal. □

The above result establishes Equation (9) in the statement of Proposition 1. Overall, the above arguments establish that

$$\bar{\psi} = \max_{\psi_t \geq \bar{\psi}} \min_{\theta \in [\underline{\theta}, \bar{\theta}]} \left(\mu + J'(W_t)\theta\right) a(\psi_t, \theta) + J'(W_t)\gamma W_t + J''(W_t)\frac{1}{2}\psi_t^2.$$

In particular, we have

$$\begin{aligned} rJ(W_t) \geq\ & \mu a(\psi_t, \theta(\psi_t, W_t)) + J'(W_t)\left(\gamma W_t + \theta(\psi_t, W_t) a(\psi_t, \theta(\psi_t, W_t))\right) \\ & + \frac{1}{2} J''(W_t)\psi_t^2 \end{aligned}$$

for any $\psi_t \geq 0$. It remains to show that this solution equals the principal's payoff in the optimal contract. This follows from a verification argument, suitably modified to account for ambiguity aversion. We present this argument next.

Lemma 8. *The solution to the HJB equation (10) equals the principal's value function in Equation (6).*

Proof. We now verify that the contract is optimal.[63] Consider an arbitrary incentive compatible contract α and let $\theta(\alpha) = \{\theta_t(\alpha)\}_{t\geq 0}$ be the minimizer in Equation (6). Denote with $P^{a,\theta(\alpha)}$ the measure under the pair $(a, \theta(\alpha))$. Define the value from following this contract until time $t \leq \tau$ and then changing to the candidate for the optimal contract as

$$G_t(\alpha, \theta(\alpha)) = \int_0^t e^{-rs}(dX_s - dc_s) + e^{-rt}J(W_t). \tag{31}$$

Here, $J(W)$ is the solution to the principal's HJB equation (10). Note that $J(W_\tau) = L$ for $t = \tau$ since $W_\tau = 0$. Formally, we have

$$dW_s = \begin{cases} (\gamma W_s + \theta_s(\alpha))\, ds - dc_s + \psi_s dB_s^a & \text{for} \quad s \leq t \\ (\gamma W_s + \theta_s^*)\, ds - dc_s^* + \bar{\psi} dB_s^{a^*} & \text{for} \quad s > t \end{cases} \tag{32}$$

and we define processes $\tilde{\alpha}_s = \mathbb{1}\{s \leq t\}\alpha_s + \mathbb{1}\{s > t\}\alpha_s^*$ and $\tilde{\theta}_s = \mathbb{1}\{s \leq t\}\theta_s(\alpha) + \mathbb{1}\{s > t\}\theta_s^*$ for $s \leq \tau$.

[63]The proof is similar to the one in DeMarzo and Sannikov (2006), p. 2712f. It is modified to account for ambiguity.

Using Itô's Lemma under $P^{a,\theta(\alpha)}$ together with Equation (32) yields

$$\begin{aligned} dG_t(\alpha, \theta(\alpha)) &= e^{-rt}\Big(\mu a_t dt + \sigma dB_t^a - dc_t + J'(W_t)\left((\gamma W_t + \theta_t(\alpha) a_t) dt - dc_t + \psi_t dB_t^a\right) \\ &\quad + J''(W_t)\frac{1}{2}\psi_t^2 dt - rJ(W_t) dt\Big). \end{aligned}$$

The solution of the HJB equation (10) is strictly concave on $[0, \bar{W})$ and satisfies $J'(W) = -1$ for $W \geq \bar{W}$, since $J(W) = J(\bar{W}) - (W - \bar{W})$ for $W \geq \bar{W}$. It follows that $-(1 + J'(W_t)) dc_t \leq 0$ for any $W_t \in (0, \bar{W})$ and $-(1 + J'(W_t)) dc_t = 0$ for $W_t \geq \bar{W}$. Also, we have

$$\mu a_t + J'(W_t)(\gamma W_t + \theta_t(\alpha) a_t) + J''(W_t)\frac{1}{2}\psi_t^2 \leq rJ(W_t),$$

which follows from Equation (10). Taken together, these inequalities imply that

$$\begin{aligned} & e^{rt} E^{a,\theta(\alpha)}\left[dG_t(\alpha, \theta(\alpha))\right] \\ = \ & E^{a,\theta(\alpha)}\Bigg[\left(\mu a_t + J'(W_t)(\gamma W_t + \theta_t(\alpha) a_t) + J''(W_t)\frac{1}{2}\psi_t^2 - rJ(W_t)\right) dt \\ & -(1 + J'(W_t)) dc_t + (\sigma + \psi_t J'(W_t)) dB_t^a\Bigg] \\ \leq \ & E^{a,\theta(\alpha)}\left[(\sigma + \psi_t J'(W_t)) dB_t^a\right]. \end{aligned}$$

The solution to the principal's HJB equation $J(W)$ is strictly concave on $[0, \bar{W})$, $J'(W) = -1$ for $W \geq \bar{W}$ and $J'(W) \geq -1$ for $W < \bar{W}$. Moreover, Lemma 6 implies that the solution to the HJB equation has a finite starting slope $J'(0)$. Thus, $J'(W)$ is bounded for all $W \geq 0$ and the term $\psi_t J'(W_t)$ is square integrable. This implies that

$$E^{a,\theta(\alpha)}\left[(\sigma + \psi_t J'(W_t)) dB_t^a\right] = 0,$$

since B^a is a Brownian Motion under $P^{a,\theta(\alpha)}$. Thus, we have shown that

$$E^{a,\theta(\alpha)}\left[dG_t(\alpha, \theta(\alpha))\right] \leq 0$$

so that $G_t(\alpha, \theta(\alpha))$ is a $P^{\alpha,\theta(\alpha)}$-supermartingale. This implies that

$$G_0(\alpha, \theta(\alpha)) \geq E^{a,\theta(\alpha)}\left[G_{t\wedge\tau}(\alpha, \theta(\alpha))\right]$$

for any fixed $t \geq 0$. Taking $t \to \infty$ yields

$$G_0(\alpha, \theta(\alpha)) \geq E^{a,\theta(\alpha)}\left[G_\tau(\alpha, \theta(\alpha))\right].$$

By construction of $G_t(\alpha, \theta(\alpha))$ in Equation (31), we have $G_0(\alpha, \theta(\alpha)) = J(W_0)$ for any $(\alpha, \theta(\alpha))$. Thus,

$$J(W_0) = G_0(\alpha, \theta(\alpha)) \geq E^{a,\theta(\alpha)}\left[G_\tau(\alpha, \theta(\alpha))\right] \geq \inf_{\theta \in \Theta} E^{a,\theta}\left[G_\tau(\alpha, \theta)\right],$$

where the last inequality follows from taking the infimum over $\theta \in \Theta$. Taking the supremum over contracts $\alpha \in \mathcal{A}$ then yields

$$J(W_0) \geq \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \Theta} E^{a,\theta}\left[G_\tau(\alpha, \theta)\right]. \tag{34}$$

We now establish the reverse inequality to (34). Consider the optimal contract α^* , with effort $a^* = (a_t^*)_{t \geq 0}$ with $a_t^* = 1$ for all $t < \tau$, and an arbitrary density generator $\theta \in \Theta$. Define $G_t(\alpha^* \theta)$ analogously as

$$G_t(\alpha^*, \theta) = \int_0^t e^{-rs}(dX_s - dc_s) + e^{-rt} J(W_t)$$

for $t \leq \tau$. Under $P^{a^*,\theta}$, Ito's Lemma yields

$$\begin{aligned} dG_t(\alpha^*, \theta) &= e^{-rt}\Big(\mu dt + \sigma dB_t^{a^*} - dc_t^* + J'(W_t)\left((\gamma W_t + \theta_t)\, dt - dc_t^* + \bar{\psi} dB_t^{a^*}\right) \\ &\quad + J''(W_t)\frac{1}{2}\bar{\psi}^2 dt - rJ(W_t)\, dt\Big). \end{aligned}$$

For any $W \in (0, \bar{W})$, we have

$$\begin{aligned} rJ(W_t) &= \min_{\theta \in [\underline{\theta}, \bar{\theta}]} \mu + J'(W_t)(\gamma W_t + \theta) + J''(W_t)\frac{1}{2}\bar{\psi}^2 \\ &\leq \mu + J'(W_t)(\gamma W_t + \theta_t) + J''(W_t)\frac{1}{2}\bar{\psi}^2 \end{aligned}$$

where $\theta_t \in [\underline{\theta}, \bar{\theta}]$ is arbitrary in the second line. Then, by a similar argument as in the previous case, $G_t(\alpha^*, \theta)$ is a $P^{a^*,\theta}$-submartingale and we have $G_0(\alpha^*, \theta) \leq E^{a^*,\theta}\left[G_{t \wedge \tau}(\alpha^*, \theta)\right]$. Taking $t \to \infty$ then yields $G_0(\alpha^*, \theta) \leq E^{a^*,\theta}[G_\tau(\alpha^*, \theta)]$. As in the previous case, we have $J(W_0) = G_0(\alpha^*, \theta)$ and therefore $J(W_0) \leq E^{a^*,\theta}\left[G_\tau(\alpha^*, \theta)\right]$. Taking the infimum over $\theta \in \Theta$

implies that

$$J(W_0) \leq \inf_{\theta\in\Theta} E^{a^*,\theta}\left[G_\tau(\alpha^*,\theta)\right] \leq \sup_{\alpha\in\mathcal{A}}\inf_{\theta\in\Theta} E^{a,\theta}\left[G_\tau(\alpha,\theta)\right]. \tag{35}$$

Taken together, Equations (34) and (35) imply that

$$\begin{aligned} J(W_0) &= \sup_{\alpha\in\mathcal{A}}\inf_{\theta\in\Theta} E^{a,\theta}\left[G_\tau(\alpha,\theta)\right] \\ &= \sup_{\alpha\in\mathcal{A}}\inf_{\theta\in\Theta} E^{a,\theta}\left[\int_0^\tau e^{-rt}(dX_t - dc_t) + e^{-r\tau}L\right]. \end{aligned}$$

The last equation follows because $\tau = \inf\{t : W_t = 0\}$ and $J(W_\tau) = L$.

It remains to verify that under the optimal contract α^* and the density generator $\theta^* = \{\theta_t^*\}_{t\geq 0}$, where $\theta_t^* = \theta(\bar{\psi}, W_t)$ for all $t < \tau$, the process $G_t(\alpha^*,\theta^*)$ is a P^{a^*,θ^*}-martingale. Under P^{a^*,θ^*}, we have

$$\begin{aligned} dG_t(\alpha^*,\theta^*) &= e^{-rt}\Big(\mu dt + \sigma dB_t^{a^*} - dc_t^* + J'(W_t)\left(\left(\gamma W_t + \theta\left(\bar{\psi}, W_t\right)\right)dt - dc_t^* + \bar{\psi} dB_t^{a^*}\right) \\ &\quad + J''(W_t)\frac{1}{2}\bar{\psi}^2 dt - rJ(W_t)\,dt\Big). \end{aligned}$$

The definition of the HJB equation (10) implies that

$$\begin{aligned} rJ(W_t) &= \max_{\psi_t\geq\bar{\psi}}\min_{\theta_t\in[\underline{\theta},\bar{\theta}]} \mu a(\psi_t,\theta_t) + J'(W_t)\left(\gamma W_t + \theta_t a(\psi_t,\theta_t)\right) + \frac{1}{2}J''(W_t)\psi_t^2 \\ &= \mu + J'(W_t)\left(\gamma W_t + \theta_t\left(\bar{\psi}, W_t\right)\right) + \frac{1}{2}J''(W_t)\bar{\psi}^2 \end{aligned}$$

and we have $dc_t \geq 0$ whenever $J'(W_t) = -1$, which is the case for $W_t \geq \bar{W}$, and $dc_t = 0$ otherwise. Thus, we have

$$\begin{aligned} & e^{rt}E^{a^*,\theta^*}\left[dG_t(\alpha^*,\theta^*)\right] \\ =\; & E^{a^*,\theta^*}\Bigg[\left(\mu + J'(W_t)\left(\gamma W_t + \theta_t\left(\bar{\psi}, W_t\right)\right) + \frac{1}{2}J''(W_t)\bar{\psi}^2 - rJ(W_t)\right)dt \\ & -dc_t\left(J'(W_t) + 1\right) + \sigma dB_t^{a^*}\left(\sigma + \bar{\psi}J'(W_t)\right)\Bigg] \\ =\; & E^{a^*,\theta^*}\left[\sigma dB_t^{a^*}\left(\sigma + \bar{\psi}J'(W_t)\right)\right] \\ =\; & 0, \end{aligned}$$

i.e. $G_t(\alpha^*,\theta^*)$ is a P^{a^*,θ^*}-martingale. This implies that $J(W_0) = G_0 = E^{a^*,\theta^*}\left[G_\tau(\alpha^*,\theta^*)\right]$. Thus, α^* is the optimal contract and θ^* is the worst-case effort cost. □

A.4 Proof of Proposition 2

Let (a^*, c^*, τ^*) denote the robust optimal contract and θ^* the worst-case effort cost. Similarly, let (a, c, τ) denote the optimal contract without ambiguity aversion. The latter contract is identical to the optimal contract in DeMarzo and Sannikov (2006). In either setting, we can write any continuation value W_t for $t \leq \min\{\tau^*, \tau\}$ as

$$W_t = E^{a^*,\theta^*}\left[\int_t^{\tau^*} e^{-\gamma(s-t)} dc_s^* \middle| \mathcal{F}_t^{X,\kappa}\right] - E^{a^*,\theta^*}\left[\int_t^{\tau^*} e^{-\gamma(s-t)} \theta_s^* ds \middle| \mathcal{F}_t^{X,\kappa}\right]$$

and

$$W_t = E\left[\int_t^{\tau} e^{-\gamma(s-t)} dc_s \middle| \mathcal{F}_t^{X,\kappa}\right] - E\left[\int_t^{\tau} e^{-\gamma(s-t)} \theta_0 ds \middle| \mathcal{F}_t^{X,\kappa}\right].$$

Denote the expected compensation under the worst-case and the contract without ambiguity as

$$g^*(W_t) = E^{a^*,\theta^*}\left[\int_t^{\tau} e^{-\gamma(s-t)} dc_s^* \middle| \mathcal{F}_t^{X,\kappa}\right]$$

and

$$g(W_t) = E\left[\int_t^{\tau} e^{-\gamma(s-t)} dc_s \middle| \mathcal{F}_t^{X,\kappa}\right].$$

The continuation value W_t is a sufficient statistic for expected compensation, i.e. the functions $g(\cdot)$ and $g^*(\cdot)$ depend on the contract through W_t alone (see Proposition 1). The proposition claims that there is a value of θ_0 such that

$$g^*(W_t) < g(W_t)$$

whenever W_t is sufficiently low and that the inequality is reversed whenever W_t is sufficiently high.

We first record basic properties of $g(W)$ and $g^*(W)$. Since the agent is fired at $W = 0$ and hence receives no further payments, we have $g(0) = g^*(0) = 0$. Moreover, the construction of $g^*(W_t)$ and $g(W_t)$ imply, respectively

$$g^*(W_t) = W_t + E^{a^*,\theta^*}\left[\int_t^{\tau^*} e^{-\gamma(s-t)} \theta_s^* ds \middle| \mathcal{F}_t^{X,\kappa}\right]$$

and

$$g(W_t) = W_t + E\left[\int_t^{\tau} e^{-\gamma(s-t)} \theta_0 ds \middle| \mathcal{F}_t^{X,\kappa}\right],$$

so that $g^*(W_t) > W_t$ and $g(W_t) > W_t$ for all $W_t > 0$. Under P^{a^*,θ^*}, $g^*(W)$ is the unique solution to the ODE

$$\gamma g^*(W) = g^{*\prime}(W)(\gamma W + \theta^*(W)) + g^{*\prime\prime}(W)\frac{1}{2}\bar{\psi}^2 \tag{36}$$

with boundary conditions $g^*(0) = 0$ and $g^{*\prime}(\bar{W}^*) = 1$, where $\bar{W}^*$ is the payout boundary in the optimal contract under ambiguity. This follows directly from applying the Feynman-Kac Theorem under measure P^{a^*,θ^*}. Similarly, $g(W)$ is the unique solution to

$$\gamma g(W) = g'(W)(\gamma W + \theta_0) + g''(W)\frac{1}{2}\psi^2 \tag{37}$$

with $g(0) = 0$ and $g'(\bar{W}) = 1$, where $\psi = \theta_0\sigma/\mu$ and $\bar{W}$ is the payout boundary in the contract without ambiguity. Both $g(W)$ and $g^*(W)$ are linear functions with slope one for continuation values right of their respective payout boundaries, i.e. $g'(W) = 1$ for $W \geq \bar{W}$ and $g^{*\prime}(W) = 1$ for $W \geq \bar{W}^*$. Standard arguments imply that $g(W)$ is uniformly continuous in its parameters on any bounded interval $[0, W_{max}]$. In particular, $g(W)$ is uniformly continuous in θ_0. In the following, we write $g(W, \theta_0)$ to denote the dependence of g on the parameter θ_0. Similarly, $g^*(W)$ is uniformly continuous in $\bar{\theta}$ and $\underline{\theta}$.

Moreover, $g(W)$ is strictly concave on $[0, \bar{W}]$. To see this, differentiate Equation (37) on $[0, \bar{W}]$ to yield

$$g'''(W)\frac{1}{2}\psi^2 = g''(W)(\gamma W + \theta_0),$$

so that $g'''(W)$ and $g''(W)$ have the same sign. This implies that $g''(W)$ cannot cross zero on $[0, \bar{W}]$. Thus, $g(W)$ is either strictly concave, strictly convex, or linear. Since $g(W) > W$ for $W > 0$, it must be the case that $g'(0) > 1 = g'(\bar{W})$. Thus, $g(W)$ is strictly concave.

We now present the main argument. Assume that $\underline{\theta} = 0$. If $\theta_0 = \bar{\theta}$, then $g(W, \bar{\theta}) > g^*(W)$ for all $W > 0$. Intuitively, $g(W, \bar{\theta})$ has the same volatility but a higher drift than $g^*(W)$, so the agent gets paid more in expectation. If $\theta_0 = 0$, then the unique solution to the ODE (37) is $g(W) = W$. This is immediate, since for $\theta_0 = 0$ the agent faces no effort cost and the contract in DeMarzo and Sannikov (2006) pays him an immediate lump sum of W. In particular, if $\theta_0 = 0$, then $\bar{W} = 0$. Moreover, when $\underline{\theta} = 0$, $g^*(W)$ is linear on $[0, W^{0*})$, where W^{0*} is the point at which $\theta^*(W)$ exhibits a jump. This follows by plugging in $\theta^*(W) = 0$ for $W < W^{0*}$ into (36) and picking any $\hat{g} > W^{0*}$ as a boundary condition, i.e. $g^*(W) = C \cdot W$ with $C > 1$ is the unique solution to the ODE (36) on $[0, W^{0*}]$ with boundary conditions $g^*(0) = 0$ and $g^*(W^{0*}) = \hat{g} > W^{0*}$.

Since $g^*(W)$ does not depend on θ_0, it holds that $g(W, \bar{\theta}) > g^*(W)$ for $W \in (0, W^{0*}]$ and that $g(W, 0) = W < g^*(W)$ for $W \in (0, W^{0*}]$. We now consider the following cases (1) $\bar{W} \leq W^{0*}$ for $\theta_0 = \bar{\theta}$ and (2) $\bar{W} > W^{0*}$ for $\theta_0 = \bar{\theta}$. In the latter case, since $g(W, \theta_0)$ is uniformly continuous in θ_0, there exists a θ_0 such that $\bar{W} = W^{0*}$. Suppose that $g(W^{0*}, \theta_0) \leq g^*(W^{0*})$. Then, since $g^*(W)$ is linear on $[0, W^{0*}]$ and $g(W, \theta_0)$ is strictly concave, there exists a $\hat{W} \leq W^{0*}$ such that $g(W, \theta_0) > g^*(W)$ for $W \in (0, \hat{W})$ and $g(W, \theta_0) < g^*(W)$ for $W \in (\hat{W}, W^{0*})$.[64] We now show that $g^*(W) > g(W, \theta_0)$ for $W > W^{0*}$. Since $\bar{W} = W^{0*}$, we have $g'(W, \theta_0) = 1$ for all $W > \bar{W} = W^{0*}$. On the interval $[W^{0*}, \bar{W}^*]$, $g^*(W)$ is strictly concave. To see this, differentiate Equation (36) for $W \in (W^{0*}, \bar{W}^*)$ to obtain

$$g^{*\prime\prime\prime}(W) \frac{1}{2} \bar{\psi}^2 = g^{*\prime\prime}(W) \left(\gamma W + \bar{\theta}\right),$$

which implies that $g^{*\prime\prime}(W)$ cannot cross zero. Thus, $g^*(W)$ is either strictly concave, strictly convex, or linear on $(W^{0*}, \bar{W}^*)$. Since $g^{*\prime}(W^{0*}) > 1 = g^{*\prime}(\bar{W}^*)$, $g^*(W)$ must be strictly concave on $(W^{0*}, \bar{W}^*)$. But then, we must have $g^{*\prime}(W) > 1 = g'(W, \theta_0)$ for all $W > W^{0*}$. Since $g^*(W^{0*}) > g(W^{0*}, \theta_0)$, it follows that $g^*(W) > g(W, \theta_0)$ for $W \geq W^{0*}$.

Now, suppose that $g(W^{0*}, \theta_0) > g^*(W^{0*})$ if $\bar{W} = W^{0*}$. Then, there exists a θ_0 such that $\bar{W} < W^{0*}$, $\hat{W} = \inf \{W > 0 : g(W, \theta_0) = g^*(W)\}$ exists, and $\hat{W} < W^{0*}$. Since $g(W, \theta_0)$ is concave and $g^*(W)$ is linear on $[0, W^{0*}]$, we have indeed $g(W, \theta_0) > g^*(W)$ for all $W \in (0, \hat{W})$ and $g(W, \theta_0) < g^*(W)$ for all $W \in (\hat{W}, W^{0*}]$. Since $\bar{W} < W^{0*}$, it holds that $g'(W^{0*}, \theta_0) = 1$ and $g(W^{0*}, \theta_0) < g^*(W^{0*})$. Then, the same argument as in the previous case establishes that $g(W, \theta_0) < g^*(W)$ for all $W > W^{0*}$.

Overall, we have just shown that $g^*(W) < g(W, \theta_0)$ for $W < \hat{W}$ and $g^*(W) > g(W, \theta_0)$ for $W > \hat{W}$ whenever $\bar{W} > W^{0*}$, which is the result we set out to establish. Now, consider the case $\bar{W} \leq W^{0*}$ for $\theta_0 = \bar{\theta}$. Fix some $\hat{W} \in (0, W^{0*})$. By uniform continuity of $g(W, \theta_0)$ in θ_0, there exists a θ_0 such that $g(\hat{W}, \theta_0) = g^*(\hat{W})$. Since $g(W, \theta_0)$ is strictly concave and $g^*(W)$ is linear on $[0, W^{0*}]$, we have $g(W, \theta_0) > g^*(W)$ for $W \in (0, \hat{W})$ and $g(W, \theta_0) < g^*(W)$ for $W \in (\hat{W}, W^{0*})$. The remainder of the proof is then the same as in the previous case.

Thus, we have shown that for $\underline{\theta} = 0$, there exists a θ_0 such that $g(W, \theta_0) > g^*(W)$ for $W \in (0, \hat{W})$ and $g(W, \theta_0) < g^*(W)$ for $W > \hat{W}$ for some value $\hat{W}$. The result in the proposition statement then follows since $g^*(W)$ is uniformly continuous in $\underline{\theta}$, so the result continues to hold when $\underline{\theta} > 0$ is sufficiently small.

A.5 Proof of Proposition 3

Suppose that the firm has long term debt D, a credit line C_L with constant interest γ, and equity. Suppose that the equity share of the agent is $\Psi = \bar{\psi}\frac{1}{\sigma}$, the amount of long term debt

[64] The latter interval is empty if $\hat{W} = W^{0*}$.

equals

$$D = \frac{1}{r}\left(\mu - \frac{h}{\Psi} - \gamma C_L\right)$$

for some $h \in \left(\underline{\theta}, \bar{\theta}\right)$, and the long term debt is performance sensitive. Whenever $W_t > W^0$ the interest on the debt is

$$r_1 = \frac{\mu - \gamma C_L - \frac{\bar{\theta}}{\Psi}}{D} < r,$$

while when $W_t \leq W^0$ the interest is

$$r_2 = \frac{\mu - \gamma C_L - \frac{\underline{\theta}}{\Psi}}{D} > r.$$

Let p_t be the instantaneous debt repayment rate. We have $p_t = r_1 D$ if $W_t > W^0$ and $p_t = r_2 D$ otherwise. Denote with M_t the draw on the credit line, which follows

$$dM_t = \gamma M_t dt + p_t dt + dDiv_t - dX_t,$$

where $dDiv_t$ are dividend payments, and dX_t is the change in the firm's output. Plugging in the values for p_t yields

$$dM_t = \gamma M_t dt + \left(\mu - \frac{1}{\Psi}\theta^*\left(W_t\right) - \gamma C_L\right) dt + dDiv_t - dX_t,$$

where $\theta^*(W_t)$ is given by Equation (9). Using guess and verify, the agent's continuation value W_t in Equation (7) solves

$$\gamma W_t = \max_{a\in\{0,1\}} -\Psi\left(\gamma M_t + \mu - \frac{1}{\Psi}\theta^*\left(W_t\right) - \gamma C_L + \frac{dDiv_t}{dt} - \mu a\right) + \frac{dc_t}{dt} - \theta^*\left(W_t\right)a.$$

Setting $dc_t = \Psi dDiv_t$ implies that the optimal contract is implemented and $W_t = \Psi\left(C_L - M_t\right)$, where $C_L = \frac{\bar{W}}{\Psi}$ is the credit limit.

A.6 Proof of Proposition 4

We now solve the agent's problem when the naive principal offers the misspecified contract. We prove Proposition 4 by a standard approach. First, Lemmas 9 and 10 show that the agent's HJB equation (13) has a unique solution with the properties stated in Proposition 4. Then, in Lemma 11, we verify that the solution is indeed the agent's value function. The arguments in this section are similar but simpler than those in Appendix A.3, because the boundary $\bar{W}$ is exogenous in the agent's problem.

Lemma 9. *The agent's HJB equation* (13) *with the boundary conditions* $V(0) = 0$ *and* $V'(\bar{W}) = 1$ *has a solution which solves the equation for all points in* $\left[0, \bar{W}\right] \setminus \{W^0\}$*, is* C^1 *everywhere on* $\left[0, \bar{W}\right]$ *and* C^2 *on* $\left[0, \bar{W}\right] \setminus \{W^0\}$*.*

Proof. Equation (13) is equivalent to

$$V''(W) + \frac{2}{\bar{\psi}^2} H(W, u, p) = 0,$$

where

$$H(W, u, p) = \max_{a \in \{0,1\}} -\gamma u - \theta_0 a + p \cdot \left(\gamma W + \left(\theta^*(W) - \bar{\theta}\right) + \bar{\theta} a\right).$$

The function H is Lipschitz continuous in (u, p) for all $W \in \left[0, \bar{W}\right]$. For two appropriately chosen constant functions $\bar{V}(W) = \bar{V} < 0$ and $\underline{V}(W) = \underline{V} > 0$, the inequalities

$$\bar{V}''(W) + H\left(W, \bar{V}(W), \bar{V}'(W)\right) \geq 0$$

and

$$\underline{V}''(W) + H\left(W, \underline{V}(W), \underline{V}'(W)\right) \leq 0$$

can be verified to hold. Moreover, $H(W, u, p)$ satisfies

$$|H(W, u, p)| \leq M(1 + |p|)$$

for all $u \in \left[\underline{V}, \bar{V}\right]$ and $W \in \left[0, \bar{W}\right]$. Then, a variant of Thompson (1996), Theorem 1 establishes the existence of a solution which solves the equation almost everywhere, and has absolutely continuous first derivative. Since $\theta^*(W)$ is constant on $[0, W^0)$ and $(W^0, \bar{W}]$, $H(W, u, p)$ is continuous on these regions, and it can be shown that $V(W)$ is twice continuously differentiable on either region.[65] Thus, $V(W)$ satisfies (13) at all points except W^0. □

Lemma 10. *The solution to Equation* (13) *with boundary conditions* $V(0) = 0$ *and* $V'\left(\bar{W}\right) = 1$ *is unique.*

Proof. Consider two solutions U, V on $\left[0, \bar{W}\right]$, and define $Z(W) = U(W) - V(W)$, which

[65] Precisely, we can construct a separate boundary value problem for each of the regions, and by Strulovici and Szydlowski (2015), each has a twice continuously differentiable solution on its respective domain.

satisfies

$$\begin{aligned}
\gamma Z(W) &= \max_{a\in\{0,1\}}\left(-\theta_0+U'(W)\bar{\theta}\right)a-\max_{a\in\{0,1\}}\left(-\theta_0+V'(W)\bar{\theta}\right)a \\
&\quad +\left(U'(W)-V'(W)\right)\left(\gamma W+\theta^*(W)-\bar{\theta}\right)+\frac{1}{2}\bar{\psi}^2\left(U''(W)-V''(W)\right) \\
&= \max_{a\in\{0,1\}}\left(-\theta_0+U'(W)\bar{\theta}\right)a-\max_{a\in\{0,1\}}\left(-\theta_0+V'(W)\bar{\theta}\right)a \\
&\quad +Z'(W)\left(\gamma W+\theta^*(W)-\bar{\theta}\right)+\frac{1}{2}\bar{\psi}^2 Z''(W)
\end{aligned}$$

on $[0,\bar{W}]\backslash\{W^0\}$. The boundary conditions $U(0)=V(0)=0$ and $U'(\bar{W})=V'(\bar{W})=1$ imply that $Z(0)=Z'(\bar{W})=0$. The equation above implies that on $[0,W^0)$, $Z(W)$ can neither have a strictly positive local maximum nor a strictly negative local minimum. Specifically, suppose that $Z(W)>0$ is a local maximum for some $W\in(0,W^0)$. Then, $Z'(W)=0$, which implies that $U'(W)=V'(W)$ and that

$$\max_{a\in\{0,1\}}\left(-\theta_0+U'(W)\bar{\theta}\right)a=\max_{a\in\{0,1\}}\left(-\theta_0+V'(W)\bar{\theta}\right)a.$$

But then, we have

$$\gamma Z(W)=\frac{1}{2}\bar{\psi}^2 Z''(W)\leq 0,$$

since $Z''(W)\leq 0$ whenever W is a maximum, which is a contradiction to $Z(W)>0$. The proof that $Z(W)$ cannot have a strictly negative local minimum is analogous. The same argument establishes that $Z(W)$ cannot have a strictly positive local maximum or a strictly negative local minimum on $(W^0,\bar{W}]$. Thus, since $Z(0)=0$, either (1) $Z(W)=Z'(W)=0$ for all $W\in(0,W^0)$, (2) $Z(W)>0$ and $Z'(W)>0$ for all $W\in(0,W^0)$ or (3) $Z(W)<0$ and $Z'(W)<0$ for all $W\in(0,W^0)$. If $Z(W)<0$ for all $W\in(0,W^0)$, then by continuity of $Z(.)$ and $Z'(.)$, it holds that $Z(W^0)<0$ and $Z'(W^0)<0$. But then, since $Z(W)$ cannot have a strictly negative minimum, we must have $Z'(W)<0$ for all $W\in(W^0,\bar{W}]$. In particular, this implies that $Z'(\bar{W})<0$, which is a contradiction to the boundary condition $Z'(\bar{W})=0$. The case when $Z(W)>0$ for $W\in(0,W^0)$ is analogous. Thus, we must have $Z(W)=0$ for $W\in(0,W^0)$, in which case continuity guarantees that $Z(W^0)=Z'(W^0)=0$, so that $Z(W)=Z'(W)=0$ for all $W\in(W^0,\bar{W}]$.

Overall, the argument establishes that $Z(W)=0$ is the unique solution to the above equation, so that $U(W)=V(W)$ for all $W\leq\bar{W}$. □

Lemma 11. *The solution to Equation (13) equals the agent's value function.*

Proof. The proof follows from the Feynman-Kac Theorem, which applies because $V(W)$ is continuously differentiable and twice differentiable almost everywhere in W. Take $F(W_t)=$

$e^{-\gamma t}V(W_t)$ and let $a(W_t)$ be the optimal effort choice of the agent in Equation (13). By Itô's Lemma for semimartingales and Equation (12), we have

$$\begin{aligned} dF(W_t) &= e^{-\gamma t}\left(-\gamma V(W_t) + \left(\gamma W_t + \left(\theta^*(W) - \bar{\theta}\right) + \bar{\theta}a(W_t)\right)V'(W_t) + V''(W_t)\frac{1}{2}\bar{\psi}^2\right)dt \\ &\quad + e^{-\gamma t}V'(W_t)\left(-dc_t + \bar{\psi}dB_t^a\right) \\ &= e^{-\gamma t}\left(\theta_0 a(W_t)\,dt + V'(W_t)\left(-dc_t + \bar{\psi}dB_t^a\right)\right). \end{aligned}$$

Since the firing time τ satisfies $\tau = \inf\{t : W_t = 0\}$ we have for any constant time $T > 0$,

$$\begin{aligned} E^{a,\theta_0}\left[F(W_{\tau\wedge T}) - F(W_0)\right] &= E^{a,\theta_0}\left[e^{-\gamma(\tau\wedge T)}V(W_{\tau\wedge T}) - V(W_0)\right] \\ &= E^{a,\theta_0}\left[\int_0^{\tau\wedge T} e^{-\gamma t}\left(\theta_0 a(W_t)\,dt - V'(W_t)\,dc_t\right)\right]. \end{aligned}$$

Because $dc_t = 0$ whenever $W_t \neq \bar{W}$, we can replace $V'(W_t)\,dc_t$ with $V'(\bar{W})dc_t$, so that

$$V(W_0) = E^{a,\theta_0}\left[\int_0^{\tau\wedge T} e^{-\gamma t}\left(V'(\bar{W})dc_t - \theta_0 a_t dt\right)\right] + E^{a,\theta_0}\left[e^{-\gamma(\tau\wedge T)}V(W_{\tau\wedge T})\right]. \qquad (38)$$

The transversality condition

$$\lim_{T\to\infty} E^{a,\theta_0}\left[e^{-\gamma T}V(W_T)\right] = 0$$

holds because $V(W)$ is bounded. Applying the conditions $W_\tau = 0$, $V(0) = 0$ and $V'(\bar{W}) = 1$ to Equation (38) and letting T go to infinity then yields the result. □

Before we proceed, we record a technical result which will be used later.

Lemma 12. *We have $V(W) > 0$ for all $W \in (0, \bar{W}]$ and $V'(W) > 0$ for all $W \in \left[0, \bar{W}\right]$.*

Proof. The agent can guarantee himself a strictly positive payoff for all $W > 0$ by always shirking, which implies that $V(W) > 0$ for all $W \in (0, \bar{W}]$. In turn, this implies that $V'(0) \geq 0$. If $V'(0) = 0$, then the agent shirks at $W = 0$ the agent's HJB equation implies that (13) $V''(0) = 0$. Similarly, all higher order derivatives are zero and the agent's value is a constant function, i.e. $V(W) = 0$ for all $W \leq \bar{W}$, a contradiction. Thus, it must be the case that $V'(0) > 0$. To show that $V'(W) > 0$ for all $W \in [0, \bar{W}]$, suppose by way of contradiction that $\hat{W} = \inf\{W : V'(W) \leq 0\}$ exists. Then, $V'(\hat{W}) = 0$ by continuity of $V'(W)$ and since $V'(\hat{W}) = 0 < \frac{\theta_0}{\bar{\theta}}$, the agent optimally shirks at $\hat{W}$ (see Equation (13)).

Thus, the HJB equation (13) implies that

$$V''\left(\hat{W}\right)\frac{1}{2}\bar{\psi}^2=\gamma V\left(\hat{W}\right)>0.$$

This is a contradiction. By construction of $\hat{W}$, we have $V'(W)>0$ for $W\in(\hat{W}-\varepsilon,\hat{W})$. But since $V''(\hat{W})>0$, the function $V'(W)$ cannot cross zero from above at $\hat{W}$. Thus, $V'(W)>0$ for all $W\in[0,\bar{W}]$. □

We next study the properties of the value function and the agent's optimal effort. Lemma 13 contains a technical result. Lemmas 14 and 16 show that either the agent works for all W or that there exists a cutoff $W_s\in(0,W^0]$ such that the agent shirks for $W<W_s$ and works for $W>W_s$. Lemma 18 shows that for θ_0 sufficiently small, the agent works for all W and that for θ_0 sufficiently large, he shirks for $W<W_s$.

In the remainder of the section, we work with the scaled value function

$$\tilde{V}(W)=V(W)-\frac{\theta_0}{\bar{\theta}}W \tag{39}$$

to simplify notation. It is optimal for the agent to exert effort whenever $\tilde{V}'(W)>0$, so that it is more convenient to study $\tilde{V}$ instead of V. By Lemma 9, $\tilde{V}(W)$ is continuously differentiable everywhere on $[0,\bar{W}]$ and twice continuously differentiable on $[0,\bar{W}]\setminus\{W^0\}$.

Lemma 13. *For any $n\in\mathbb{R}$ denote*

$$I_n(W)=\int_0^W\frac{\exp\left(-\frac{1}{\frac{1}{2}\bar{\psi}^2}\int_0^u(\gamma y+n)\,dy\right)}{\left(u+\frac{n}{\gamma}\right)^2}du,$$

which is finite on any interval not containing $W=-\frac{n}{\gamma}$. On any interval $[W_1,W_2]\subset[0,W^0]$ on which the agent shirks, the solution to Equation (13) satisfies[66]

$$V_{ls}(W)=\left(W-\frac{\bar{\theta}-\underline{\theta}}{\gamma}\right)\left(C_{ls1}+C_{ls2}I_{\underline{\theta}-\bar{\theta}}(W)\right)+\frac{\theta_0}{\bar{\theta}}W-\frac{\theta_0}{\gamma}\frac{\bar{\theta}-\underline{\theta}}{\bar{\theta}},$$

and $C_{ls2}=0$ whenever $\frac{\bar{\theta}-\underline{\theta}}{\gamma}\in[W_1,W_2]$. If the agent works on $[W_1,W_2]\subset[0,W^0]$, the solution satisfies

$$V_{lw}=\left(W+\frac{\underline{\theta}}{\gamma}\right)\left(C_{lw1}+C_{lw2}I_{\underline{\theta}}(W)\right)+\frac{\theta_0}{\bar{\theta}}W-\frac{\theta_0}{\gamma}\frac{\bar{\theta}-\underline{\theta}}{\bar{\theta}}.$$

[66] The constants C_{ls1} and C_{ls2} denote shirking when θ_t is low. Similarly, C_{lw1} and C_{lw2} denote working when θ_t is low, and C_{hs1} and C_{hs2} denote shirking when θ_t is high.

On any interval $[W_1, W_2] \subset \left[W^0, \bar{W}\right]$ *on which the agent shirks we have*

$$V_{hs}(W) = W\left(C_{hs1} + C_{hs2} I_0(W)\right) + \frac{\theta_0}{\bar{\theta}} W,$$

and if the agent works we have

$$V_{hw}(W) = \left(W + \frac{\bar{\theta}}{\gamma}\right)\left(C_{hw1} + C_{hw2} I_{\bar{\theta}}(W)\right) + \frac{\theta_0}{\bar{\theta}} W.$$

Proof. We only provide a proof for the first case; the proofs for the other cases are analogous. Consider the transformation $\tilde{V}(W) = V(W) - \frac{\theta_0}{\bar{\theta}} W$. When $W < W^0$ (i.e., $\theta^*(W) = \underline{\theta}$) and the agent shirks, $V(W)$ solves the equation

$$\gamma V(W) = V'(W)\left(\gamma W - \left(\bar{\theta} - \underline{\theta}\right)\right) + V''(W)\frac{1}{2}\bar{\psi}^2,$$

and since $\tilde{V}'(W) = V'(W) - \frac{\theta_0}{\bar{\theta}}$ and $\tilde{V}''(W) = V''(W)$ by construction, $\tilde{V}(W)$ solves

$$\gamma \tilde{V}(W) = \tilde{V}'(W)\left(\gamma W - \left(\bar{\theta} - \underline{\theta}\right)\right) - \frac{\theta_0}{\bar{\theta}}\left(\bar{\theta} - \underline{\theta}\right) + \tilde{V}''(W)\frac{1}{2}\bar{\psi}^2.$$

On any interval $[W_1, W_2]$ with $W_2 < W^0$ a particular solution to the homogeneous equation

$$\gamma \tilde{V}(W) = \tilde{V}'(W)\left(\gamma W - \left(\bar{\theta} - \underline{\theta}\right)\right) + \tilde{V}''(W)\frac{1}{2}\bar{\psi}^2$$

is $\tilde{V}(W) = W - \frac{\bar{\theta} - \underline{\theta}}{\gamma}$, and any general solution is of the form

$$\tilde{V}(W) = \left(W - \frac{\bar{\theta} - \underline{\theta}}{\gamma}\right)\left(C_{ls1} + C_{ls2} I_{\underline{\theta} - \bar{\theta}}(W)\right) - \frac{\theta_0}{\gamma}\frac{\bar{\theta} - \underline{\theta}}{\bar{\theta}}$$

for some constants $C_{ls1}, C_{ls2} \in \mathbb{R}$, which are determined from the boundary conditions at W_1 and W_2.[67] Since the agent's value function is finite on $\left[0, \bar{W}\right]$, $C_{ls2} = 0$ if $W_2 > \frac{\bar{\theta} - \underline{\theta}}{\gamma}$. □

Corollary 1. *On any interval* $[W_1, W_2] \subset \left[0, \bar{W}\right]$ *with* $W^0 \notin [W_1, W_2]$ *on which the agent's effort is constant, i.e. he either works for all* $W \in [W_1, W_2]$ *or shirks for all* $W \in [W_1, W_2]$, $V(W)$ *is either strictly concave, strictly convex, or linear.*

Proof. Differentiate $\tilde{V}(W)$ twice to get for any $n \in \left\{-\left(\bar{\theta} - \underline{\theta}\right), 0, \underline{\theta}, \bar{\theta}\right\}$ and $W \neq -\frac{n}{\gamma}$

$$V''(W) \;=\; -C_2 \frac{\gamma}{\frac{1}{2}\bar{\psi}^2} \exp\left(-\frac{1}{\frac{1}{2}\bar{\psi}^2}\int_0^W (\gamma y + n)\, dy\right).$$

[67] See for example Polyanin and Zaitsev (2002).

Since $\tilde{V}''(W) = V''(W)$ by construction, if $C_2 > 0$, $V(W)$ is concave, if $C_2 < 0$, $V(W)$ is convex, and if $C_2 = 0$ it is linear. □

Using the explicit characterization in Lemma 13 and Corollary 1, we characterize the agent's effort policy on $\left[0, \bar{W}\right]$.

Lemma 14. *If the agent works at W^0, he works for all $W \in [W^0, \bar{W}]$.*

Proof. If the agent works at W^0, one of the following cases must hold: (1) $\tilde{V}'(W^0) > 0$, (2) $\tilde{V}'(W^0) = 0$ and $\tilde{V}''_+(W^0) > 0$, (3) $\tilde{V}'(W^0) = 0$ and $\tilde{V}''_+(W^0) < 0$, and (4) $\tilde{V}'(W^0) = \tilde{V}''_+(W^0) = 0$. If $\tilde{V}'(W^0) > 0$, then by continuity of $\tilde{V}'(W)$ there exists an interval of positive length $(W^0, \widetilde{W}]$ on which the agent works. The same holds when $\tilde{V}'(W^0) = 0$ and $\tilde{V}''_+(W^0) > 0$. If $\widetilde{W} = \bar{W}$, we have established the result. By way of contradiction, suppose that $\widetilde{W} < \bar{W}$. Then, we can define wlog $\widetilde{W} = \inf\left\{W \in [W^0, \bar{W}] : \tilde{V}'(W) \leq 0\right\}$. By continuity of $\tilde{V}'(W)$, it holds that $\tilde{V}'(\widetilde{W}) = 0$. Since $\tilde{V}'(\widetilde{W}) = 0 < \tilde{V}'(W^0)$, Corollary 1 implies that $\tilde{V}''(W) < 0$ on $(W^0, \widetilde{W}]$, i.e. $\tilde{V}(W)$ is strictly concave on the entire interval $(W^0, \widetilde{W}]$. Moreover, $\tilde{V}''(\widetilde{W} + \varepsilon) < 0$ for some $\varepsilon > 0$, since by Lemma 9 $\tilde{V}''(W)$ is continuous on $(W^0, \bar{W}]$. This implies that $\tilde{V}'(\widetilde{W} + \varepsilon) < 0$, so the agent shirks on some interval $(\widetilde{W}, \widetilde{W} + \varepsilon)$. On this interval, $\tilde{V}(W)$ is concave, again by Corollary 1. Therefore, for all $W > \widetilde{W}$ the agent shirks and $\tilde{V}'(W) < 0$ for all $W \in [\widetilde{W}, \bar{W}]$. In particular, this implies that $\tilde{V}'(\bar{W}) < 0$. However, by Equation (39) the boundary condition $V'(\bar{W}) = 1$ is equivalent to $\tilde{V}'(\bar{W}) > 0$ and we have a contradiction. If $\tilde{V}'(W^0) = 0$ and $\tilde{V}''(W^0) < 0$, then the previous argument establishes that the agent shirks for all $W > W^0$, which again is a contradiction to the boundary condition at $\bar{W}$. Finally, if $\tilde{V}'(W^0) = \tilde{V}''_+(W^0) = 0$, then Corollary 1 establishes that the agent's value is linear on $[W^0, \bar{W}]$, so the agent works on that interval. □

Lemma 15. *If the agent works at $W = 0$, he works for all $W \in \left[0, \bar{W}\right]$.*

Proof. Suppose that the agent works at $W = 0$, which implies that $\tilde{V}'(0) \geq 0$. We distinguish the cases $\tilde{V}'(0) = 0$ and $\tilde{V}'(0) > 0$. If $\tilde{V}'(0) > 0$, then since $\tilde{V}'(W)$ is continuous, there exists an interval of positive length on which the agent works. On this interval, $\tilde{V}(W)$ takes the form

$$\tilde{V}(W) = \left(W + \frac{\underline{\theta}}{\gamma}\right)\left(C_{lw1} + C_{lw2} I_{\underline{\theta}}(W)\right) - \frac{\theta_0}{\gamma}\frac{\bar{\theta} - \underline{\theta}}{\bar{\theta}}$$

by Lemma 13. The boundary condition at zero is $\tilde{V}(0) = 0$ which implies $C_{lw1} = \frac{\theta_0}{\underline{\theta}}\frac{\bar{\theta} - \underline{\theta}}{\bar{\theta}} > 0$. If $C_{lw2} < 0$, then by Corollary 1 $\tilde{V}(W)$ is convex and $\tilde{V}'(W) > \tilde{V}'(0) > 0$ for all $W \in [0, W^0]$. Thus, the agent works on $[0, W^0]$. Since $\tilde{V}'(W^0) > 0$ by continuity of $\tilde{V}'(W)$ (see Lemma

9), Lemma 14 implies that the agent also works for all $W \in [W^0, \bar{W}]$. If $C_{lw2} > 0$, then

$$\tilde{V}'(W) = C_{lw1} + C_{lw2} I_{\underline{\theta}}(W) + C_{lw2} \frac{\exp\left(-\frac{1}{\frac{1}{2}\bar{\psi}^2}\int_0^W (\gamma y + \underline{\theta})\, dy\right)}{\left(W + \frac{\underline{\theta}}{\gamma}\right)} > 0$$

for all $W \in [0, W^0]$, and the agent works for all $W \leq W^0$. Since the agent works at W^0, again by Lemma 14, the agent also works for all $W \in [W^0, \bar{W}]$.

Finally, if $\tilde{V}'(0) = 0$ and the agent works at $W = 0$, then the agent's HJB equation (13) and the boundary condition $\tilde{V}(0) = 0$ imply that $V''(0) = \tilde{V}''(0) > 0$, so that $\tilde{V}'(W) > 0$ for some interval of positive length. The remainder of the proof is then the same as in the previous case. □

Lemma 16. *If the agent shirks at $W = 0$, he shirks for all $W \in [0, W_s)$ and works for all $W \in [W_s, \bar{W}]$, where $W_s = \inf\left\{W \in [0, \bar{W}] : a(W) = 1\right\}$ and $W_s \leq W^0$. The agent's value function is convex on $[0, W_s]$.*

Proof. If the agent shirks at $W = 0$, then $\tilde{V}'(0) \leq 0$. We distinguish the cases $\tilde{V}'(0) < 0$ and $\tilde{V}'(0) = 0$. First, consider the case when $\tilde{V}'(0) < 0$. If the agent shirks for all $W \in [0, \bar{W}]$, then we have $\tilde{V}'(\bar{W}) \leq 0$, which is inconsistent with the boundary condition $V'(\bar{W}) = 1$, since the boundary condition implies that $\tilde{V}'(\bar{W}) = 1 - \frac{\theta_0}{\theta} > 0$ by Equation (39). Thus, the point $W_s = \inf\left\{W \in [0, \bar{W}] : a(W) = 1\right\}$ exists and $\tilde{V}'(W_s) = 0$ by continuity of $\tilde{V}'(W)$.

Suppose that $W_s \leq W^0$. By Corollary 1, this implies that $\tilde{V}''(W) > 0$ for all $W \in [0, W_s]$, since $\tilde{V}'(0) < 0 = \tilde{V}'(W_s)$. If $W_s < W^0$, then for $W \in [W_s, W^0)$, Corollary 1 implies that $\tilde{V}''(W) > 0$, since $\tilde{V}''(W_s) > 0$. Thus, $\tilde{V}'(W^0) > \tilde{V}'(W_s) = 0$, so that the agent works at W^0. Then, Lemma 14 implies that the agent works for all $W \in [W^0, \bar{W}]$.

If $W_s = W^0$, we must distinguish three cases: (1) $\tilde{V}_+''(W^0) > 0$, (2) $\tilde{V}_+''(W^0) < 0$ and (3) $\tilde{V}_+''(W^0) = 0$. If $V_+'''(W^0) > 0$, then the agent works on some interval $[W^0, W^0 + \varepsilon]$. By Lemma 14, the agent works for all $W \in [W^0, \bar{W}]$. If $\tilde{V}_+''(W^0) < 0$, the agent shirks on some interval $(W^0, W^0 + \varepsilon)$. Then, Corollary 1 implies that $\tilde{V}''(W) < 0$ on that interval, which implies that the agent shirks for all $W \geq W^0$, which again is a contradiction to the boundary condition at $\bar{W}$. Finally, when $\tilde{V}_+''(W^0) = 0$, then Corollary 1 implies that the agent's value function is linear on $[W^0, \bar{W}]$ and we have $\tilde{V}(W) = \tilde{V}'(W^0) = 0$ for all $W \in [W^0, \bar{W}]$. This is another contradiction, since the boundary condition at $\bar{W}$. Thus, the only possible case is that the agent works for all $W \in [W^0, \bar{W}]$ if $W_s = W^0$.

Next, suppose that $W_s \in (W^0, \bar{W})$. We will show that this results in a contradiction. In that case, $\tilde{V}'(W^0) < 0$ since $W^0 < W_s$, and the agent shirks on (W^0, W_s). At W_s, the agent's HJB equation (13) yields

$$\gamma\tilde{V}\left(W_s\right)=\tilde{V}'\left(W_s\right)\gamma W_s+\tilde{V}''\left(W_s\right)\frac{1}{2}\bar{\psi}^2$$

or equivalently

$$\gamma\tilde{V}\left(W_s\right)=\tilde{V}''\left(W_s\right)\frac{1}{2}\bar{\psi}^2,$$

since $\tilde{V}'(W_s)=0$. By construction of W_s, $\tilde{V}'\left(W\right)<0$ for all $W\in(0,W_s)$. Together with $\tilde{V}\left(0\right)=0$, this implies that $\tilde{V}(W_s)<0$. Then, the equation above implies that $\tilde{V}''(W_s)<0$. This is a contradiction, since when $\tilde{V}''(W_s)<0$, the function $\tilde{V}'\left(W\right)$ cannot cross zero from below at W_s. Thus, it must be the case that $W_s\leq W^0$.

Next, consider the case $\tilde{V}'\left(0\right)=0$. If the agent shirks at $W=0$, then the agent's HJB equation (13) at $W=0$ implies that

$$\gamma\tilde{V}\left(0\right)=-\frac{\theta_0}{\bar{\theta}}\left(\bar{\theta}-\underline{\theta}\right)-\tilde{V}'\left(0\right)\left(\bar{\theta}-\underline{\theta}\right)+\tilde{V}''\left(0\right)\frac{1}{2}\bar{\psi}^2.$$

Plugging in $\tilde{V}\left(0\right)=\tilde{V}'\left(0\right)=0$ then yields $\tilde{V}''\left(0\right)>0$. Then, a variant of Lemma 14 implies that the agent works for all $W>0$ and we have $W_s=0$.

Overall, we have shown that $W_s\leq W^0$ exists and that the agent's value is convex on $[0,W_s]$. □

Lemma 17. *The agent's value function is concave on* $\left[W^0,\bar{W}\right]$.

Proof. By the preceding Lemma, the agent works for all $W\in[W_s,\bar{W}]$, where $W_s\leq W^0$. For $W\in(W_s,\bar{W})$, we have $V'\left(W\right)\in(\theta_0/\bar{\theta},1)$ by construction. Suppose by way of contradiction that $V''_+(W^0)>0$. Then, by Corollary 1, $V''\left(W\right)>0$ on $(W^0,\bar{W})$, since $V''\left(W\right)=\tilde{V}''\left(W\right)$ for all $W\in[0,\bar{W}]\setminus\{W^0\}$. Moreover, the agent's HJB equation (13) implies that

$$\gamma V\left(W^0\right)=-\theta_0+V'\left(W^0\right)\left(\gamma W^0+\bar{\theta}\right)+V''_+\left(W^0\right)\frac{1}{2}\bar{\psi}^2$$

and

$$\gamma V\left(W^0\right)=-\theta_0+V'\left(W^0\right)\left(\gamma W^0+\underline{\theta}\right)+V''_-\left(W^0\right)\frac{1}{2}\bar{\psi}^2,$$

which together imply that

$$\frac{1}{2}\bar{\psi}^2\left(V''_+\left(W^0\right)-V''_-\left(W^0\right)\right)=-V'\left(W^0\right)\left(\bar{\theta}-\underline{\theta}\right)<0,$$

since $V'(W^0)>0$ by Lemma 12. Thus, we have $0<V''_+\left(W^0\right)<V''_-\left(W^0\right)$. Corollary 1 then implies that $V''\left(W\right)>0$ for all $W\in(W_s,W^0)$. By Lemma 16, $V''\left(W\right)>0$ for $W\in[0,W_s]$.

Thus, overall, we have $V''(W) > 0$ for all $W \in [0, \bar{W}] \setminus \{W^0\}$.

On $(W^0, \bar{W})$, the agent's HJB equation (13) implies that

$$\gamma V(W) = -\theta_0 + V'(W)\left(\gamma W + \bar{\theta}\right) + V''(W)\frac{1}{2}\bar{\psi}^2$$

or equivalently

$$V''(W)\frac{1}{2}\bar{\psi}^2 = \gamma\left(V(W) - V'(W)W\right) + \left(\theta_0 - V'(W)\bar{\theta}\right).$$

Since $V'(W) > \theta_0/\bar{\theta}$ on $(W^0, \bar{W})$, the second term is negative. The first term is negative as well. To see this, define

$$Z(W) = V(W) - V'(W)W$$

and note that $Z(0) = 0$ and $Z'(W) = -V''(W)W$. Above, we have shown that $V''(W) > 0$ for all $W \in [0, \bar{W}]$ and thus $Z'(W) < 0$. Since $Z(0) = 0$, this implies that $Z(W) < 0$. Thus, we have $V''(W) < 0$ on $(W^0, \bar{W})$, which is a contradiction.

Hence, it must be the case that $V''_+(W^0) \leq 0$. Then, Corollary 1 implies that $V''(W) \leq 0$ for all $W \in [W^0, \bar{W}]$, which is what we set out to prove. □

We thus have established that there are only two possible regimes. Either the agent works for all $W \in [0, \bar{W}]$, or he shirks on $[0, W_s)$ and works on $[W_s, \bar{W}]$. Which of the regimes arises depends solely on the initial slope of $\tilde{V}(W)$, which in turn depends on the effort cost θ_0.

Lemma 18. *There exists a cutoff $\hat{\theta}_0$ such that for $\theta_0 > \hat{\theta}_0$ the agent shirks for $W < W_s$ and for $\theta_0 < \hat{\theta}_0$ the agent works for all W.*

Proof. Let $V_{\theta_0}(W)$ be the value function given effort cost parameter θ_0 and $V_{\theta_0}(W, a)$ the value function given a certain effort policy $a \equiv \{a_t\}_{t\geq 0}$ with $a_t = 1$ on a set of time with strictly positive measure. We have for any $\theta_0' > \theta_0$

$$V_{\theta_0'}(W, a) \leq V_{\theta_0}(W, a)$$

for any such a, and the inequality is strict for $W > 0$. Let

$$a_{\theta_0} \in \arg\sup_a V_{\theta_0}(W, a).$$

Then,

$$V_{\theta_0'}(W) = V_{\theta_0'}\left(W, a_{\theta_0'}\right) < V_{\theta_0'}\left(W, a_{\theta_0'}\right) \leq V_{\theta_0}\left(W, a_{\theta_0}\right) = V_{\theta_0}(W)$$

for $W > 0$. Thus, the value function $V_{\theta_0}(W)$ is decreasing in θ_0, uniformly for all $W > 0$.

Since $V_{\theta_0}(0) = V_{\theta_0'}(0) = 0$ for all $\theta_0', \theta_0 > 0$ by the boundary condition at zero, we must have $V_{\theta_0}'(0) > V_{\theta_0'}'(0)$ for $\theta_0' > \theta_0$.

Consider the case $\theta_0 = \underline{\theta}$. We have $V_{\underline{\theta}}(W) > W$ for $W > 0$ and hence $V'(0) > 1 > \frac{\underline{\theta}}{\theta}$. By Lemma 15, the only solution for an agent with $\theta_0 = \underline{\theta}$ is to work at all W. Similarly, for $\theta_0 = \bar{\theta}$, we must have $V_{\bar{\theta}}(W) < W$ for $W > 0$. For this agent, $V'(0) < 1 = \frac{\theta_0}{\theta}$ and hence the agent shirks at $W = 0$. By Lemma 16, the agent works for all $W \geq W_s$.

Given the existence of types which work at all W and shirk at $W = 0$, the result follows from continuity and monotonicity of $V_{\theta_0}'(0)$ in θ_0. □

A.7 Proof of Proposition 5

Since the agent exerts effort whenever $\psi_t \geq \psi \equiv \frac{\sigma}{\mu}\kappa$, the principal's HJB equation for any incentive compatible contract is given by

$$
\begin{aligned}
rJ(W) &= \max_{\psi' \geq \psi} \min_{\theta \in [\underline{\theta}, \bar{\theta}]} \mu + \theta + J'(W)(\gamma W + \kappa) + J''(W)\frac{1}{2}\psi'^2 \\
&= \max_{\psi' \geq \psi} \mu + \underline{\theta} + J'(W)(\gamma W + \kappa) + J''(W)\frac{1}{2}\psi'^2
\end{aligned}
$$

with boundary conditions $J(0) = L$, $J'(\bar{W}) = -1$ and $J''(\bar{W}) = 0$. This equation has a unique, twice continuously differentiable and strictly concave solution, which follows from a similar argument as in the proof of Proposition 1. Setting ψ to $\bar{\psi}$ is optimal since J is concave. The verification argument for this HJB equation is similar to the one in Proposition 1.

A.8 Proof of Proposition 6

The agent's value satisfies the HJB equation

$$
\gamma V(W) = \max_{a \in \{0,1\}} -\kappa a + V'(W)\left(\gamma W + \kappa a - \frac{\psi}{\sigma}\underline{\theta}\right) + V''(W)\frac{1}{2}\psi^2
$$

on $[0, \bar{W}]$ with boundary conditions $V(0) = 0$ and $V(\bar{W}) = \bar{W}$. This equation has a unique twice continuously differentiable solution. Guessing $V(W) = W + C$ for some constant C yields

$$
\gamma W + \gamma C = \max_{a \in \{0,1\}} -\kappa a + \gamma W + \kappa a - \frac{\psi}{\sigma}\underline{\theta}.
$$

Thus, $a = 1$ is optimal for all $W \in \left[0, \bar{W}\right]$ and we have

$$C = -\frac{\psi}{\gamma\sigma}\underline{\theta} \; = -\frac{\kappa}{\gamma\mu}\underline{\theta} > 0.$$

This establishes the proposition.

A.9 Proof of Proposition 7

That $\theta_t^A = \bar{\theta}$ for all t is the agent's worst case can be seen directly from Equation (16). For any contract α, we have

$$E^{a,\theta^A}\left[\int_0^\tau e^{-\gamma t}\left(dc_t - a_t d\kappa_t\right)\right] \geq E^{a,\bar{\theta}}\left[\int_0^\tau e^{-\gamma t}\left(dc_t - a_t d\kappa_t\right)\right],$$

where $\bar{\theta} = \left\{\bar{\theta}\right\}_{t\geq 0}$ is the constant process $\bar{\theta}_t = \bar{\theta}$. We can now apply the martingale representation theorem under $P^{a,\bar{\theta}}$ (e.g. Shreve (2004), Corollary 5.3.2, p. 222) to get

$$dW_t = \left(\gamma W_t + \bar{\theta} a_t\right) dt - dc_t + \psi_t dB_t.$$

Exerting effort is incentive compatible whenever

$$\psi_t \geq \bar{\psi} \equiv \frac{\mu}{\sigma}\bar{\theta}.$$

Suppose first that the principal is sophisticated. Then, her problem is given by Equation (18). Her own worst-case θ^P does not affect the agent's (true) value in Equation (16). Thus, her HJB equation is

$$\begin{aligned} rJ(W) &= \max_{\psi_t \geq \bar{\psi}} \min_{\theta^P \in \left[\underline{\theta}, \bar{\theta}\right]} \mu + J'(W)\left(\gamma W + \bar{\theta}\right) + \frac{1}{2}\bar{\psi}^2 J''(W) \\ &= \max_{\psi_t \geq \bar{\psi}} \mu + J'(W)\left(\gamma W + \bar{\theta}\right) + \frac{1}{2}\bar{\psi}^2 J''(W) \end{aligned}$$

with boundary conditions $J(0) = L$, $J'(\bar{W}) = -1$, and $J''(\bar{W}) = 0$. This equation is equivalent to the one in DeMarzo and Sannikov (2006) and their characterization applies. The optimal contract is then identical to theirs.

Now, suppose that the principal is naive, i.e. she offers the contract of Proposition 1 and believes that the agent evaluates this contract under P^{a^*,θ^*}. Since the agent's worst-case is given by $\theta_t^A = \bar{\theta}$, we can redefine $\theta_0 = \bar{\theta}$ in Proposition 4, which then characterizes the agent's optimal effort.

B Additional Results

B.1 Sufficient Condition for Effort

We now provide a sufficient condition so that implementing $a_t = 1$ for all t is optimal.

Proposition 9. *Implementing $a_t = 1$ is optimal as long as for all $W \in [0, \bar{W}]$,*

$$\bar{\psi} \in \arg\max_{\psi \geq 0} \min_{\theta \in \{\underline{\theta}, \bar{\theta}\}} \left(\mu + J'(W)\,\theta\right) a(\psi, \theta) + J''(W)\,\frac{1}{2}\psi^2.$$

This holds whenever

$$J'(0) \leq \frac{\mu}{\bar{\theta} - \underline{\theta}}$$

and

$$\gamma r L \geq (\gamma - r)\,\frac{\mu}{r}.$$

Since $J(W)$ is uniformly bounded below $\frac{\mu}{r}$, the first condition holds if $\bar{\theta} - \underline{\theta}$ is sufficiently small. The second condition guarantees that effort is always optimal,[68] and is satisfied for γ sufficiently close to r.

The proof is similar to the one in DeMarzo and Sannikov (2006), p. 2721f, which establishes an analogous result in their model. It is hence omitted.

B.2 Sophisticated Principal

The agent's value is given by

$$W_0(\alpha) = E^{a,\theta_0}\left[\int_0^\tau e^{-\gamma t}\left(dc_t - a_t d\kappa_t\right)\right].$$

Under measure P^{a,θ_0}, the martingale representation theorem (e.g. Shreve (2004), Corollary 5.3.2, p. 222) implies that

$$dW_t = (\gamma W_t + \theta_0 a_t)\,dt - dc_t + \psi_t dB_t^a \tag{40}$$

and the contract is incentive compatible whenever

$$\psi_t \geq \frac{\sigma}{\mu}\theta_0. \tag{41}$$

[68] See DeMarzo and Sannikov (2006), Proposition 8.

The principal's problem is

$$J_0 = \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \Theta} E^{a,\theta} \left[\int_0^{\tau} e^{-rt} \left(dX_t - dc_t \right) \right]$$

subject to incentive compatibility and participation for the agent, i.e. $W_0(\alpha) \geq 0$ and Equations (40) and (41). Since θ does not affect the agent's problem, it is irrelevant for the principal. Thus, the principal's problem is equivalent to the one in DeMarzo and Sannikov (2006) and their characterization applies.

References

Aliprantis, C. D. and K. Border (2006). *Infinite dimensional analysis: A hitchhiker's guide*. Springer Science & Business Media.

Bailey, P., L. Shampine, and P. Waltman (1968). *Nonlinear two-point boundary value problems*. Academic Press.

Battaglini, M. (2005). Long-term contracting with Markovian consumers. *American Economic Review 95*(3), 637–658.

Bebchuk, L. and J. Fried (2006). *Pay without performance: The unfulfilled promise of executive compensation*. Harvard Univ Press.

Bergemann, D. and K. Schlag (2008). Pricing without priors. *Journal of the European Economic Association 6*(2-3), 560–569.

Bergemann, D. and K. Schlag (2011). Robust monopoly pricing. *Journal of Economic Theory 146*(6), 2527–2543.

Berger, P. G., E. Ofek, and D. L. Yermack (1997). Managerial entrenchment and capital structure decisions. *Journal of Finance 52*(4), 1411–1438.

Bertrand, M. and S. Mullainathan (2003). Enjoying the quiet life? Corporate governance and managerial preferences. *Journal of Political Economy 111*(5), 1043–1075.

Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica 78*(1), 73–118.

Chen, Z. and L. Epstein (2002). Ambiguity, risk, and asset returns in continuous time. *Econometrica 70*(4), 1403–1443.

Cvitanic, J. and J. Zhang (2012). *Contract theory in continuous-time models*. Springer Science & Business Media.

DeMarzo, P., M. Fishman, Z. He, and N. Wang (2012). Dynamic Agency and the q Theory of Investment. *Journal of Finance 67*(6), 2295–2340.

DeMarzo, P. and Y. Sannikov (2006). Optimal security design and dynamic capital structure in a continuous-time agency model. *Journal of Finance 61*(6), 2681–2724.

Di Tillio, A., N. Kos, and M. Messner (2016). The design of ambiguous mechanisms. *The Review of Economic Studies 84*(1), 237–276.

Dicks, D. L. and P. Fulghieri (2018). Uncertainty and Contracting in Organizations. *Kenan Institute of Private Enterprise Research Paper* (19-1).

Dumav, M. (2017). Continuous-Time Contracting with Ambiguous Perceptions. *Working Paper*.

Elliott, R. J. (1977). The optimal control of a stochastic system. *SIAM Journal on Control and Optimization 15*(5), 756–778. Publisher: SIAM.

Farber, H. S. and R. Gibbons (1996). Learning and wage dynamics. *Quarterly Journal of Economics 111*(4), 1007–1047.

Flabbi, L. and A. Ichino (2001). Productivity, seniority and wages: new evidence from personnel data. *Labour Economics 8*(3), 359–387.

Fleming, W. H. and H. M. Soner (2006). *Controlled Markov Processes and Viscosity Solutions*, Volume 25. Springer Science & Business Media.

Fleming, W. H. and P. E. Souganidis (1989). On the existence of value functions of two-player, zero-sum stochastic differential games. *Indiana Univ. Math. J 38*(2), 293–314.

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

Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics 123*(1), 49–100.

Garrett, D. and A. Pavan (2012). Managerial turnover in a changing world. *Journal of Political Economy 120*(5), 879–925.

Garrett, D. F. (2014). Robustness of simple menus of contracts in cost-based procurement. *Games and Economic Behavior 87*, 631–641.

Garrett, D. F. and A. Pavan (2015). Dynamic managerial compensation: A variational approach. *Journal of Economic Theory 159*, 775–818. Publisher: Elsevier.

Giat, Y., S. Hackman, and A. Subramanian (2010). Investment under Uncertainty, Heterogeneous Beliefs, and Agency Conflicts. *Review of Financial Studies 23*(4), 1360–1404.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Hansen, L. P. and T. J. Sargent (2012). Three types of ambiguity. *Journal of Monetary Economics 59*(5), 422–445. Publisher: Elsevier.

Hansen, L. P., T. J. Sargent, G. Turmuhambetova, and N. Williams (2006). Robust control and model misspecification. *Journal of Economic Theory 128*(1), 45–90.

Hartman, P. (2002). *Ordinary Differential Equations*. Classics in Applied Mathematics. SIAM.

He, Z. (2009). Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion. *Review of Financial Studies 22*(2), 859–892.

He, Z., B. Wei, J. Yu, and F. Gao (2017). Optimal long-term contracting with learning. *Review of Financial Studies 30*(6), 2006–2065. Publisher: Oxford University Press.

Hermalin, B. E. and M. S. Weisbach (1988). The determinants of board composition. *RAND Journal of Economics 19*(4), 589–606.

Holmström, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica 55*(2), 303–328.

Jeanblanc, M., M. Yor, and M. Chesney (2009). *Mathematical methods for financial markets*. Springer Science & Business Media.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review 76*(2), 323–329.

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*(4), 305–360.

Karatzas, I. and S. E. Shreve (1991). *Brownian motion and stochastic calculus*. Springer Verlag.

Laffont, J. J. and J. Tirole (1988). The dynamics of incentive contracts. *Econometrica 56*(5), 1153–1175.

Lazear, E. P. and P. Oyer (2012). Personnel Economics. *The Handbook of Organizational Economics*, 479 – 519.

Manso, G., B. Strulovici, and A. Tchistyi (2010). Performance-sensitive debt. *Review of Financial Studies 23*(5), 1819–1854.

Marinovic, I. and F. Varas (2019). CEO Horizon, Optimal Pay Duration, and the Escalation of Short-Termism. *Journal of Finance 74*(4), 2011–2053. Publisher: Wiley Online Library.

Miao, J. and A. Rivera (2016). Robust contracts in continuous time. *Econometrica 84*(4), 1405–1440.

Murphy, K. J. (2003). The trouble with stock options. *Journal of Economic Perspectives 17*, 49–70.

Øksendal, B. K. (2003). *Stochastic differential equations: an introduction with applications*. Springer Verlag.

Piskorski, T. and A. Tchistyi (2011). Stochastic house appreciation and optimal mortgage lending. *Review of Financial Studies 24*(5), 1407–1446.

Polyanin, A. D. and V. F. Zaitsev (2002). *Handbook of exact solutions for ordinary differential equations*. Chapman & Hall/CRC.

Prat, J. and B. Jovanovic (2014). Dynamic contracts when the agent's quality is unknown. *Theoretical Economics 9*(3), 865–914.

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

Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*, Volume 11. Springer Science & Business Media.

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

Spear, S. E. and S. Srivastava (1987). On repeated moral hazard with discounting. *Review of Economic Studies 54*(4), 599–617.

Strulovici, B. and M. Szydlowski (2015). On the smoothness of value functions and the existence of optimal strategies in diffusion models. *Journal of Economic Theory 159*, 1016–1055.

Sung, J. (2015). Optimal Contracting Under Mean-Volatility Ambiguity Uncertainties. *Available at SSRN*.

Szydlowski, M. (2012). Ambiguity in Dynamic Contracts. Technical Report 1543, Center for Mathematical Studies in Economics And Management Science, Northwestern University.

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

Thompson, H. B. (1996). Second order ordinary differential equations with fully nonlinear two-point boundary conditions. II. *Pacific Journal of Mathematics 172*(1), 279–297.

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

Zhu, J. Y. (2016). Renegotiation of Dynamically Incomplete Contracts. *Working Paper*.

Zwiebel, J. (1996). Dynamic capital structure under managerial entrenchment. *American Economic Review 86*(5), 1197–1215.