Optimal Long-term Contracting with Learning*

Zhiguo He† Bin Wei‡ Jianfeng Yu§

First draft: January 2012
This draft: March 2013

Abstract

We introduce uncertainty into the Holmstrom and Milgrom (1987) infinite-horizon model to study optimal dynamic contracting with endogenous learning. The agent's potential belief manipulation leads to a hidden information problem. We characterize the optimal contract by solving a dynamic programming problem with one state variable. We find that in the optimal contract, the optimal effort decreases over time, exhibiting a front-loaded pattern. Furthermore, the optimal contract exhibits an option-like feature in that incentives increase after good performance.

Keywords: Executive Compensation, Moral Hazard, Bayesian Learning, Hidden Information, Belief Manipulation, Private Savings, Continuous-Time Optimal Contracting.

*We thank Philip Dybvig, Johannes Horner, Navin Kartik, Mark Loewenstein, Julien Prat, Yuliy Sannikov, and participants at Cheung Kong Graduate School of Business, the 2013 Econometric Society at San Diego, 2012 Society of Economic Dynamics Meeting, and 2012 China International Conference in Finance for helpful comments. This work does not necessarily reflect the views of the Federal Reserve System or its staff. All errors are our own. Jianfeng Yu gratefully acknowledges financial support from the Dean's Small Research Grant from the Carlson School of Management at the University of Minnesota.

†University of Chicago, Booth School of Business, and NBER. 5807 South Woodlawn Ave., Chicago, IL 60637. Phone: (1)773-834-3769. Email: zhiguo.he@chicagobooth.edu.

‡Board of Governors of the Federal Reserve System, Washington, DC 20551. Phone: (1)202-452-2693. Email: bin.wei@frb.gov.

§University of Minnesota, Carlson School of Management, CSOM 3-122, 321 19th Avenue South, Minneapolis, MN 55455. Phone: (1)612-625-5498. Email: jianfeng@umn.edu.

1 Introduction

Many long-term contractual relationships feature learning, as uncertainty arises if either the project's quality or the agent's ability is unknown when a contract is signed. We introduce stationary learning into a continuous-time-infinite-horizon variation of the classic Holmstrom and Milgrom (1987) model with a Constant-Absolute-Risk-Aversion (CARA) agent to study optimal long-term contracting with endogenous learning. More specifically, the principal signs a long-term contract with the agent, where the observable output dY_t each period is the sum of the agent's unobservable effort μ_t and the project's unknown profitability θ_t, plus some noise σdB_t:

$$dY_t = (\mu_t + \theta_t)\, dt + \sigma dB_t. \tag{1}$$

To focus on learning only, we assume that both the principal and agent share the same prior on the initial project profitability θ_0, and they learn the project profitability θ_t dynamically from the observable output.

Dynamic contracting with uncertainty and learning is theoretically challenging. Without learning, our setting is similar to the standard *"hidden action"* dynamic agency models where the agent's unobservable shirking has only a *short-lived* effect. In these pure "hidden action" models the history dependence of optimal contracting is fully summarized by the agent's continuation payoff, and, as a result, incentive provisions are essentially independent across periods.[1] Once we allow the output to also be affected by an uncertain profitability component (i.e., θ_t in equation 1), dynamic learning renders an intriguing link among incentive provisions across different times in long-term contracting.

The intertemporally linked incentive provisions are rooted in the following *persistent* "*hidden information*" problem. Along the equilibrium path the principal knows as much as the agent since both start with a common prior. However, along off-equilibrium paths the agent knows more, because only the agent knows his actual effort that is not observable to the principal and may deviate from the recommended level (i.e., "*hidden action*"). Imagine that the agent has followed the recommended effort policy in the past, and thus both parties share the same correct belief today. Suppose that the agent shirks today, i.e., exerting some effort below the recommended level. The lower effort decreases today's output on average. Furthermore, with Bayesian learning, the principal who anticipates a higher effort would mistakenly attribute today's weak performance to worse profitability. Thus, by shirking today the agent can distort downward the principal's inference about profitability from today on, which is long-lasting (i.e., persistent hidden information). This *belief manipulation* effect is beneficial to the agent, as the principal will mistakenly reward the agent later whenever

[1] See, for example, Holmstrom and Milgrom (1987), Spear and Srivastava (1987), Sannikov (2008), etc.

the future performance beats the principal's downward distorted expectations. We call this potential benefit due to off-equilibrium private information the agent's *information rent*, which increases with future pay-performance sensitivities. Because the information rent tarnishes the agent's current working incentives, incentive provisions at all dates are interlinked through this belief-manipulation effect.

We characterize the optimal contract in such a dynamic agency model with both "hidden action" (i.e., unobservable effort) and "hidden information" (i.e., potentially diverging beliefs about profitability). In Section 3 we first solve the agent's problem by characterizing the first-order optimality condition for any given compensation contract: the agent's optimal effort is simply the instantaneous incentive (i.e., pay-performance sensitivity) minus the information rent due to the belief manipulation effect. As mentioned, the information rent captures the marginal benefit of the agent distorting the principal belief downward, and can be conveniently expressed as the sum of properly discounted future incentives. The higher the future incentives, the greater the information rent, and the lower the agent's current working motivations. Further, Proposition 1 shows that, given the CARA preference, the local first-order optimality condition is sufficient to guarantee the global optimality of the agent's problem.

We then solve the principal's problem of designing the optimal contract in Section 4. Given the stationary nature of learning, there is no mechanical time effect. We reformulate the optimal contracting problem as a dynamic programming problem with two state variables: the agent's continuation value and the information rent. The CARA preference features no wealth effect. Although the agent is promised with different levels of continuation values after certain history of performance shocks, these promised continuation values themselves can be viewed as different promised wealth levels to the agent and thus do not affect the contracting problem looking forward. As a result, the optimal contract can be fully characterized by an ordinary differential equation (ODE) with the information rent as the only state variable, as we show in Section 5. Relative to Prat and Jovanovic (2012) and DeMarzo and Sannikov (2012) who focus on the contract that implements a constant first-best level of effort, we solve for the optimal effort path endogenously jointly with the optimal incentives in the long-term contract.

We find that the optimal effort policy exhibits two interesting features. First, we find that in our model the optimal effort policy, which is always distorted downward relative to the first-best benchmark, has a negative drift, thus exhibiting a front-loaded or time-decreasing pattern. In fact, in Section 5.2.1 we solve in closed form the optimal deterministic contract (i.e., the optimal one among the contracts in the subspace that implement deterministic policies only), and shows analytically that the optimal deterministic effort policy decreases over

time. This front-loaded pattern comes from the belief manipulation effect. As mentioned, later incentives increase the agent's current information rent for shirking. This implies that future pay-performance sensitivities hurt the agent's motive for working in earlier periods, but not the other way around. Given that later incentives are more costly, the optimal contract implements lower effort in later periods.

Interestingly, the pattern of time-decreasing effort policy in our paper with *post-contracting* information asymmetry[2] is opposite to the dynamic contracting setting with *pre-contracting* asymmetric information in Garrett and Pavan (2012): in that paper, the agent privately observes his productivities when signing the contract; and under the assumption that the effect of the initial productivity on his future productivity is declining over time, the optimal effort policy is time-increasing. Intuitively, in Garrett and Pavan (2012), downward distortion required for rent extraction is more severe in earlier periods when the major friction is pre-contracting private information. It is intriguing that pre-contracting private information and post-contracting information have opposite predictions for the time-series pattern of effort distortion, but the difference also lies on the agent being risk-neutral without wealth constraint in Garrett and Pavan (2012).[3] Relatedly, Sannikov (2012) allows for the possibility of the agent's current effort to affect future fundamentals, and finds that the optimal effort policy is increasing over time as well.

Second, we find that the optimal effort policy is stochastic with higher incentives after good performance, exhibiting an *option-like* feature.[4] The intuition comes from reducing the agent's belief manipulation in a long-term relationship. For a risk-averse agent, the amount of information rent not only depends on future pay-performance sensitivities, but also the agent's marginal utilities at these future states when receiving those manipulation benefits. Raising incentives after good performance introduces a negative correlation between pay-for-performance and marginal utility, i.e., allocating greater belief manipulation benefits in states where the agent cares less. Hence, the option-like compensation contract lowers the agent's information rent standing today.

This result of effort policy being history-dependent is more surprising given our setting.

[2] To be more precise, in our paper the post-contracting information asymmetry is caused by unobservable deviations in recommended effort levels, and thus exists only along off-equilibrium paths.

[3] In Garrett and Pavan (2012), since the agent is risk neutral and has sufficient wealth (or, no limited liability), without pre-contracting private information the principal can sell the project to the risk-neutral agent to achieve the first-best outcome, and there is no downward distrotion at all. In our model, it is efficient for the principal to offer insurance to the risk-averse agent. We expect the front-loaded pattern of optimal effort policy to be attenuated if we allow for the agent's pre-contracting private information, because of the effect in Garrett and Pavan (2012).

[4] Dittmann and Maug (2007) show that the optimal contract implied by standard principal-agent models almost never contains any options, a prediction contrary to practice. From this perspective, the endogenous *option-like* result is particularly interesting.

With a standard CARA-Normal setting and learning, as the posterior variance only changes over time deterministically (in our stationary setting, it is a constant), the resulting equilibrium effort profile is usually deterministic (e.g., Gibbons and Murphy, 1992; Holmstrom, 1999). In contrast, in our paper with profitability uncertainty and learning the optimal long-term contract has an option-like feature in that pay-for-performance rises following good performance. We emphasize that the combination of long-term contracting and learning drives both results. In Section 5.1, we first show that with long-term contracting but no learning, the model is a simple extension of Holmstrom and Milgrom (1987) and a constant effort policy is optimal. On the other hand, with short-term relationships and learning, the belief manipulation effect is still present so that the agent works less today if future pay-performance sensitivities are higher. However, the absence of commitment in the short-term relationships implies that principals at different times will not take this belief manipulation effect into account. This again implies a constant effort process in equilibrium, thanks to the Gaussian setting with stationary Bayesian learning (as in Holmstrom, 1999).

Though we rely on the specific setting (e.g., CARA preference, Gaussian processes, etc.) to characterize fully the optimal long-term contract with learning, the above two results are likely to hold qualitatively in a more general setting. Robustness exists because the economic forces driving these results do not depend on the CARA preference or Gaussian processes. In any long-term contracting environment with learning, the agent's information rent due to belief manipulation—that is, the agent would like to shirk to distort the principal's future belief downward—is general. The time-decreasing effort policy comes from the fact that later incentives enter the agent's forward-looking information rent in earlier periods (but not the other way around); and the option-like feature only relies on the agent's risk-aversion, so that the marginal value of earning future (potential) belief manipulation benefit is lower after the agent with good performance has received higher compensation.

Our paper is closest to DeMarzo and Sannikov (2012) and Prat and Jovanovic (2012). Both restrict attention to the optimal contract that implements a constant first-best level of effort. In contrast, we solve for the optimal effort policy endogenously joint with the optimal long-term compensation contract, and emphasize the general economic mechanisms that shape the optimal effort policy in long-term optimal contracting. Learning-wise, DeMarzo and Sannikov (2012), like our paper, consider a framework with stationary learning, while Prat and Jovanovic (2012) study "parameter uncertainty," which implies that the project profitability is an unknown constant rather than a moving target.[5] Theoretically speaking,

[5] On the assumption of the agent's preference, the agent in DeMarzo and Sannikov (2012) is risk neutral. Prat and Jovanovic (2012) cover general utility functions for the agent's problem, but then specialize exponential utility in deriving the optimal contract.

the exact learning technology is less important, as both of the above learning technologies lead to a long-lasting belief manipulation effect, which is the major challenge in solving for the optimal contract. Relatedly, the long-lasting belief manipulation also exists in Bergemann and Hege (1998, 2005) and Horner and Samulson (2012).

There are other papers that are related to learning but do not deal with the belief manipulation effect. Adrian and Westerfield (2009) focus on the disagreement between the principal and the agent about the agent's ability and the agent is dogmatic about his belief (i.e., the agent never updates his posterior belief about profitability from past performance), which eliminates the belief manipulation effect. In that paper, although the agent could distort the principal's belief by shirking, the dogmatic agent (who does not realize that the firm's profitability is, in fact, above the one perceived by the principal) will not gain anything from this channel, and as a result there is no belief distortion effect.[6]

The topic of optimal contracting with endogenous learning also relates to the recent literature studying optimal long-term contracts with adverse selection and moral hazard (e.g., Baron and Besanko, 1984; Sung, 2005; Sannikov, 2007; and Cvitanic, Wan, and Yang, 2012). Garrett and Pavan (2012) further allow agent's ability to be time-varying, and study the optimal retention policy. In Gershakov and Perry (2012), the agent whose skills are his private information faces a finite sequence of tasks and learns their level of difficulty. Halac, Kartik, and Liu (2012) introduce interaction between the agent's ability and project's profitability to study optimal experimentation. In general, when the agent has pre-contracting private information that is persistent, a mechanism design approach naturally arises; see Pavan, Segal and Toikka (2012) and Golosov, Troshkin, and Tsyvinski (2012), who use the first-order-approach to solve the agent's problem.[7] However, because our paper focuses on the problem without pre-contracting private information, we do not need to solve for the optimal menu for agent's truthful reporting when signing the contract.

The rest of the paper is organized as follows. Section 2 lays out the model, and Section 3 solves the agent's problem. Section 4 reformulates the principal's problem which allows for the use of dynamic programming techniques. Section 5 solves for the optimal contract and discusses its implications. Section 6 concludes.

[6] More recently, Cosimano, Speight, and Yun (2011) study the long-term contracting problem with binary unobservable productivity states, and show that the optimal contract tends to be sticky. They assume that the agent's effort is observable but not contractible, and hence both the principal and the agent always have the same information set, both on- and off-equilibrium paths.

[7] This is the same approach used in Williams (2009, 2011) and Zhang (2009), who study persistent information in a continuous-time principal-agent setting. We also use the first-order-approach to solve the agent's problem, and verify the global optimality of the agent's first-order-condition.

2 The Model

Consider a continuous-time infinite-horizon principal-agent model with a common constant discount rate $r > 0$. The project generates a cumulative output Y_t up to time t, which evolves according to

$$dY_t = (\mu_t + \theta_t)\, dt + \sigma dB_t, \tag{2}$$

where $\{B_t\}$ is a standard Brownian motion on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, μ_t is the agent's unobservable effort level, and the constant $\sigma > 0$ is the volatility of cash flows.

Relative to Holmstrom and Milgrom (1987), we introduce project profitability θ_t into the output process in equaiton (2). Equivalently, we can also interpret θ_t as the agent's unknown ability. Project profitability θ_t is unobservable and thus calls for learning. If project profitability θ_t is perfectly observable (or absent), there is no need for learning; In Section 5.1.1 we show that this case reduces to Holmstrom and Milgrom (1987). For tractability, we assume that unknown profitability affects output dY_t additively.

We assume that profitability $\{\theta_t\}$ follows a martingale process whose innovations are driven by another Brownian motion $\left\{B^\theta\right\}$ that is independent to $\{B\}$, i.e.,

$$d\theta_t = \phi\sigma dB_t^\theta \text{ and } B \perp B^\theta,$$

where $\phi > 0$ is a constant. At time 0, the principal and the agent share the common normal prior that $\theta_0 \sim \mathcal{N}\left(0, \Sigma_0^\theta\right)$, where we have normalized the prior mean to 0.

We mainly focus on stationary learning, except that in Section 5.2.2 we briefly discuss non-stationary learning for robustness check. For learning to be stationary, the prior uncertainty is assumed to satisfy $\Sigma_0^\theta = \sigma^2\phi$, so that the posterior variance $\Sigma_t^\theta = \Sigma_0^\theta$ for all t and Bayesian updating is time-independent. When $\phi = 0$, our model features no uncertainty, and thus is reduced to the benchmark model of Holmstrom and Milgrom (1987).

Both parties commit to the long-term relationship. The risk-neutral principal (*she* later on) offers the risk-averse agent (*he* later on) a contract $\{c_t, \mu_t\}$, so that the agent is recommended to take the effort policy $\mu = \{\mu_t\}$, and is compensated by the wage process $c = \{c_t\}$. Both elements are measurable to $\mathcal{Y}_t \equiv \mathcal{F}\{Y_s : 0 \leq s \leq t\}$, which is the filtration generated by the output history.

Moral hazard arises from the agent's unobservable effort choice. Furthermore, we assume that the agent can privately save (i.e., saving is unobservable) to smooth his consumption intertemporally. Thus the agent's actual consumption can differ from wage c_t. Denote the agent's actual consumption by $\widehat{c}_t$ and actual effort by $\widehat{\mu}_t$. Following Holmstrom and Milgrom (1987), we assume that the agent has an exponential or CARA preference:

$$u\left(\widehat{c}_t, \widehat{\mu}_t\right) = -\frac{1}{a}\exp\left[-a\left(\widehat{c}_t - g\left(\widehat{\mu}_t\right)\right)\right],$$

where $a > 0$ is the agent's absolute risk aversion coefficient, and $g(\widehat{\mu}_t) \equiv \frac{1}{2}\widehat{\mu}_t^2$ is the instantaneous quadratic monetary cost for effort $\widehat{\mu}_t$. The quadratic form of $g(\cdot)$ simplifies our results, but our analysis goes through as long as $g(\cdot)$ is strictly increasing and strictly convex. At the beginning when both parties sign the contract, the agent has no personal wealth and has an exogenous reservation utility of v_0. Without loss of generality, we assume the principal has all the bargaining power.

It is a celebrated result that CARA preferences do not have a wealth effect, and the issue of private savings can be easily dealt with (e.g., Fudenberg, Holmstrom, and Milgrom, 1990; Williams, 2009; He, 2011). In fact, private savings allow us to separate the agent's continuation payoff from another state variable in deriving the optimal contract, which renders the tractability of our problem.

2.1 Bayesian Learning and Effort

Recall that at time 0, the principal and the agent share the common normal prior $\theta_0 \sim \mathcal{N}\left(0, \Sigma_0^\theta\right)$. Both parties update their beliefs based on their own information sets respectively. Recall that $\mathcal{Y}_t = \mathcal{F}\{Y_s : 0 \leq s \leq t\}$ is the augmented filtration generated by output path Y. Given any contract $\{c_t, \mu_t\}$, the principal's information set at time t is $\mathcal{F}\{Y_s, \mu_s : 0 \leq s \leq t\}$, as the principal knows the recommended effort policy $\mu \equiv \{\mu_t\}$. However, the agent's information set also includes his actual effort policy $\widehat{\mu} \equiv \{\widehat{\mu}_t\}$, i.e., $\mathcal{F}\{Y_s, \mu_s, \widehat{\mu}_s : 0 \leq s \leq t\}$. Intuitively, relative to the principal, the agent knows (weakly) more because he knows his actual past effort choices $\widehat{\mu}$, which may deviate from the recommended policy μ. This distinction is important for our analysis.

If the agent follows the recommended effort policy μ, the principal's posterior belief about θ_t is correct and fully summarized by the first two moments:

$$m_t^\mu \equiv \mathbb{E}\left[\theta_t \left|\mathcal{Y}_t, \mu\right.\right] \text{ and } \Sigma_t^{\theta,\mu} \equiv \mathbb{E}\left[\left(\theta_t - m_t\right)^2 \left|\mathcal{Y}_t, \mu\right.\right].$$

Standard filtering argument (e.g., Theorem 12.2 in Liptser and Shiryayev, 1977) implies that

$$dm_t^\mu = \Sigma_t^{\theta,\mu} \frac{dY_t - (\mu_t + m_t^\mu)\, dt}{\sigma^2} = \sigma\phi dB_t^\mu, \text{ with } m_0 = 0, \tag{3}$$

$$\Sigma_t^{\theta,\mu} = \sigma^2\phi \text{ for all } t, \tag{4}$$

where B_t^μ is a standard Brownian motion under the measure induced by the effort policy μ:

$$dB_t^\mu = \frac{dY_t - (\mu_t + m_t^\mu)\, dt}{\sigma}. \tag{5}$$

Conditional on the actual effort policy $\{\widehat{\mu}_t\}$, the agent forms his posterior belief as

$$m_t^{\widehat{\mu}} \equiv \mathbb{E}\left[\theta_t \left|\mathcal{Y}_t, \widehat{\mu}\right.\right] \text{ and } \Sigma_t^{\theta,\widehat{\mu}} \equiv \mathbb{E}\left[\left(\theta_t - m_t^{\widehat{\mu}}\right)^2 \left|\mathcal{Y}_t, \widehat{\mu}\right.\right].$$

The superscript $\widehat{\mu}$ is used here to emphasize the dependence on the agent's actual effort policy $\widehat{\mu}$ (which the principal does not know). A similar argument yields

$$\begin{aligned} dm_t^{\widehat{\mu}} &= \Sigma_t^{\theta,\widehat{\mu}} \frac{dY_t - \left(\widehat{\mu}_t + m_t^{\widehat{\mu}}\right) dt}{\sigma^2} = \sigma\phi dB_t^{\widehat{\mu}}, \text{ with } m_0 = 0, \\ \Sigma_t^{\theta,\widehat{\mu}} &= \sigma^2\phi \text{ for all } t, \end{aligned} \tag{6}$$

where $B_t^{\widehat{\mu}}$ is a standard Brownian motion under the measure induced by the actual effort policy $\widehat{\mu}$:

$$dB_t^{\widehat{\mu}} \equiv \frac{1}{\sigma}\left(dY_t - \left(\widehat{\mu}_t + m_t^{\widehat{\mu}}\right) dt\right). \tag{7}$$

As we will show next, the potential belief divergence $m_t^{\widehat{\mu}} - m_t^{\mu}$ gives rise to the agent's incentive to manipulate the principal's belief, which plays an important role in the design of the optimal long-term contract with learning.

2.2 Effort and Belief Distortion

The difficulty of introducing learning into the dynamic moral hazard problem is not learning per se. Rather, the challenge is to deal with the issue of belief manipulation: the agent, simply by shirking from the recommended effort today, can distort the principal's future beliefs about project profitability downward.

Consider the following thought experiment. Suppose that at time t the agent chooses an effort level $\widehat{\mu}_t$ below the recommended effort μ_t, and thus output is lower than what is expected by the principal. Crucially, however, the principal thinks the agent is exerting an effort of μ_t, and, thus, through learning, mistakenly attributes lower output to a lower value of profitability θ_t. In contrast, the agent updates profitability θ_t based on his true effort level $\widehat{\mu}_t$, leading to a positive wedge $m_t^{\widehat{\mu}} - m_t^{\mu}$ between the beliefs of the agent and principal. In other words, by shirking, the agent makes the principal (mistakenly) underestimate project profitability. This belief manipulation is beneficial to the agent—when future outputs turn out to be high, the agent gets rewarded for high profitability (based on the agent's correct information set) rather than his effort.

We now formalize this effect. When the agent deviates from the recommended effort path μ by choosing effort policy $\widehat{\mu}$, the principal's posterior mean estimate about θ_t is distorted downward, and we denote this distortion by

$$\Delta_t \equiv m_t^{\widehat{\mu}} - m_t^{\mu} = \text{agent's belief - principal's belief.}$$

In the Gaussian framework with stationary learning, the posterior variances of both parties coincide to be a constant $\phi\sigma^2$. According to equations (3) and (6), the increment of Δ_t is

given by:

$$\begin{aligned} d\Delta_t &= dm_t^{\widehat{\mu}} - dm_t^{\mu} = \phi\left(dY_t - \left(\hat{\mu}_t + m_t^{\hat{\mu}}\right)dt\right) - \phi\left(dY_t - \left(\mu_t + m_t^{\mu}\right)dt\right) \\ &= \phi\left(\mu_t - \widehat{\mu}_t - \Delta_t\right)dt, \end{aligned}$$

which allows us to solve for Δ_t to be

$$\Delta_t = \phi \int_0^t e^{\phi(s-t)}\left(\mu_s - \widehat{\mu}_s\right)ds. \tag{8}$$

Here, we have used the fact that both the principal and the agent share the same belief when signing the contract at $t = 0$, i.e., $\Delta_0 = 0$ as $m_0^{\mu} = m_0^{\widehat{\mu}} = m_0$. Intuitively, equation (8) says that the current belief distortion is a properly-weighted cumulative effort deviation in the past. When $\phi = 0$, the zero prior uncertainty $\Sigma_0^{\theta} = \sigma^2\phi = 0$ eliminates any belief divergence and the issue of belief manipulation is absent.

Suppose that the employment contract relies on the agent's "unexpected" performance along the equilibrium path

$$dY_t - \left(\mu_t + m_t^{\mu}\right)dt, \tag{9}$$

which is a martingale increment σdB_t^{μ} under the equilibrium measure. For the agent who deviates by exerting $\widehat{\mu} \neq \mu$, under his information set the above performance is no longer a martingale incremental. Imagine that the agent has deviated before t so that $\widehat{\mu}_s \neq \mu_s$ where $s < t$. Even if the agent exerts the same effort at time t so that $\mu_t = \widehat{\mu}_t$, equation (7) implies that

$$dY_t - \left(\widehat{\mu}_t + m_t^{\widehat{\mu}}\right)dt = dY_t - \left(\mu_t + m_t^{\widehat{\mu}}\right)dt \tag{10}$$

is a martingle under the agent's information set. Comparing equation (9) to equation (10), we observe that the performance $dY_t - \left(\mu_t + m_t^{\mu}\right)dt$, which is "unexpected" for the principal along the equilibrium path, is no longer a marginale increment under the agent's information set. Rather, it displays a positive drift of belief divergence, because

$$\begin{aligned} dY_t - \left(\mu_t + m_t^{\mu}\right)dt &= dY_t - \left(\mu_t + m_t^{\widehat{\mu}}\right)dt + \left(\mu_t + m_t^{\widehat{\mu}}\right) - \left(\mu_t + m_t^{\mu}\right)dt \\ &= \left[dY_t - \left(\mu_t + m_t^{\widehat{\mu}}\right)dt\right] + \Delta_t dt, \end{aligned}$$

where the first term is a martingale increment under the agent's information set. Intuitively, if the agent shirked a bit in the past $\widehat{\mu}_s < \mu_s$ with $s < t$, then later on the principal would mistakenly think the project is worse than it actually is (under the agent's correct measure), i.e., $\Delta_t > 0$. As a result, the agent can easily beat the principal's expectation, which explains the positive drift $\Delta_t > 0$. As we emphasize later, this translates to a positive information rent to the agent, as long as the principal is providing incentives at time t to reward the agent for the "unexpected" performance.

2.3 Formulating the Optimal Contracting Problem

We first state the agent's problem. Denote by S_t the balance of the agent's savings account, which earns interest at the constant rate r. Given the contract $c = \{c_t, \mu_t\}$ the agent's problem is

$$\max_{\{\widehat{c},\widehat{\mu}\}} \mathbb{E}^{\widehat{\mu}} \left[\int_0^\infty e^{-rt} u\left(\widehat{c}_t, \widehat{\mu}_t\right) dt \right] \tag{11}$$
$$\text{s.t. } dY_t = \left(\widehat{\mu}_t + m_t^{\widehat{\mu}}\right) dt + \sigma dB_t^{\widehat{\mu}},$$
$$dS_t = rS_t dt + c_t dt - \widehat{c}_t dt \text{ with } S_0 = 0,$$

with the transversality condition $\lim_{T\to\infty} \mathbb{E}^{\widehat{\mu}} \left[e^{-rT} S_T\right] = 0$. Here, $\mathbb{E}^{\widehat{\mu}}[\cdot]$ denotes the probability measure induced by the agent's effort policy $\{\widehat{\mu}_t\}$, and $\{\widehat{c}_t\}$ is the agent's actual consumption policy. Denote the optimal solution to equation (11) by $\{c_t^\star, \mu_t^\star\}$.

We call the contract $\{c_t, \mu_t\}$ *incentive-compatible and no-savings* if, given the contract $\{c_t, \mu_t\}$, the solution to the agent's problem in equation (11) is $c_t^\star = c_t$ and $\mu_t^\star = \mu_t$, which further implies $S_t = 0$ for any t (i.e., no private savings at any time). In other words, the agent finds it optimal to consume his wages and work as recommended. As typical in the literature, the following lemma shows that there is no loss of generality when we restrict our attention to incentive-compatible and no-savings contracts. The basic idea is similar to the revelation principle. The principal who can fully commit to the contract can save for the agent, and, once she knows the agent's actual effort policy, she will perform correct Bayesian updating based on that policy.

Lemma 1 *It is without loss of generality to focus on contracts that are incentive-compatible and no-savings.*

Proof. *See Appendix A.* ■

The optimal contract solves the principal's problem:

$$\max_{\{c_t,\mu_t\} \text{ is incentive-compatible and no-savings}} \quad \mathbb{E}_t^{\mu} \left[\int_0^\infty e^{-rt} \left(dY_t - c_t dt\right) \right] \tag{12}$$
$$\text{s.t.} \quad dY_t = \left(\mu_t + m_t^{\mu}\right) dt + \sigma dB_t^{\mu}; \tag{13}$$
$$\mathbb{E}_0^{\mu} \left[\int_0^\infty e^{-rt} u\left(c_t, \mu_t\right) dt \right] = v_0. \tag{14}$$

In equation (13), the principal's Bayesian updating of m_t^μ and dB_t^μ is based on the correct optimal effort policy $\{\mu_t\}$ taken by the agent in the incentive-compatible contract. Equation (14) is the participation constraint, as the contract must attract the agent with a reservation value v_0 at $t = 0$. As standard in the optimal contracting literature with CARA agents, this participation constraint must bind.

3 The Agent's Problem

In this section we analyze the agent's problem. We illustrate heuristically the necessary conditions for a contract $\{c_t, \mu_t\}$ to be incentive-compatible and induce no private savings. We further establish the sufficiency of the necessary conditions. Formal proofs are delegated to Appendix A.

3.1 Continuation Value and Incentives

Given the incentive-compatible and no-savings compensation contract $\{c_t, \mu_t\}$, the agent's continuation value is defined as:

$$v_t \equiv \mathbb{E}_t^{\mu}\left[\int_t^{\infty} e^{-r(s-t)} u\left(c_s, \mu_s\right) ds\right]. \tag{15}$$

According to the standard martingale representation argument (e.g., Sannikov, 2008), there exists some progressively measurable process $\{\beta_t\}$ so that

$$\begin{aligned} dv_t &= rv_t dt - u\left(c_t, \mu_t\right) dt + \beta_t\left(-arv_t\right)\left(dY_t - m_t^{\mu} dt\right) \\ &= rv_t dt - u\left(c_t, \mu_t\right) dt + \beta_t\left(-arv_t\right) \sigma dB_t^{\mu}. \end{aligned} \tag{16}$$

Intuitively, $\beta_t\left(-arv_t\right)$ can be interpreted as the incentive loading (in utils) on the agent's unexpected performance $dY_t - m_t^{\mu} dt$. As we show shortly, $-arv_t > 0$ is the agent's marginal utility from consumption at time t, i.e., $u_c\left(c_t, \mu_t\right)$. Thus we can interpret β_t as dollar incentives on the agent's unexpected performance. Later we refer to pay-performance sensitivities $\{\beta_t\}$ simply as incentives.

3.2 No Savings

We first show that, under CARA preference, no-savings conditions imply

$$rv_t = u\left(c_t, \mu_t\right) = -\frac{1}{a} \exp\left[-a\left(c_t - g\left(\mu_t\right)\right)\right]. \tag{17}$$

The argument is similar to He (2011). To see this, we first present the following lemma, where we use Π to denote any compensation contract.

Lemma 2 *At any time $t \geq 0$, consider a deviating agent who has some arbitrary savings S_t and faces the continuation contract Π_t. Denote by $v_t\left(S; \Pi\right)$ his deviation continuation value. We have*

$$v_t\left(S; \Pi\right) = v_t\left(0; \Pi\right) \cdot e^{-arS} = v_t \cdot e^{-arS}, \tag{18}$$

where we have used the fact that $v_t\left(0; \Pi\right)$ is the agent's continuation value v_t along the no-savings path defined in equation (15).

Proof. *See Appendix A.* ■

The driving force behind this result is simple. Due to the CARA preference, the agent's problem is translation-invariant with respect to his underlying wealth level, as evidenced by $u(c_s + rS, \mu_s) = e^{-arS} u(c_s, \mu_s)$. Thus, for a CARA agent, given the extra savings S, his new optimal policy is to take the optimal consumption-effort-learning policy without savings, and consume an extra rS more for all future dates. As suggested by equation (18), the optimal deviation value with savings S is just the original value without savings, multiplied by the adjusting factor e^{-arS} for extra consumption.

By the optimality of the agent's consumption-savings policy in equation (11), his marginal utility from consumption must equal his marginal value of wealth, and equation (18) implies that:

$$u_c(c_t, \mu_t) = \left.\frac{\partial v_t(S; \Pi)}{\partial S}\right|_{S=0} = -arv_t. \tag{19}$$

Thus, equation (17) follows immediately from equation (19) by using the fact that under CARA preference, the agent's utility level is linear in his marginal utility:

$$au(c_t, \mu_t) = -u_c(c_t, \mu_t). \tag{20}$$

Plugging equation (17) into equation (16), we find that for no-saving contracts, v_t follows an exponential martingale:

$$\begin{aligned} & dv_t = \beta_t(-arv_t)\sigma dB_t^\mu \\ \Leftrightarrow \quad & v_s = v_t \exp\left(-\int_t^s ar\beta_u \sigma dB_u^\mu - \frac{1}{2}\int_t^s a^2 r^2 \beta_u^2 \sigma^2 du\right) \text{ for } s > t. \end{aligned} \tag{21}$$

We can also understand this result that v_t is a martingale by combining two observations: first, the agent can smooth out his consumption intertemporally so that his marginal utility has to follow a martingale; second, his continuation value v_t is linear in his marginal utility u_c because of equations (19) and (20).

3.3 Incentive-Compatibility Constraint and Intuition

3.3.1 Solution to the agent's problem

Proposition 1 characterizes the agent's incentive compatibility constraint, along with the equilibrium consumption and continuation value heuristically derived above. In Appendix A, we provide a rigorous proof for Proposition 1 based on the Pontryagin maximum principle. There, we show that the policy (c, μ) satisfying the (local) necessary conditions is also globally optimal. In the concurrent paper by Prat and Jovanovic (2012), the global optimality is guaranteed with certain sufficient conditions, which are dependent on some endogenous variables and, thus, might be hard to verify. With the CARA preference and private savings

considered in this paper, we are able to establish the sufficiency of the incentive compatibility constraint.

Proposition 1 *For the contract* $\{c_t, \mu_t\}$ *to be incentive-compatible and no-savings,* $\{\beta_t\}$ *must satisfy*

$$\mu_t = \underbrace{\beta_t}_{\textit{instantaneous incentive}} - \underbrace{\frac{1}{-arv_t}\mathbb{E}_t^{\mu}\left[\int_t^{\infty} \phi e^{-(\phi+r)(s-t)}\beta_s\left(-arv_s\right)ds\right]}_{\textit{information rent}} \tag{22}$$

$$= \beta_t - \mathbb{E}_t^{\mu}\left[\int_t^{\infty} \phi e^{-(\phi+r)(s-t)}\beta_s \exp\left(-\int_t^s ar\beta_u\sigma dB_u^{\mu} - \frac{1}{2}\int_t^s a^2r^2\beta_u^2\sigma^2 du\right)ds\right] \tag{23}$$

In addition, equation (17) implies that consumption (or wage) follows

$$c_t = g\left(\mu_t\right) - \frac{\ln\left(-arv_t\right)}{a}, \tag{24}$$

and the continuation payoff from the contract is

$$v_t = v_0 \exp\left(-\int_0^t ar\beta_s\sigma dB_s^{\mu} - \frac{1}{2}\int_0^t a^2r^2\beta_s^2\sigma^2 ds\right). \tag{25}$$

Proof. *See Appendix A.* ■

We devote the rest of this section to explain the intuition behind the effort optimality condition in equation (22). In a standard dynamic agency problem without profitability uncertainty (e.g., $\phi = 0$), the agent's effort μ_t at time t should only depend on the time-t incentive β_t offered by the contract (i.e., $\mu_t = \beta_t$). With learning and associated belief-manipulation, the agent's effort decisions across periods are interlinked, as suggested by the forward-looking nature of the second downward adjustment term in equation (22).

The forward-looking downward adjustment term represents the *information rent* to the agent. Intuitively, this term captures the marginal benefit of manipulating the principal's belief downward from then on.[8] Because the principal rewards β_{t+s} to the agent's future "unexpected" positive performance at $t+s$ for all $s > 0$, loosely speaking, the total benefit is the expected (properly) discounted future incentives $\left\{\beta_{t+s} : s > 0\right\}$. The discount rate is $\phi + r$ due to the combination of stationary learning discounting and time discounting. In addition, as explained later, we also adjust for marginal utilities (which is $-arv_s$) at future states.

[8] This information rent term captures the *marginal* rent that the agent may enjoy by deviating from the recommended effort slightly, rather than the rent that the agent actually enjoys in equilibrium. It is because in equilibrium the principal knows the agent's actual effort and hence the agent does not know more than the principal does. However, as any typical moral hazard models, the marginal deviation benefit (marginal rent) is important in characterizing the agent's incentive-compatibility condition.

3.3.2 Intuition for incentive-compatibility constraint

Consider the agent who reduces his effort to slightly below the recommended effort level μ_t, say $\mu_t - \epsilon$, only at the time interval $[t, t+dt]$. In other words, given the recommended policy $\{\mu_s\}$, the deviation effort policy is

$$\mu^\epsilon = \begin{cases} \mu_s & \text{for } s \notin [t, t+dt] \\ \mu_s - \epsilon & \text{otherwise} \end{cases}. \tag{26}$$

What is the impact on the agent's total payoff onwards?

Standing at time t, the agent's total payoffs include his instantaneous utility $u(c_t, \mu_t)$ and his continuation payoff. More specifically, we write the agent's total payoff from time t onwards as follows:[9]

$$\begin{aligned}
& u(c_t, \mu_t - \epsilon)\, dt + v_t + \mathbb{E}_t^{\mu^\epsilon}\left[\int_t^\infty e^{-r(s-t)} dv_s\right] \\
= \; & u(c_t, \mu_t - \epsilon)\, dt + v_t + \mathbb{E}_t^{\mu^\epsilon}\left\{\begin{array}{c} \beta_t(-arv_t)\left(dY_t(\mu_t - \epsilon) - m_t^\mu dt\right) + \\ \int_{t+dt}^\infty e^{-r(s-t)}\beta_s(-arv_s)\left[dY_s - (\mu_s + m_s^\mu)\, ds\right] \end{array}\right\} \\
= \; & \underbrace{u(c_t, \mu_t - \epsilon)\, dt}_{\text{saving effort cost instantaneously}} + v_t + \\
& \mathbb{E}_t^{\mu^\epsilon}\left\{\begin{array}{c} \underbrace{\beta_t(-arv_t)\left(dY_t(\mu_t - \epsilon) - m_t^\mu dt\right)}_{\text{hurting performance instantaneously}} + \\ \underbrace{\int_{t+dt}^\infty e^{-r(s-t)}\beta_s(-arv_s)\left[\underbrace{\left(dY_s - \left(\mu_s^\epsilon + m_s^{\mu^\epsilon}\right) dt\right)}_{\text{martingale under info set generated by } \mu} + \underbrace{\Delta_s ds}_{\text{future belief divergences } \Delta_s}\right]}_{\text{creating belief divergence persistently}} \end{array}\right\}
\end{aligned} \tag{27}$$

In the first equality, $\mathbb{E}_t^{\mu^\epsilon}$ implies that the agent forms his expectation based on his information set induced by μ^ϵ. In the second equality, we have used the result in equation (21). There should be another correction term in $(\mu_s^\epsilon - \mu_s)\, ds$ in the third equality, but it is zero because of (26), i.e., we consider an one-shot deviation at time t from the equilibrium effort policy.

As the second line in equation (27) shows, there are two channels through which shirking at time t affects future changes in the agent's continuation payoffs. The first channel captures the instantaneous performance effect, i.e., the agent's effort affects instantaneous performance dY_t and, thus, his continuation payoff. To see this, write performance $dY_t(\mu_t)$ over $[t, t+dt]$ as a function of time-t effort μ_t. Exerting effort $\mu_t - \epsilon$ will then hurt the short-term performance over $[t, t+dt]$ because

$$dY_t(\mu_t - \epsilon) = (\mu_t - \epsilon)\, dt + m_t^\mu dt + \sigma dB_t^\mu = dY_t(\mu_t) - \epsilon dt.$$

[9] To see the first line is the agent's total payoff from time t onwards given any effort policy μ and c, define $G(t) \equiv \int_0^t e^{-rs} u_s ds + e^{-rt} v_t$ and $G(\infty)$ is the agent's total payoff. Due to private savings, $u_s = rv_s$, and we have $dG(t) = e^{-rt} dv_t$. Therefore the total payoff (inflated by e^{rt}) is $\mathbb{E}_t^\mu\left[e^{rt} G(\infty)\right] = e^{rt} G(t) + \mathbb{E}_t^\mu\left[\int_t^\infty e^{-r(s-t)} dv_s\right]$, which is $u(c_t, \mu_t)\, dt + v_t + \mathbb{E}_t^\mu\left[\int_t^\infty e^{-r(s-t)} dv_s\right]$ by ignoring utilities occuring before t. Under equilibrium effort, dv_s is martingale increment and thus $\mathbb{E}_t^\mu\left[\int_t^\infty e^{-r(s-t)} dv_s\right] = 0$.

Modulated by incentives, this leads to a drop in the agent's continuation payoff by $\beta_t(-arv_t)\cdot \epsilon dt$, via the channel of "hurting performance instantaneously."

The second channel is the persistent effect due to belief manipulation: as discussed in Section 2.2, the agent's shirking at time t shifts the belief divergence path $\{\Delta_s\}$ away from the equilibrium path $\{\Delta_s = 0\}$ for $s > t$. We have seen from equation (8) that once shirking creates a positive time-t belief divergence between the principal and the agent, this belief divergence will persist in future contractual relations.

Now we show that the incentive-compatibility constraint in equation (22) is implied by equation (27). By "reducing effort cost instantaneously" in equation (27), the agent's marginal gain from shirking at t is $-u_\mu(c_t, \mu_t)\cdot \epsilon dt$. Since $u_\mu(c_t, \mu_t) = -u_c(c_t, \mu_t)\mu_t = arv_t\mu_t$, this marginal gain is $(-arv_t)\mu_t \cdot \epsilon dt$. On the other hand, shirking "hurts performance instantaneously" in equation (27), which gives rise to a marginal cost of $\beta_t(-arv_t)\cdot \epsilon dt$. In standard models without belief manipulation, these two forces fully determine the agent's trade-off in choosing his optimal effort at time t.

Now we analyze the novel term "creating belief divergence persistently" in equation (27). There, because $dY_s - \left(\mu_s^\epsilon + m_s^{\mu^\epsilon}\right)dt$ are martingale increments, the term is reduced to

$$\mathbb{E}_t^{\mu^\epsilon}\left[\int_t^\infty e^{-r(s-t)}\beta_s(-arv_s)\Delta_s ds\right]. \tag{28}$$

As equation (26) suggests, the agent is shirking only at time $[t, t+dt]$, but never before or after. Thus, using equation (8) we derive the belief divergence in any future time $s > t$ to be

$$\Delta_s = \phi\int_0^s e^{\phi(s-t)}(\mu_t - \mu_t^\epsilon)dt = \phi e^{-\phi(s-t)}\cdot \epsilon dt. \tag{29}$$

As new information flows in, the belief divergence persists but decays over time exponentially. Plugging equation (29) into equation (28), the marginal impact of shirking via the channel of belief manipulation is

$$\mathbb{E}_t^{\mu^\epsilon}\left[\int_t^\infty \phi e^{-(\phi+r)(s-t)}\beta_s(-arv_s)ds\right]\cdot \epsilon dt = \mathbb{E}_t^{\mu}\left[\int_t^\infty \phi e^{-(\phi+r)(s-t)}\beta_s(-arv_s)ds\right]\cdot \epsilon dt + o(\epsilon dt)$$

where the second term is because the difference between the expectation terms $\mathbb{E}_t^{\mu^\epsilon}[\cdot]$ and $\mathbb{E}_t^{\mu}[\cdot]$ is (at most) in the order of ϵdt.[10] Intuitively, if the principal mistakenly believes that the project is less profitable than it should be, the agent's normal performance will be considered superb. The higher-powered the incentives $\{\beta_s\}$, the greater the information rent that the agent can enjoy. Finally, for a risk averse agent, the information rent depends on

[10] As we are mainly for illustrating intuition, this statement is heuristic. For rigorous argument, see the Appendix for the proof of Proposition 1.

the agent's future marginal utility $(-arv_s)$ when receiving those manipulation benefits, a result that is useful later in Section 5.4.

Combining three pieces together and canceling ϵdt (and ignoring higher order terms), the agent's local incentive-compatibility constraint is

$$(-arv_t)\,\mu_t - \beta_t\,(-arv_t) + \mathbb{E}_t^{\mu}\left[\int_t^{\infty} \phi e^{-(\phi+r)(s-t)}\beta_s\,(-arv_s)\,ds\right] = 0.$$

After dividing both sides by time-t marginal utility $(-arv_t)$ and using equation (25), we reach the key incentive-compatibility condition in equation (23). Proposition 1 further establishes that this local incentive-compatibility constraint guarantees the global optimality of the agent's solution.

4 The Principal's Problem

From now on we focus on incentive-compatible contracts such that both parties will have the same information set. For ease of notation we use dB_t to denote dB_t^{μ}, and write $\mathbb{E}^{\mu}$ as $\mathbb{E}$.

4.1 Rewriting the Principal's Problem

In light of Proposition 1, we first rewrite the principal's problem in equation (12). Proposition 1 establishes an important link between recommended effort $\{\mu_t\}$ and incentives $\{\beta_t\}$ in any incentive-compatible contracts. Moreover, the principal can choose the optimal $\{\beta_t^*\}$ to maximize her value, and the corresponding optimal consumption process $\{c_t^*\}$ and the optimal effort policy $\{\mu_t^*\}$ are determined by equation (24) and equation (22), respectively. Therefore, we can rewrite the principal's problem in equation (12) as

$$\max_{\{\beta_t\}} \mathbb{E}\left[\int_0^{\infty} e^{-rt}\,(dY_t - c_t dt)\right] \tag{30}$$

$$s.t. \quad dY_t = (\mu_t + m_t)dt + \sigma dB_t \text{ and } dm_t = \phi\sigma dB_t, \tag{31}$$

$$c_t = g\,(\mu_t) - \frac{\ln\,(-arv_t)}{a}, \text{ where } g\,(\mu_t) = \frac{1}{2}\mu_t^2, \tag{32}$$

$$dv_t = \beta_t\,(-arv_t)\,\sigma dB_t, \text{ given } v_0; \tag{33}$$

$$\mu_t = \beta_t - \mathbb{E}_t\left[\int_t^{\infty} \phi\beta_s e^{-(\phi+r)(s-t)} \exp\left(-\int_t^s ar\beta_u\sigma dB_u - \frac{1}{2}\int_t^s a^2r^2\beta_u^2\sigma^2 du\right) ds\right] \tag{34}$$

Here, equation (31) describes the dynamics of output and posterior belief; equations (32)-(34) are derived from equations (23)-(25) in Proposition 1.

We will use the technique of dynamic programming to solve the above problem. The formulation above suggests that we need to keep track of two state variables: one is the

agent's continuation payoff v_t and the other is the the second term "information rent"

$$\mathbb{E}_t\left[\int_t^\infty \phi\beta_s e^{-(\phi+r)(s-t)}\exp\left(-\int_t^s ar\beta_u\sigma dB_u - \frac{1}{2}\int_t^s a^2r^2\beta_u^2\sigma^2 du\right)ds\right]$$

on the right-hand side of equation (34). As we discussed before, the information rent is also the agent's marginal benefit of shirking in manipulating the principal's belief.

Importantly, we observe that in equation (34) the information rent is independent of the agent's continuation value v_t. Indeed, thanks to the CARA preference, the optimal contracting problem can be rewritten without v_t. We start from the principal's objective in equation (30). For outputs, using equation (31) we have (m_t is a martingale with an initial value $m_0 = 0$):

$$\mathbb{E}\left[\int_0^\infty e^{-rt}dY_t\right] = \mathbb{E}\left[\int_0^\infty e^{-rt}(\mu_t + m_t)dt\right] = \mathbb{E}\left[\int_0^\infty e^{-rt}\mu_t dt\right],$$

In Appendix A, for compensation cost, we use equation (32) and integration by parts to show that

$$\begin{aligned}\mathbb{E}\left[e^{-rt}c_t dt\right] &= \mathbb{E}\left[\int_0^\infty e^{-rt}\left(g(\mu_t) - \frac{1}{a}\ln(-arv_t)\right)dt\right] \\ &= \underbrace{-\frac{\ln(-arv_0)}{ar}}_{\text{certainty equivalent of outside option } v_0} + \mathbb{E}\left[\int_0^\infty e^{-rt}\left(\underbrace{g(\mu_t)}_{\text{effort cost}} + \underbrace{\frac{1}{2}ar\sigma^2\beta_t^2}_{\text{risk compensation}}\right)dt\right].\end{aligned}$$

The total compensation cost is the certainty equivalent $-\frac{\ln(-arv_0)}{ar}$ of delivering the agent's outside option v_0, plus the (monetary) effort cost $g(\mu_t) = \mu_t^2/2$ and the discounted risk compensation due to incentive provisions. Thus, the certainty equivalent $-\frac{\ln(-arv_0)}{ar}$ separates from the problem, and the optimal solution $\{\beta_t^*\}$ will be independent of the agent's initial outside option v_0. This result comes from the translation invariance of CARA preference.

In sum, the principal's problem can be simplified to

$$\max_{\{\beta_t\}}\mathbb{E}\left[\int_0^\infty e^{-rt}\left(\mu_t - \frac{1}{2}\mu_t^2 - \frac{1}{2}ar\sigma^2\beta_t^2\right)dt\right] \tag{35}$$

$$\begin{aligned} s.t.\ \ \mu_t &= \beta_t - \mathbb{E}_t\left[\int_t^\infty \phi\beta_s e^{-(\phi+r)(s-t)}\exp\left(-\int_t^s ar\beta_u\sigma dB_u - \frac{1}{2}\int_t^s a^2r^2\beta_u^2\sigma^2 du\right)ds\right], \\ \beta_t &\in [-M, M] \text{ for some sufficiently large } M.\end{aligned}$$

We impose the last constraint as the transversality condition on our infinite horizon problem. Essentially, we restrict the feasible incentive slopes $\{\beta_t\}$ to be bounded, i.e., there exists some sufficiently large constant M such that $\beta_t \in [-M, M]$. Importantly, as we will prove later, the derived optimal policy is independent of M when M is sufficiently large.

4.2 Recursive Formulation of the Problem

4.2.1 Information rent as the unidimensional state variable

In light of the incentive-compatibility constraint in equation (23), define the information rent p_t as

$$p_t \equiv \mathbb{E}_t \left[\int_t^\infty \phi \beta_s e^{-(\phi+r)(s-t)} \exp\left(-\int_t^s ar\beta_u \sigma dB_u - \int_t^s \frac{1}{2} (ar\beta_u \sigma)^2 du \right) ds \right]. \tag{36}$$

As shown shortly, the information rent p_t serves as the only state variable for the principal to design the optimal contract. Define

$$\pm M_p \equiv \pm \frac{\phi M}{\phi + r}.$$

The following lemma shows that the information rent p_t is bounded in $[-M_p, -M_p]$ if incentives β_t is bounded in $[-M, M]$, and the boundaries for p_t are absorbing. This also gives us a simple boundary condition for the value function V.

Lemma 3 *Suppose that $\beta_t \in [-M, M]$ where M is a given constant. Then the state variable p_t reaches $\pm M_p$ if and only if $\beta_s = \pm M, \forall s \geq t$, which implies that $\pm M_p$ are absorbing states for p. As a result, when $p = \pm M_p$, $V(\pm M_p)$ is quadratic in M:*

$$V(p = \pm M_p) = \pm \frac{M}{\phi + r} - \frac{r}{2} \left[\frac{1}{(\phi + r)^2} + a^2 \sigma^2 \right] M^2. \tag{37}$$

Proof. *See Appendix A.* ■

When the state variable $p_t \in (-M_p, M_p)$ is away from boundaries, using equation (36) in Appendix A we show that the dynamics of p_t follows

$$dp_t = [(\phi + r) p_t + \beta_t (ar\sigma\sigma_t^p - \phi)] dt + \sigma_t^p dB_t, \tag{38}$$

where σ_t^p is some progressively measurable process derived from the martingale representation of equation (36). From now on, we interpret $\{\sigma_t^p, \beta_t\}$ as our control because the pair determines the drift and diffusion of p_t in equation (38). Note that we will derive σ_t^p and β_t as a function of the auxiliary state p_t. Combined with the boundary condition that $\beta_s = \pm M$ for $s > t$ whenever p_t hits $\pm M_p$, the control pair $\{\sigma_t^p, \beta_t\}$, as functions of the auxiliary state p_t, gives the full history of $\{\beta_t : t \geq 0\}$ that we are after.

Now we are able to write the problem (35) in the standard dynamic programming language. Substituting $\mu_t = \beta_t - p_t$ in the principal's objective, we have

$$\max_{\{\beta_t, \sigma_t^p\}} \mathbb{E} \left\{ \int_0^\infty e^{-rt} \left[(\beta_t - p_t) - \frac{1}{2} (\beta_t - p_t)^2 - \frac{1}{2} ar\sigma^2 \beta_t^2 \right] dt \right\}$$

$$\begin{aligned} s.t.\ dp_t &= [(\phi + r) p_t + \beta_t (ar\sigma\sigma_t^p - \phi)] dt + \sigma_t^p dB_t \text{ for all } t > 0, \text{ and } p_0 = p \\ \beta_s &\in [-M, M], p_t \in [-M_p, +M_p], \text{ and } p_t = \pm M_p \text{ are absorbing.} \end{aligned}$$

We are after the optimal policy $\{\beta_t^*, \sigma_t^{p,*}\}$ as functions of the state variable p_t.

4.2.2 Relaxed problem

We first consider the principal's relaxed maximization problem given M (and $M_p = \frac{\phi M}{\phi + r}$):

$$V(p; M) \equiv \max_{\{\beta_t, \sigma_t^p\}} \mathbb{E}\left[\int_0^\infty e^{-rt}\left(\mu_t - \frac{1}{2}\mu_t^2 - \frac{1}{2}ar\sigma^2\beta_t^2\right)dt\right] \tag{39}$$

$$\begin{aligned} s.t.\ dp_t &= \left[(\phi + r)\, p_t + \beta_t\left(ar\sigma\sigma_t^p - \phi\right)\right]dt + \sigma_t^p dB_t \text{ for all } t > 0, \text{ and } p_0 = p \\ &p_t \text{ is absorbing at } \pm M_p, \\ &\beta_t \text{ can exceed } M \text{ (but remains finite) when } p_t \in (-M_p, M_p). \end{aligned} \tag{40}$$

The problem (39) is a relaxed version of the principal's original problem (35), due to equation (40). Essentially, given M, in the original problem (35) we require $\beta_t \in [-M, M]$ for any time t; while in problem (39) we only require $\beta_t \in [-M, M]$ whenever p_t hits $\pm M_p$ in light of Lemma 3. Thus, when $p = \pm M_p$, the boundary conditions are the same between these two problems. However, this relaxation helps because the relaxed problem (39) allows us to use the interior first-order condition of β when $p \in (-M_p, M_p)$. We will show that for sufficiently large M the value achieved in the relaxed problem is the same as that in the original problem, which implies that the solution to the relaxed problem is also that to the original problem.

The Hamilton-Jacobi-Bellman (HJB) equation for problem (39) is straightforward:

$$rV = \max_{\beta, \sigma^P} (\beta - p) - \frac{1}{2}(\beta - p)^2 - \frac{ar\sigma^2}{2}\beta^2 + V_p\left[(\phi + r)\, p + \beta\left(ar\sigma\sigma^P - \phi\right)\right] + \frac{1}{2}V_{pp}\left(\sigma^P\right)^2. \tag{41}$$

Under the assumption that $1 + ar\sigma^2 + a^2r^2\sigma^2\frac{(V_p)^2}{V_{pp}} > 0$ and $V_{pp} < 0$ that we will verify later in Proposition 5, the first-order optimality conditions are given by:

$$\beta = \frac{1 + p - \phi V_p}{1 + ar\sigma^2 + a^2r^2\sigma^2\frac{(V_p)^2}{V_{pp}}} \text{ and } \sigma^p = -ar\sigma\beta\frac{V_p}{V_{pp}}. \tag{42}$$

Plugging them back into the HJB equation (41) and dropping M in $V(p; M)$ without risk of confusion, we have

$$rV = \frac{1}{2}\frac{(1 + p - \phi V_p)^2}{1 + ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}} - p - \frac{1}{2}p^2 + V_p(\phi + r)\, p. \tag{43}$$

We will solve the problem in equation (39) by analyzing the above Ordinary Differential Equation (ODE) in equation (43) with the boundary condition in equation (37).

5 Optimal Contracting

5.1 Two Important Benchmarks

Before we move on to solve the optimal long-term contract with learning, we present results for two benchmark cases: one is observable profitability without learning while the other is with learning but only short-term contractual relations.

5.1.1 When Profitability is Observable

We first consider the benchmark case where the profitability θ_t is observable. This case is almost the same as the standard Holmstrom and Milgrom (1987) model, except that the optimal contract always benchmarks the agent's performance to θ_t. Using the incentive constraint $\mu_t = \beta_t$, the optimal solution is

$$\mu_t^{HM} = \frac{1}{1 + ar\sigma^2},$$

and the principal's value is $V^{HM} = 1/\left(2r\left(1 + ar\sigma^2\right)\right)$. The optimal contract can be implemented by a constant equity share $1/\left(1 + ar\sigma^2\right)$ (with proper benchmarking). Because the principal can ignore the direct information about θ_t, the value V^{HM} derived in the observable profitability case serves as an upper bound for our value function $V(p)$ when profitability is unobservable:

$$V(p) \leq V^{HM} = \frac{1}{2r\left(1 + ar\sigma^2\right)}. \tag{44}$$

5.1.2 When Contractual Relation is Short-term

Consider another polar case where the contractual relationship is short-term. Imagine the following setting, where a long-lived agent with unknown ability θ_t is working for a continuum of principals. At any time $t > 0$ there is one principal who signs a short-term incentive contract with the agent. The relation, however, only lasts for the interval $[t, t + dt]$. The short-term contract consists of a fixed wage α_t, an incentive β_t, and the recommended effort μ_t, so that at the beginning of $t + dt$ the agent receives a compensation flow of

$$\alpha_t dt + \beta_t \left(dY_t - \mu_t dt - m_t dt\right),$$

given date t belief, $\mathbb{E}_t[\theta_t] = m_t$, and output performance dY_t over $[t, t + dt]$. At the end of period $t+dt$, the relation breaks, and the agent is signing another contract $\left\{\alpha_{t+dt}, \beta_{t+dt}\right\}$ with another principal indexed by $t + dt$. Importantly, short-term relations rule out inter-period commitment, so that each principal takes other principals' equilibrium offers as given.

For simplicity, to determine the history of fixed wages $\{\alpha_t\}$, we assign all the bargaining power to principals. The following proposition states that the equilibrium incentives and effort are constant over time.

Proposition 2 *When a contractual relationship is short-term and principals have all the bargaining power, incentives are constant over time:*

$$\beta_t = \beta^{ST} = \frac{\phi + r}{r + ar\sigma^2 (\phi + r)} \text{ for all } t,$$

and the equilibrium effort is constant over time as well

$$\mu_t = \frac{r}{\phi + r}\beta^{ST} = \frac{r}{r + ar\sigma^2 (\phi + r)} \text{ for all } t.$$

Proof. *See Appendix A.* ■

Here is the intuition. When the principals have all the bargaining power, Proposition 1 still applies to the agent's problem.[11] Thus, given today's incentive β_t and continuation incentives $\left\{\beta_{t+s} : s > 0\right\}$, the agent sets $\mu_t = \beta_t - p_t$, where p_t is discounted future incentives as in equation (36). The principal at time t who takes p_t as given will maximize the expected output $\mu_t + m_t = \beta_t - p_t + m_t$, minus the total compensation which is the sum of the effort cost $g(\beta_t - p_t)$ and the risk compensation $\frac{1}{2}ar\sigma^2\beta_t^2$. Ignoring the given project quality m_t, the principal's maximization problem at time t is only maximizing the flow payoff in equation (35) or equation (41)

$$\max_{\beta_t} (\beta_t - p_t) - \frac{1}{2}(\beta_t - p_t)^2 - \frac{ar\sigma^2}{2}\beta_t^2 \tag{45}$$

This important observation implies that, with short-term contracting, each principal ignores the long-term consequence of his short-term incentives. The optimality condition for this simple quadratic problem (45) yields

$$\beta_t^{ST} = \frac{1 + p_t^{ST}}{1 + ar\sigma^2},$$

where p_t^{ST} is defined as in equation (36) with future incentives $\left\{\beta_{t+s}^{ST}\right\}$. Stationarity implies both β_t^{ST} and p_t^{ST} are constants; intuitively, each principal will solve the same problem, giving rise to the constant equilibrium outcome in Proposition 2.[12]

[11] When the agent does not have any bargaining power, the proof in Proposition 2 shows that for the agent's problem, the short-term incentives $\{\beta_t\}$ here play the same role as the incentives $\{\beta_t\}$ in long-term contracts analyzed in Proposition 1. For more explanation, see footnote 12.

[12] Holmstrom (1999) considers learning when contractual relations are short-term, and when learning is stationary the equilibrium outcome is constant over time. Comparing to our analysis, the substantial difference is that Holmstrom (1999) studies wage contracts without incentives, and principals are competitive

Thus, without commitment, in short-term contracting each principal at different points of time solves his individual myopic problem in equation (45). In contrast, with long-term contracting, a single principal not only maximizes the flow payoff in equation (45), but also takes into account the effect of β_{t+s} on the forward-looking information rent p_t.[13] The next two subsections show that, the full commitment in long-term contracting with learning makes the optimal effort policy time-decreasing and stochastic.

5.2 Optimal Deterministic Contract

Before we characterize the optimal contract fully, in order to gain better intuition we first constrain the incentive slopes $\{\beta\}$ to be deterministic. Within the class of deterministic contracts, we solve the optimal contract in closed form.

5.2.1 Time-decreasing optimal effort policy

When $\{\beta\}$ is deterministic, we can move the conditional expectation in equation (36) inside the integral. Because $\exp\left(-\int_t^s ar\beta_v\sigma dB_v^\mu - \frac{1}{2}\int_t^s a^2r^2\beta_v^2\sigma^2 dv\right)$ is a martingale, the information rent p_t is simply

$$p_t = \phi\int_t^\infty e^{-(\phi+r)(s-t)}\beta_s ds.$$

In other words, deterministic incentives imply deterministic information rents, i.e., the volatility σ^p is constrained to zero.

Denote by $V^d(p)$ the value function with deterministic policies. Plugging $\sigma^p = 0$ into equation (41), we have $\beta^d(p) = \left(1 + p - \phi V_p^d\right)/(1 + ar\sigma^2)$, with the HJB equation as:

$$rV^d(p) = \frac{1}{2}\frac{\left(1 + p - \phi V^d(p)\right)^2}{1 + ar\sigma^2} - p - \frac{1}{2}p^2 + V_p^d(p)(\phi + r)p.$$

(so the agent has all the bargaining power). Thus, every period competitive principals always pay the agent some fixed wage equal to his/her expected ability, and agents work for reputation concerns.

Even if we introduce short-term incentives in Holmstrom (1999), where reputation concern is present with competitive principals (or, the agent has all the bargaining power), our stationary setting should again lead to constant equilibrium incentives/effort profiles. Our assumption of principals having all the bargaining power eliminates this reputation concern, as the agent always gets zero continuation payoff for future relationships. As a result, we only have the belief manipulation effect. Further, this implies that the current short-term contract is the only source of variation in the agent's continuation value, and thus the short-term incentive β_t here has the same interpretation of β_t in that of long-term contracting. This allows us to highlight the role played by commitment.

[13] This result is in contrast to Fudenberg, Holmstrom, and Milgrom (1990), who show that with moral hazard, under CARA preferences the optimal long-term contract can be implemented by short-term ones. In a way, their result suggests that commitment itself is not that important. In contrast, our model shows that the commitment in long-term contracting is important because of the long-lasting belief manipulation effect with endogenous learning.

We solve this ODE in closed-form:

$$V^d(p) = -\frac{1}{2}A^d p^2 + B^d p, \tag{46}$$

where A^d and B^d are constants so that $B^d = 1/\phi$, and

$$A^d = \frac{1}{2\phi^2}\left((2\phi + r)\, ar\sigma^2 + r + \sqrt{(2\phi + r)^2 a^2 r^2 \sigma^4 + 2ar\sigma^2\left[(\phi + r)^2 + \phi^2\right] + r^2}\right) > 0.$$

We have the following proposition.

Proposition 3 *Within the class of deterministic policies, the value function for the optimal contract is given by equation (46), and it is optimal to set time-0 information rent to be*

$$p_0^d = \overline{p}^d \equiv \frac{B^d}{A^d} > 0,$$

where $\overline{p}^d$ *satisfies* $V_p^d\left(\overline{p}^d\right) = 0$. *The evolution of information rent* p_t *is given by*

$$p_t^d = \overline{p}^d \exp(-\lambda t), \text{ with } \lambda \equiv -\phi - r + \frac{1 + A^d\phi}{1 + ar\sigma^2}\phi > 0.$$

In the deterministic optimal contract, the incentive is given by

$$\beta_t^d(p) = \frac{1 + A^d\phi}{1 + ar\sigma^2}p_t^d = \frac{1 + A^d\phi}{1 + ar\sigma^2}\overline{p}^d \exp(-\lambda t),$$

and the optimal effort policy is

$$\mu_t^d = \beta_t^d - p_t^d = \frac{A^d\phi - ar\sigma^2}{1 + ar\sigma^2}\overline{p}^d \exp(-\lambda t). \tag{47}$$

Proof. *See Appendix A.* ■

The above proposition shows that in the optimal deterministic contract, the information rent p_t^d, the incentive β_t^d, and the optimal effort μ_t^d all follow certain exponentially decaying paths (toward zero), with the same rate of $-\lambda$. Moreover, at $t = 0$, from equation (47) we have

$$\mu_0^d = \frac{A^d\phi - ar\sigma^2}{1 + ar\sigma^2}p_0^d = \frac{1 - ar\sigma^2\overline{p}^d}{1 + ar\sigma^2} < \frac{1}{1 + ar\sigma^2} = \mu_0^{HM}.$$

Thus, the initial optimal effort level is below the observable profitability benchmark, and the optimal effort path is decreasing over time.

The optimality of the front-loaded effort policies comes from the forward looking nature of information rents. From the agent's incentive-compatibility condition in equation (22), the belief manipulation effect implies that giving incentives later tends to make the agent shirk earlier, but not the other way around. This implies that later incentives are more

costly than early ones, and, consequently, the optimal contract implements higher effort in earlier periods. Clearly, this result relies on the commitment ability in long-term contracting: Section 5.1.2 shows that when relationships are short-term so that each principal does not take the forward-looking information rent into account, the resulting incentives and effort policies are constant over time.

5.2.2 Discussion

Information rents and front-loaded effort policy Both Prat and Jovanovic (2012) and our model find front-loaded incentives to be optimal. Because Prat and Jovanovic (2012) implement a constant effort,[14] the forward looking nature of information rents implies that the compensation contract has to offer front-loaded incentives. Our model allows the optimal contract to adjust on the effort margin (not just incentives), and cheaper incentive provisions in earlier periods naturally push the optimal contract to implement a front-loaded effort profile.

The front-loaded incentive/effort policies also arise in models with career concerns (e.g., Gibbons and Murphy (1992), Holmstrom (1999)), but through a distinct mechanism. There, agents in their early careers face *higher uncertainty* in their abilities, and, thus, work harder to impress the market (but the market will not be fooled in equilibrium, a standard signal jamming problem). This force is not present in our stationary model, as the uncertainty of the profitability/ability (i.e., the posterior variance of θ_t) stays constant over time.

Perhaps more interestingly, Garrett and Pavan (2012) study a dynamic contracting problem with *pre-contracting* asymmetric information. Under the assumption that the effect of an agent's initial productivity on his future productivity is declining over time, they find that the optimal effort policy is time-increasing. This result is opposite to the front-loaded effort policy derived in our paper without pre-contracting private information (but with learning). Intuitively, in their setting, facing information asymmetry, the equilibrium effort is downward distorted for rent extraction. The downward distortion resulting from information asymmetry is more severe in earlier periods. Compared to our paper where *post-contracting* private information is always off-equilibrium, in Garrett and Pavan (2012) the *pre-contracting* private information occurs on-equilibrium. It is intriguing that these two types of private information have opposite predictions on the time-series pattern of distortion. On the other hand, the difference also lies on the agent being risk-neutral without wealth constraint in Garrett and Pavan (2012).[15] Relatedly, Sannikov (2012) allows for the possibility of the

[14] Prat and Jovanovic (2012) assume that the effort cost is linear over the feasible interval $[0, 1]$ and focus on implementing the highest effort level 1. A similar structure is assumed in DeMarzo and Sannikov (2012).

[15] In Garrett and Pavan (2012), since the agent is risk neutral and has enough wealth (or, no limited liability), without pre-contracting private information the principal can sell the project to the risk-neutral

agent's current effort to affect future fundamentals, and finds that the optimal effort policy is increasing over time as well. We leave future research for further exploration on this topic.

What if learning is non-stationary? For tractability reasons we assume stationary learning in our model as in DeMarzo and Sannikov (2012). A non-stationary learning setting is also widely adopted when economic agents face parameter uncertainty (e.g., Prat and Jovanovic, 2012): the underlying profitability θ (as a parameter) never changes, and as time passes both parties get to learn the true profitability eventually. Does the front-loaded feature of the optimal effort profile depend on whether or not learning is stationary?

When learning is non-stationary, the posterior updating rules corresponding to equations (3) and (4) are

$$dm_t = \Sigma_t^\theta \frac{dY_t - (\mu_t + m_t)\,dt}{\sigma^2} \equiv \frac{\Sigma_t^\theta}{\sigma} dB_t^\mu, \text{ and } \Sigma_t^\theta = \frac{\sigma^2 \Sigma_0^\theta}{\sigma^2 + \Sigma_0^\theta t}, \tag{48}$$

The most salient difference lies in the conditional variance Σ_t^θ. In the stationary case that we are studying, the posterior variance in equation (4) is constant over time, while in the stationary case, the posterior variance decreases over time and vanishes asymptotically.

If learning is non-stationary, on equilibrium path eventually everyone learns θ perfectly, which implies that over time there are less and less noises in output dY_t. This is important, as optimal contracting is essentially an inference problem (Holmstrom, 1979). When time goes by, the signal-to-noise ratio of dY_t goes up, which implies that *information quality* goes up in designing the optimal contract. In other words, this information quality effect potentially push the optimal effort policy to be back-loaded (i.e., getting closer to the first-best level) to later periods with less uncertainty. This is an opposite force against the information rent effect which favors front-loaded effort policies.[16] Which force is stronger, the information rent effect or the information quality effect? We have the following analytical result.

Proposition 4 *Restrict attention to the deterministic policies. With non-stationary learning so that the posterior variance follows equation (48), the optimal effort policy is decreasing over time.*

Proof. *See Appendix B.* ■

agent to achieve the first-best outcome, and there is no downward distrotion at all. In our model, it is efficient for the principal to offer insurance to the risk-averse agent. We expect the front-loaded pattern of optimal effort policy to be attenuated if we allow for the agent's pre-contracting private information, because of the effect in Garrett and Pavan (2012).

[16] Although both parties learn profitability θ perfectly asymptotically so that we are in Holmstrom and Milgrom (1987), the optimal contract might not implement the level implied by Holmstrom and Milgrom (1987) in the distant future, simply because it could leave too much information rents to the agent in earlier periods.

Thus, under deterministic policies, even with non-stationary learning the optimal effort policy is still decreasing with time, suggesting that the information rent effect dominates the information quality effect. We leave future research to investigate the general optimal contract with non-stationary learning.

5.3 Optimal Contract

5.3.1 Characterizing the value function

We now solve the optimal contract without the restriction of incentives $\{\beta_t\}$ being deterministic. Based on the technique of dynamic programming, we analyze the ODE in equation (43) with the boundary condition in equation (37), which gives the solution to the constrained problem (39).

In order to ensure the concavity of $V(p)$, we impose the following parametric condition throughout:

$$\frac{\phi}{r} > a\left[2\left(\frac{\phi}{r}+1\right)^3 - \phi\sigma^2\right]. \tag{49}$$

The following proposition characterizes the properties of the value function under the optimal policy $\{\beta^*, \sigma^{p,*}\}$:

Proposition 5 *When M is sufficiently large, we have the following properties for $V(p;M) \in \mathbb{C}^2$.*

1. *$V(0;M) = 0$ and $V_p(0;M) = 1/\phi$.*
2. *$V(p;M)$ is strictly concave over the interval $\left[-\frac{\phi M}{\phi+r}, \frac{\phi M}{\phi+r}\right]$, and $1 + ar\sigma^2 + a^2r^2\sigma^2\frac{(V_p)^2}{V_{pp}} > 0$.*
3. *There exists $\overline{p} \in \left(0, \overline{p}^d\right)$ such that $V_p(\overline{p};M) = 0$. Under the optimal policy, $\overline{p}$ is an upper entrance-no-exit boundary, and 0 is a lower absorbing boundary 0 with $V(0;M) = 0$. This implies that under the optimal policy the endogenous state variable p never exits the interval $[0, \overline{p}]$.*
4. *The endogenous upper boundary $\overline{p}$ is independent of M.*
5. *Thus, when M is sufficiently large, the value function $V(p;M) = V(p)$ and associated optimal policies $\{\beta^*, \sigma^{p,*}\}$ are independent of M.*

Proof. *See Appendix A.* ■

In the optimal contract, at date 0 the principal sets $p_0^* = \overline{p}$. Afterwards the information rent p_t^* as the state variable evolves according to equation (38). The optimal control is characterized by equation (42). Interestingly, property 3 in Proposition 5 states that the information rent p_t^* will never wander out of an endogenous interval $[0, \overline{p}]$. Intuitively, because the information rent p_t^* is the (properly) discounted promised future incentives as in equation (36), it is suboptimal to promise too much or too little future incentives in optimal contracting. This result is reminiscent of Holmstrom and Milgrom (1987): without learning, the problem is essentially isolated each period, and the optimal incentives $\{\beta_t^*\}$ remain constant. In our model with learning, incentive provisions are connected across periods, and the optimal incentives $\{\beta_t^*\}$ become stochastic. Nevertheless, since the model primitives are stationary (CARA-Normal setting, additive technology in equation (31), and stationary learning), the (properly) discounted information rents $\{p_t^*\}$ remain endogenously bounded in the optimal contract.

The next theorem shows that for sufficiently large M the value achieved in the relaxed problem is the same as the value achieved in the original problem. Hence, the solution to the relaxed problem is also the solution to the original problem, which is the optimal contract that we are after.

Theorem 1 *When the constant M is sufficiently large, the solution to the relaxed problem in equation (39) is the solution to the original problem in equation (35). Hence the solution in Proposition 5 characterizes the optimal contract.*
Proof. *See Appendix A.* ■

5.4 Implications of optimal contracting

We study the implication of the optimal contract now. From now on we always refer to the optimal policies, without risk of confusion we omit the superscript "*".

5.4.1 Value function and optimal policies

Figure 1 plots the value function $V(p)$, the optimal control $\{\beta(p), \sigma^p(p)\}$, and the associated optimal policy $\mu_t(p) = \beta_t(p) - p$ in solid lines. For comparison, in each panel we also plot the corresponding deterministic counterparts in dashed lines, and the Holmstrom and Milgrom (1987) benchmark in dotted lines.

The value delivered by the optimal stochastic contract must exceed the one under the deterministic counterpart, as shown in Panel A in Figure 1. Panel B plots the volatility of the agent's information rent, σ^p, which is zero by definition when the contract is restricted to be deterministic. A positive σ^p in the optimal stochastic contract implies that the information

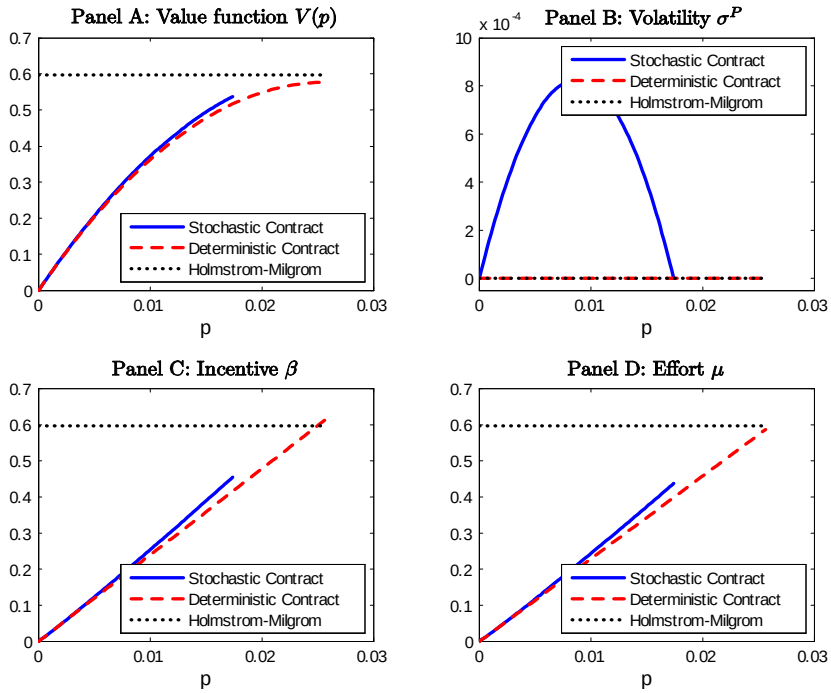

Figure 1: Value function and optimal policies in the optimal contract. Solid lines correspond to the optimal stochastic contract, dashed lines correspond to the optimal deterministic contract, and dotted lines correspond to the Holmstrom-Milgrom (1987) benchmark. The parameters are $r = 0.5, a = 0.6, \sigma = 1.5, \phi = 0.022$.

rent rises after good performance shocks, an interesting property that will be discussed shortly.

What drives the stochastic contract to be superior to the deterministic one? It is because the stochastic contract implements a more efficient effort policy, closer to the higher Holmstrom and Milgrom (1987) effort benchmark. In Panel C, we show that the incentive $\beta(p)$ sits above the deterministic counterpart for almost the entire range (and, thus, gets closer to the Holmstrom and Milgrom (1987) benchmark level), except for low p's that are close to zero. A similar pattern holds for the implemented effort $\mu(p) = \beta(p) - p$ in Panel D of Figure 1.

Interestingly, though not evident in Figure 1, when p is close to zero both the incentive slope $\beta(p)$ and the effort $\mu(p)$ drop below their deterministic counterparts. This result is a robust feature of the model. Indeed, with the aid of asymptotic analysis in equation (97) in Appendix C, one can analytically verify that the adjustment term between the deterministic and stochastic contracts is negative by setting $p \simeq 0$. The seemingly counter-intuitive result is rooted in the "option-like" feature in the optimal contract, to which we turn next.

5.4.2 Option-like incentives in optimal contracting

In our model it is optimal to implement a history dependent effort policy. This is surprising: because the posterior variance only changes over time deterministically with a standard CARA-Normal setting and learning (in our stationary setting, it is a constant in particular), usually the resulting equilibrium effort profile is a deterministic process as well (e.g., Holmstrom, 1999).

To understand the economic mechanism that drives this result, we study how the history-dependent effort policies improve over deterministic policies. To this end, we investigate the response of pay-performance sensitivity β and the implemented effort μ to unexpected shocks, which is captured by the diffusion term of $d\beta(p_t)$ (or, $d\mu(p_t)$), i.e., $\beta'(p_t)\sigma^p dB_t$ (or, $(\beta'(p_t)-1)\sigma^p dB_t$). As shown in the top panels in Figure 2, these diffusion terms are positive. Thus, in contrast to the Holmstrom and Milgrom (1987) benchmark where the optimal contract features a constant equity share, with learning the optimal contract has an option-like feature. More specifically, pay-for-performance rises following good performance, suggesting that the optimal contract is "convex" in output. The same pattern holds for the optimal effort policy.

The optimality of this option-like feature is a result of reducing the agent's information rent in a long-term relation. As explained in Section 3.3, the thrust of endogenous learning in dynamic contracting is that the agent can manipulate the principal's future belief downward by shirking today, and thus enjoy the potential information rent:

$$p_t = \frac{1}{u_c(c_t,\mu_t)}\mathbb{E}_t\left[\int_t^\infty \phi e^{-(\phi+r)(s-t)}\beta_s u_c(c_s,\mu_s)\,ds\right].$$

The information rent captures those additional future rewards to the agent when the principal mistakenly attributes the higher-profitability-driven good performance to the agent's effort. This explains why future pay-performance sensitivities $\{\beta_s\}$ enter the agent's information rent. Equally important, for a risk-averse agent, the amount of information rents also depends on his marginal utilities $u_c(c_s,\mu_s)$ when receiving manipulation benefits at future states. Following a positive output shock, the agent becomes wealthier, which drives down his marginal utility $u_{c,s} = -arv_s$.[17] Then, by making the optimal contract option-like, the principal raises incentives after good performance and, thus, imposes a negative correlation between pay-performance-sensitivity and marginal utility. As a result, by allocating greater belief manipulation benefits in states where the agent cares less, the option-like feature reduces the agent's information rent standing today.

[17] Formally, we have the evolution of marginal utility to be $d(-arv_t) = -ar\beta_t(-arv_t)\,dB_t$, which has a negative diffusion on the performance shock.

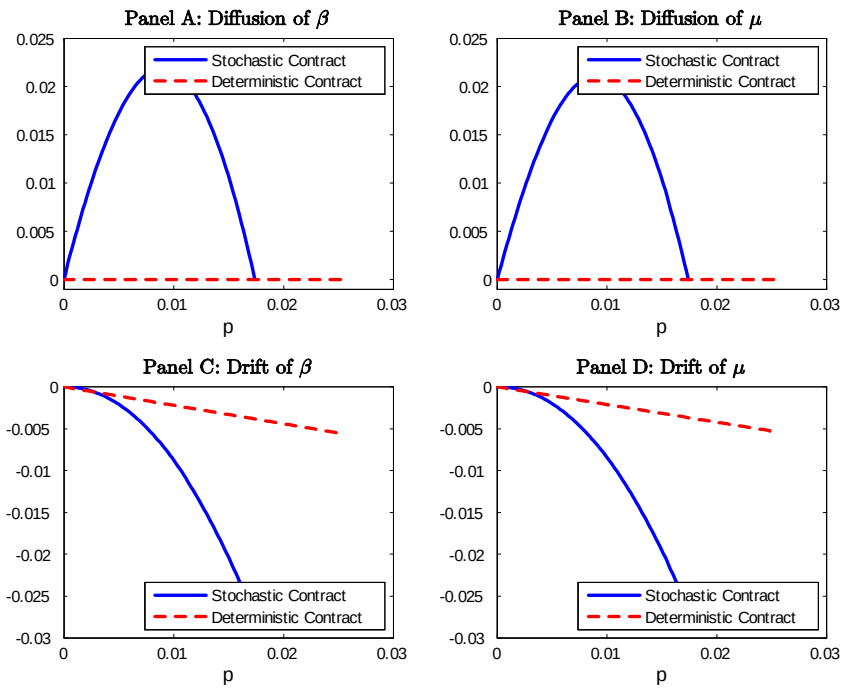

Figure 2: Diffusion and drift for incentives and effort in the optimal contract. Solid lines correspond to the optimal stochastic contract, and dashed lines correspond to the optimal deterministic contract. The parameters are $r = 0.5, a = 0.6, \sigma = 1.5, \phi = 0.022$.

Now let us get back to the seemingly counter-intuitive result in the bottom-left panel of Figure 1, i.e., when p is close to zero the optimal stochastic contract implements a lower effort than the deterministic one. Here is the intuition. A positive diffusion of incentive β (effort μ) implies that the optimal contract allocates lower incentives in the states with poor historical performance (and hence high marginal utility). Because the information rent is positively correlated with performance as indicated by Panel B in Figure 2, the stochastic optimal contract imposes lower incentives in the states where p gets close to zero.

5.4.3 Time-decreasing effort policies

We also plot the drift of incentives β and effort μ in two bottom panels in Figure 2. We have shown analytically in Section 5.2.1 that, because of the forward-looking information rent, the effort policy is decreasing over time in the optimal deterministic contract. Not surprisingly, this pattern persists in the fully stochastic optimal contract. Graphically, the front-loaded effort policy is reflected by the negative drifts (dotted lines for deterministic contracts) in the bottom panels in Figure 2.

6 Conclusion

We introduce profitability uncertainty into Holmstrom and Milgrom (1987) and study the optimal long-term contracting with endogenous learning. Although along the equilibrium path the principal and the agent hold the same belief about the project profitability, the agent's potential deviation by exerting effort below the recommended level leads to potential long-lasting belief divergence between both parties, and thus, a "hidden information" problem. Utilizing the convenient property of CARA preference, we show that the optimal contracting can be reformulated to a dynamic programming problem with only one state variable, and characterize the optimal contract by the solution to an ODE. We show that the optimal effort decreases with tenure on average, and the optimal contract exhibits an option-like feature in the sense that incentives/effort rise after positive performance shocks. These two properties relies on the combination of learning and long-term contracting, as we show the resulting effort profile is constant over time if we have either short-term contracting with learning, or long-term contracting without learning (Holmstrom and Milgrom, 1987).

Although we are only able to give a full characterization of the optimal long-term contract with learning under a specific setting (e.g., CARA preference, Gaussian processes, etc), the above two key qualitative results, i.e., front-loaded effort policy and option-like feature, will likely be robust to more general settings. The main reason we think the results extend is that the economic force behind these results do not depend on the CARA preference or Gaussian processes. The agent's information rent due to belief manipulation, i.e., the agent's inclination to shirk to distort the principal's future belief downward, is general in any long-term contracting environment with learning. The higher costs of providing incentives later leads to the time-decreasing effort policy, simply because later incentives enter the agent's forward-looking information rent in earlier periods (but not the other way around). Additionally, the option-like feature comes from the fact that the agent is risk-averse so that the marginal value of in earning belief manipulation benefits is lower after good performance.

References

Adrian, Tobias, and Mark M. Westerfield, 2009, Disagreement and learning in a dynamic contracting model, *Review of Financial Studies* 22-10, 3873–3906.

Baron, D. P., and D. Besanko, 1984, Regulation and Information in a Continuing Relationship, *Information Economics and Policy* 1, 267–302.

Bergemann, D. and U. Hege, 1998, Venture Capital Financing, Moral Hazard, and Learning, *Journal of Banking & Finance* 22, 703–735.

Bergemann, D. and U. Hege, 2005, The financing of innovation: Learning and stopping, *RAND Journal of Economics* 36, 719–752.

Cosimano, Thomas., Adam Speight, and Hayong Yun, 2011, Executive compensation in a changing environment: cautious investors and sticky contracts, working paper, University of Notre Dame.

Cvitanic, Jakša, Xuhu Wan, and Huali Yang, 2012, Dynamics of contract design with screening, forthcoming in *Management Science*.

DeMarzo, Peter M., and Yuliy Sannikov, 2012, Learning, termination and payout policy in dynamic incentive contracts, working paper, Stanford University and Princeton University.

Dittmann, Ingold and Ernst Maug, 2007, Lower salaries and no options? On the optimal structure of executive pay, *Journal of Finance* 62, 303–343.

Fleming, W., and R. Rishel, 1975, *Deterministic and stochastic optimal control*, Springer-Verlag, New York.

Fudenberg, D., Holmstrom, B., Milgrom, P., 1990. Short-term contracts and long-term agency relationships. *Journal of Economic Theory* 51, 1–31.

Garrett, Daniel F., and Alessandro Pavan, 2012, Managerial turnover in a changing world, *Journal of Political Economy* 120-5, 879–925.

Gershkov, A. and M. Perry, 2012, Dynamic Contracts with Moral Hazard and Adverse Selection, *Review of Economic Studies* 79, 268–306.

Gibbons, Robert and Kevin J. Murphy, 1992, Optimal incentive contracts in the presence of career concerns: theory and evidence." *Journal of Political Economy* 100(3): 468–505.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski, 2012, Optimal dynamic taxes, working paper, Princeton, Yale, and Minnesota.

Halac, Marina, and Navin Kartik, Qingmin Liu, 2012, Optimal contracts for experimentation, working paper, Columbia University.

He, Zhiguo, 2011, A model of dynamic compensation and capital structure, *Journal of Financial Economics* 100, 351–366.

He, Zhiguo, 2012, Dynamic compensation contracts with private savings, *Review of Financial Studies* 25, 1494-1549.

Holmstrom, Bengt, 1999, Managerial incentive problems: a dynamic perspective, *Review of Economic Studies* 66, 169–182.

Holmstrom, Bengt, and Paul Milgrom, 1987, Aggregation and linearity in the provision of intertemporal incentives, *Econometrica* 55-2, 303–328.

Liptser, Robert S., and Albert N. Shiryaev, 1977, *Statistics of random processes II: Applications*, Volume 6 in Applications of Mathematics, Springer-Verlag, New York.

Horner, J. and L. Samuleson, 2012, Incentives for experimenting agents, working paper, Yale Univesrity.

Pavan Alessandro, Ilya Segal, and Juuso Toikka, 2012, Dynamic mechanism design, working paper, Northwestern, Stanford, and MIT.

Prat, Julien, and Boyan Jovanovic, 2012, Dynamic contracts when agent's quality is unknown, working paper, NYU.

Sannikov, Yuliy, 2007, Agency problems, screening and increasing credit lines, working paper, Princeton University.

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

Sannikov, Yuliy, 2012, Moral hazard and long-run incentives, working paper, Princeton University.

Spear, Stephen E., and Sanjay Srivastava, 1987, On repeated moral hazard with discounting, *Review of Economic Studies* 54-4, 599–617.

Sung, Jaeyoung, 2005, Optimal contracts under adverse selection and moral hazard: A continuous-time approach, *Review of Financial Studies* 18-3, 1021–1073.

Williams, Noah, 2009, On dynamic principal-agent problems in continuous time, working paper, University of Wisconsin-Madison.

Williams, Noah, 2011, Persistent private information, *Econometrica* 79-4, 1233–1274.

Zhang, Yuzhe, 2009, Dynamic contracting with persistent shocks, *Journal of Economic Theory* 144, 635–675.

A Appendix A: Proofs

A.1 Proof for Lemma 1

The argument is similar to He (2011). Consider any contract $\Pi = \{c_t, \mu_t\}$ which induces an optimal policy $\{c_t^\star, \mu_t^\star\}$ from the agent with a value $v_0^\star$. The principal knows the resulting optimal effort policy $\{\mu_t^\star\}$ and she updates her belief according to $\{\mu_t^\star\}$ rather than the recommended effort policy $\{\mu_t\}$. From the agent's budget equation, we have

$$S_t = \int_0^t e^{r(t-s)} \left(c_s - c_s^\star\right) ds,$$

which gives the agent's optimal savings path, with the transversality condition $\lim_{T\to\infty} \mathbb{E}\left[e^{-rT} S_T\right] = 0$. Note that the transversality condition holds for all measures induced by any feasible effort policies.

By invoking the replication argument similar to revelation principle, we consider giving the agent a direct contract $\Pi^\star = \{c_t^\star, \mu_t^\star\}$; notice that 1) the principal will not change her Bayesian updating because she just changed her "recommended" effort policy which has no effect on the contract at all; and 2) the principal

has the same cost to deliver this contract as $\Pi = \{c_t, \mu_t\}$. Clearly, taking consumption-effort policy $\{c_t^\star, \mu_t^\star\}$ is feasible for the agent with no private-savings. Therefore, we only need to show that $\{c_t^\star, \mu_t^\star\}$ is indeed optimal for the agent, because if it is true, then the principal still correctly knows the agent's actual optimal effort policy $\{\mu_t^\star\}$, and her payoff is the same as that under the contract $\Pi = \{c_t, \mu_t\}$.

Suppose, counter-factually, that given the contract $\Pi^\star$, the agent finds that $\{c_t', \mu_t'\}$ yields a strictly higher payoff $v_0' > v_0^\star$ in her problem, with associated savings path

$$S_t' = \int_0^t e^{r(t-s)} \left(c_s^\star - c_s'\right) ds,$$

which satisfies the transversality condition. However, consider the following saving path

$$S_t'' = S_t + S_t' = \int_0^t e^{r(t-s)} \left(c_s - c_s'\right) ds,$$

which also satisfies the transversality condition. This, however, implies that given the original contract Π, the saving rule $\{S_t''\}$ supports $\{c_t', \mu_t'\}$ but delivering a strictly higher payoff v_0'. This contradicts with the optimality of $\{c_t^\star, \mu_t^\star\}$ under the contract Π.

A.2 Proof for Lemma 2

Fix any constant S. Given any savings $S_t = S$ and a contract $\Pi = \{c\}$, from time-t on the agent's problem is

$$\begin{aligned} \max_{\{\widehat{c}_s\},\{\widehat{\mu}_s\}} \quad & \mathbb{E}^{\widehat{\mu}}\left[\int_t^\infty -\frac{1}{a} e^{-a\left(\widehat{c}_s - \frac{1}{2}\widehat{\mu}_s^2\right) - r(s-t)} ds\right] \\ s.t. \quad & dS_s = rS_s ds + c_s ds - \widehat{c}_s ds, \ S_t = S, \ s > t \\ & dY_s = \left(\widehat{\mu}_t + m_t^{\widehat{\mu}}\right) dt + \sigma dB_s^{\widehat{\mu}}, \end{aligned} \tag{50}$$

given his information set. Note that the agent will learn actively. Denote by $\{c_s^\star, \mu_s^\star\}$ the solution to the above problem, and by $v_t\left(S; \Pi\right)$ the resulting agent's value.

Now consider the problem with $S = 0$, which is the continuation payoff along the equilibrium path:

$$\begin{aligned} \max_{\{\widehat{c}_s\},\{\widehat{a}_s\}} \quad & \mathbb{E}^{\widehat{\mu}}\left[\int_t^\infty -\frac{1}{a} e^{-a\left(\widehat{c}_s - \frac{1}{2}\widehat{\mu}_s^2\right) - r(s-t)} ds\right] \\ s.t. \quad & dS_s = rS_s ds + c_s ds - \widehat{c}_s ds, \ S_t = S, \ s > t \\ & dY_s = \left(\widehat{\mu}_t + m_t^{\widehat{\mu}}\right) dt + \sigma dB_s^{\widehat{\mu}}, \end{aligned}$$

We claim that the solution to this problem is $\{c_s^\star - rS, \mu_s^\star\}$, and therefore the value is $v_t\left(0; \Pi\right) = e^{arS} v_t\left(S; \Pi\right)$. There are two steps to show this. First, this solution is feasible. Second, suppose that there exists another policy $\left\{\widehat{c}_s', \widehat{\mu}_s'\right\}$ that is superior to $\{c_s^\star - rS, \mu_s^\star\}$, so that the associated value $v_t'\left(0; \Pi\right) > e^{-arS} v_t\left(S; \Pi\right)$. Consider $\left\{\widehat{c}_s' + rS, \widehat{\mu}_s'\right\}$, which is feasible to the problem in equation (50). Under this plan, however, the agent's objective is

$$e^{-arS} \cdot \max \mathbb{E}_t^{\widehat{\mu}}\left[\int_t^\infty -\frac{1}{a} e^{-\gamma\left(\widehat{c}_s' - \frac{1}{2}\widehat{\mu}_s'^2\right) - r(s-t)} ds\right] = e^{-arS} v_t'\left(0; \Pi\right) > v_t\left(S; \Pi\right),$$

which contradicts with the optimality of $\{c_s^\star, \mu_s^\star\}$. As a result, $v_t\left(S; \Pi\right) = e^{-arS} v_t\left(0; \Pi\right)$.

A.3 Proof for Proposition 1

We apply the Pontryagin's Maximum Principle (MP). For this reason, we first consider the following agent's problem in finite horizon with terminal date T. We then will let T tend to infinity and impose transversality conditions. To preserve stationarity, we assume that the agent's utility function is

$$U(C_T) = -\frac{1}{ar}\exp(-arC_T) \tag{51}$$

which is as if the infinitely-lived agent is retired at T and consume rC_T afterwards. Clearly this treatment is innocuous as we take T to infinity.

Step 1. Describing the optimal policy

We take a guess-and-verify approach. As suggested by the main text, consider the *proposed* policy $\{c^\star, \mu^\star\}$ with the following properties. First, define the agent's continuation payoff as

$$v_t^\star \equiv \mathbb{E}_t^{\mu^\star}\left[\int_t^T e^{-r(s-t)}\left(u\left(c_s^\star, \mu_s^\star\right)\right)ds + e^{-r(T-t)}U\left(C_T^\star\right)\right].$$

Then the consumption $c_t^\star$ satisfies

$$u\left(c_t^\star, \mu_t^\star\right) = rv_t^\star \text{ so that } u_c\left(c_t^\star, \mu_t^\star\right) = -arv_t^\star. \tag{52}$$

And, according to the martingale representation theorem, there exists a progressive process $\{\beta^\star\}$ such that

$$dv_t^\star = rv_t^\star - u\left(c_t^\star, \mu_t^\star\right) + \beta_t^\star\left(-arv_t\right)dB_t^{\mu^\star} = \beta_t^\star\left(-arv_t\right)dB_t^{\mu^\star}. \tag{53}$$

Given $\beta^\star$, we can rewrite the total reward from the optimal policy in its martingale representation form:

$$I_T^{c^\star,\mu^\star} = \int_0^T e^{-rt}u\left(c_t^\star, \mu_t^\star\right)dt + e^{-rT}U\left(C_T^\star\right) = v_0^\star + \int_0^T e^{-rt}\beta_t^\star\left(-arv_t^\star\right)\sigma dB_t^{\mu^\star}. \tag{54}$$

Then the effort $\mu_t^\star$ satisfies

$$\mu_t^\star = \beta_t^\star - \mathbb{E}_t^{\mu^\star}\left[\int_t^T e^{-(r+\phi)(s-t)}\beta_s^\star\frac{v_s^\star}{v_t^\star}ds\right]. \tag{55}$$

The rest of the steps attempt to show that the above proposed policy is optimal.

Step 2. Rewriting the objective function.

Consider any alternative policy $\{c, \mu\}$, and define the associated reward process

$$I_t^{c,\mu} \equiv \mathbb{E}_t^{\mu}\left[\int_0^T e^{-rs}u\left(c_s, \mu_s\right)ds + e^{-rT}U\left(C_T\right)\right] = \int_0^t e^{-rs}u\left(c_s, \mu_s\right)ds + e^{-rt}v_t,$$

where as before $v_t = \mathbb{E}_t^{\mu}\left[\int_t^T e^{-r(s-t)}u\left(c_s, \mu_s\right)ds + e^{-r(T-t)}U\left(C_T\right)\right]$. Define the local effort deviation by $\delta_t \equiv \mu_t - \mu_t^\star$ and recall that $\Delta_t = m_t^\mu - m_t^{\mu^\star} = \phi\int_0^t e^{\phi(s-t)}\left(\mu_s^\star - \mu_s\right)ds$. Let $A_t \equiv \int_0^t e^{\phi(s-t)}\mu_s ds$, $A_t^\star \equiv \int_0^t e^{\phi(s-t)}\mu_s^\star ds$, and $\Delta_t^A \equiv A_t - A_t^\star = \frac{1}{\phi}\Delta_t$. Then from equations (7) and (5) we know that

$$\sigma dB_t^{\mu^\star} = \sigma dB_t^{\mu} + \left[\delta_t - \phi\Delta_t^A\right]dt. \tag{56}$$

Therefore, the total reward from an arbitrary policy $\{c, \mu\}$ is given by

$$\begin{aligned} I_T^{c,\mu} &= I_T^{c,\mu} - I_T^{c^\star,\mu^\star} + I_T^{c^\star,\mu^\star} = \int_0^T e^{-rt}\left[u\left(c_t, \mu_t\right) - u\left(c_t^\star, \mu_t^\star\right)\right]dt + e^{-rT}\left[U\left(C_T\right) - U\left(C_T^\star\right)\right] + I_T^{c^\star,\mu^\star} \\ &= \int_0^T e^{-rt}\left[u\left(c_t, \mu_t\right) - u\left(c_t^\star, \mu_t^\star\right)\right]dt + e^{-rT}\left[U\left(C_T\right) - U\left(C_T^\star\right)\right] \\ &\quad + v_0^\star + \int_0^T e^{-rt}\beta_t^\star\left(-arv_t^\star\right)\left[\delta_t - \phi\Delta_t^A\right]dt + \int_0^T e^{-rt}\beta_t^\star\left(-arv_t^\star\right)\sigma dB_t^{\mu}, \end{aligned} \tag{57}$$

where in the last equality we have used equations (54) and (56).

Step 3. Pontryagin's Maximum Principle (MP)

We maximize $I_0^{c,\mu} - v_0^\star$ which is equivalent to maximizing $I_0^{c,\mu}$ directly. Denote the agent's effort deviation by $\delta_t = \mu_t - \mu_t^\star$ as our (effective) effort control. The cumulative belief distance under any policy δ is $d\Delta_t^A = \left(\delta_t - \phi\Delta_t^A\right) dt$. In light of equation (57), the agent's problem can be rewritten as

$$\max_{c,\delta} \quad \mathbb{E}_0^{\delta+\mu^\star} \left\{ \begin{array}{c} \int_0^T e^{-rt}\left[u\left(c_t, \delta_t + \mu_t^\star\right) - u\left(c_t^\star, \mu_t^\star\right) dt\right] + e^{-rT}\left[U\left(C_T^\star + S_T\right) - U\left(C_T^\star\right)\right] dt \\ + \int_0^T e^{-rt}\beta_t^\star\left(-arv_t^\star\right)\left[\delta_t - \phi\Delta_t^A\right] dt \end{array} \right\} \tag{58}$$

$$s.t. \;\; \Delta_t^A = \int_0^t e^{\phi(s-t)}\delta_s ds, \; dS_t = \left(rS_t + c_t^\star - c_t\right) dt.$$

Here, the controls are consumption c_t and effort deviation $\delta_t = \mu_t - \mu_t^\star$, and the state variable vector is $x_t \equiv \left(\Delta_t^A, S_t\right)^T$, i.e., the belief divergence and private savings, with evolution of

$$d\begin{bmatrix} \Delta_t^A \\ S_t \end{bmatrix} = \begin{bmatrix} \delta_t - \phi\Delta_t^A \\ rS_t + c_t^\star - c_t \end{bmatrix} dt.$$

Using notation in Yong and Zhou (1999),

$$\begin{aligned} x &= \begin{pmatrix} \Delta_t^A \\ S_t \end{pmatrix}, \; b\left(t,x,u\right) = \begin{bmatrix} \delta_t - \phi\Delta_t^A \\ rS_t + c_t^\star - c_t \end{bmatrix}, b_x = \begin{bmatrix} -\phi & 0 \\ 0 & r \end{bmatrix} \\ f\left(t,x,u\right) &= -e^{-rt}\left(u\left(c_t, \delta_t + \mu_t^\star\right) - u\left(c_t^\star, \mu_t^\star\right)\right) + e^{-rt}\beta_t^\star\left(arv_t^\star\right)\left[\delta_t - \phi\Delta_t^A\right] \\ h\left(x\left(T\right)\right) &= -e^{-rT}\left[U\left(C_T^\star + S_T\right) - U\left(C_T^\star\right)\right]. \end{aligned}$$

Define the adjoint variables by $(y,z)^T$. Given the optimal trajectory $\left(\overline{\Delta}_t^A, \overline{S}_t\right)$ and the optimal control $\left(\overline{\delta}_t, \overline{c}_t\right)$, the evolution for adjoint variables and their terminal conditions are (where σ_t^y and σ_t^z are volatilities for (y_t, z_t) to be solved shortly)

$$d\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 0 & -r \end{bmatrix}\begin{bmatrix} y_t \\ z_t \end{bmatrix} dt + \begin{bmatrix} e^{-rt}\beta_t^\star\left(-arv_t^\star\right)\phi \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sigma_t^y \\ \sigma_t^z \end{bmatrix} dB_t^{\overline{\delta}+\mu^\star}, \tag{59}$$

$$\text{and} \begin{bmatrix} y_T \\ z_T \end{bmatrix} = \begin{bmatrix} 0 \\ e^{-rT}U'\left(C_T^\star + \overline{S}_T\right) \end{bmatrix}. \tag{60}$$

We write the Hamiltonian to be

$$\begin{aligned} H\left(t,\left(\delta,c\right),\left(y,z\right)\right) &= e^{-rt}\left[u\left(c_t, \delta_t + \mu_t^\star\right) - u\left(c_t^\star, \mu_t^\star\right) + \beta_t^\star\left(-arv_t^\star\right)\left[\delta_t - \phi\Delta_t^A\right]\right] \\ &\quad + \left(\delta_t - \phi\Delta_t^A\right) y_t + \left(rS_t + c_t^\star - c_t\right) z_t, \end{aligned} \tag{61}$$

and the first-order maximum conditions are:

$$e^{-rt}\left[u_\mu\left(\overline{c}_t, \overline{\delta}_t + \mu_t^\star\right) + \beta_t^\star\left(-arv_t^\star\right)\right] + y_t = 0, \text{ and } e^{-rt}u_c\left(\overline{c}_t, \overline{\mu}_t\right) - z_t = 0.$$

Now we verify the proposed policy $(\mu^\star, c^\star)$ satisfies the above conditions with $\overline{c}_t = c_t^\star$, $\overline{\delta}_t = 0$ (so that $\overline{\mu}_t = \mu_t^\star$), $\overline{\Delta}_t^A = 0$, and $\overline{S}_t = 0$. The backward stochastic differential equations (BSDE) for the adjoint state variables in equation (59) admit the following solutions:

$$y_t = ar\phi e^{\phi t}\mathbb{E}_t^{\mu^\star}\left[\int_t^T e^{-(r+\phi)s}\beta_s^\star v_s^\star ds\right], z_t = e^{-rt}u_c\left(c_t^\star, \mu_t^\star\right).$$

First we check y. Due to the Martingale Representaion Theorem, define the process σ_t^y so that

$$\mathbb{E}_t^{\mu^\star}\left[\int_0^T e^{-(r+\phi)s}\beta_s^\star v_s^\star ds\right] = \mathbb{E}_0^{\mu^\star}\left[\int_0^T e^{-(r+\phi)s}\beta_s^\star v_s^\star ds\right] + \int_0^t \frac{\sigma_s^y}{ar\phi e^{\phi s}} dB_s^{\mu^\star} \text{ for } t \in [0,T],$$

which implies that

$$\begin{aligned} dy_t &= \phi y_t dt + ar\phi e^{\phi t}\cdot d\left[\mathbb{E}_t^{\mu^\star}\left(\int_t^T e^{-(r+\phi)s}\beta_s^\star v_s^\star ds\right)\right] \\ &= \phi y_t dt + ar\phi e^{\phi t}\cdot d\left(\mathbb{E}_0^{\mu^\star}\left[\int_0^T e^{-(r+\phi)s}\beta_s^\star v_s^\star ds\right] + \int_0^t \frac{\sigma_s^y}{ar\phi e^{\phi s}} dB_s^{\mu^\star} - \int_0^t e^{-(r+\phi)s}\beta_s^\star v_s^\star ds\right) \\ &= \phi y_t dt - ar\phi e^{-rs}\beta_t^\star v_t^\star + \sigma_t^y dB_t^{\mu^\star}, \end{aligned}$$

which coincides with the BSDE of y in equation (59). For z, since $z_t = -e^{-rt}arv_t^\star$ from equation (52) and $dv_t^\star = \beta_t\left(-arv_t^\star\right)\sigma dB_t^{\mu^\star}$ from equation (53), we have

$$dz_t = -rz_t dt + e^{-rt}\beta_t a^2 r^2 v_t^\star \sigma dB_t^{\mu^\star},$$

hence $\sigma_t^z = e^{-rt}\beta_t a^2 r^2 v_t^\star \sigma$ in equation (59). Finally, from equation (51) we have $z_T = -e^{-rT}arv_T^\star = -e^{-rT}arU\left(C_T^\star\right) = e^{-rT}U'\left(C_T^\star\right)$ if we truncate the problem at T.

With the solutions of adjoint state variables (y,z), the necessary condition to maximize Hamiltonian in equation (61) is

$$u_\mu\left(c_t^\star,\mu_t^\star\right) + \beta_t^\star\left(-arv_t^\star\right) + e^{rt}y_t = 0, \text{ and } e^{-rt}u_\mu\left(c_t^\star,\mu_t^\star\right) = z_t,$$

which are exactly equations (52) and (55) under our proposed solution.

Step 4. Sufficiency

We now show the sufficiency of the first-order necessary condition for the optimality of agent's problem. It is easy to derive the gradient for Hamiltonian in equation (61) as

$$\begin{bmatrix} H_\Delta \\ H_S \\ H_\delta \\ H_c \end{bmatrix} = \begin{bmatrix} -\phi e^{-rt}\beta_t^\star\left(-arv_t^\star\right) - \phi y_t \\ rz_t \\ e^{-rt}\left[u_\mu\left(c_t,\delta_t+\mu_t^\star\right) + \beta_t^\star\left(-arv_t^\star\right)\right] + y_t \\ e^{-rt}u_c\left(c_t,\delta_t+\mu_t^\star\right) - z_t \end{bmatrix},$$

and the Hessian as

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-rt}u_{\mu\mu}\left(c_t,\delta_t+\mu_t^\star\right) & e^{-rt}u_{\mu c}\left(c_t,\delta_t+\mu_t^\star\right) \\ 0 & 0 & e^{-rt}u_{\mu c}\left(c_t,\delta_t+\mu_t^\star\right) & e^{-rt}u_{cc}\left(c_t,\delta_t+\mu_t^\star\right) \end{bmatrix}.$$

The Hessian matrix is negative semi-definite as $u_{\mu\mu} < 0$, $u_{cc} < 0$ and (note that $u_c = -au, u_\mu = a\mu u$)

$$u_{\mu\mu}u_{cc} - u_{\mu c}^2 = \left(a^2\mu^2 u + au\right)a^2 u - a^4\mu^2 u^2 = a^3 u^2 > 0$$

for all (c,μ), which implies Hamiltonian is globally concave. Sufficiency follows by applying Theorem 5.2 in Yong and Zhou (1999). QED.

A.4 Derivation of Expected Wage Cost in Section 4.1

We have (the last equality uses equation (33)):

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty -e^{-rt}\frac{1}{a}\ln\left(-arv_t\right)dt\right] &= \mathbb{E}\left[\int_0^\infty \frac{\ln\left(-arv_t\right)}{ar}d\left(e^{-rt}\right)\right] = -\frac{\ln\left(-arv_0\right)}{ar} - \mathbb{E}\left[\int_0^\infty e^{-rt}\frac{d\ln\left(-v_t\right)}{ar}\right] \\ &= -\frac{\ln\left(-arv_0\right)}{ar} + \mathbb{E}\left[\int_0^\infty e^{-rt}0.5ar\sigma^2\beta_t^2 dt\right].\end{aligned}$$

A.5 Proof for Lemma 3

Suppose that the control variable is constrained such that $\beta_t \in [-M, M]$, where $M > 0$ is an arbitrarily large, but fixed constant. Define $M_p \equiv \frac{\phi M}{\phi + r}$. Recall the definition of p_t, and we have

$$\begin{aligned} p &= \mathbb{E}\left[\int_0^\infty \phi\beta_t e^{-(\phi+r)t}e^{\left(-\int_0^t ar\beta_v\sigma dB_v - \frac{1}{2}\int_0^t a^2r^2\beta_v^2\sigma^2 dv\right)}dt\right] \\ &\leq \mathbb{E}\left[\int_0^\infty \phi M e^{-(\phi+r)t}e^{\left(-\int_0^t ar\beta_v\sigma dB_v - \frac{1}{2}\int_0^t a^2r^2\beta_v^2\sigma^2 dv\right)}dt\right] = \frac{\phi M}{\phi + r} = M_p,\end{aligned}$$

where the equality is obtained only if $\beta_t = M$ for all t. Thus, at any time the feasible state variable p is bounded. Similarly, we can show that $p \geq -\frac{\phi M}{\phi+r} = -M_p$. Moreover, this result implies that whenever $p_t = \pm M_p$, we must have that for $\forall s \geq t$,

$$\beta_s = \pm M, p_s = \pm M_p, \text{and } \mu_s = \beta_s - p_s = \pm\frac{rM}{\phi + r}.$$

Therefore, once p_t hits $\pm M_p$, the state p_s will stay there from then on. In this sense $\pm M_p$ are absorbing boundaries.

This result helps us in deriving the value function at these boundaries. Since

$$V(p) = \mathbb{E}_t\left[\int_t^\infty e^{-r(s-t)}\left(\mu_s - \frac{1}{2}\mu_s^2 - \frac{1}{2}a^2r^2\sigma^2\beta_s^2\right)ds\right],$$

we have

$$\begin{aligned} V(M_p) &= \int_0^\infty e^{-rt}\left(\frac{rM}{\phi + r} - \frac{1}{2}\left(\frac{rM}{\phi + r}\right)^2 - \frac{1}{2}a^2r^2\sigma^2M^2\right)ds \\ &= \frac{1}{\phi}M_p - \frac{r}{2\phi^2}M_p^2 - \frac{a^2r\sigma^2(\phi + r)^2}{2\phi^2}M_p^2,\end{aligned}$$

and similarly

$$V(-M_p) = -\frac{1}{\phi}M_p - \frac{r}{2\phi^2}M_p^2 - \frac{a^2r\sigma^2(\phi + r)^2}{2\phi^2}M_p^2.$$

Both of them are concave and quadratic in M. Finally, since M_p (M) can be arbitrarily large, $V(\pm M_p)$ can be arbitrarily small.

A.6 Derivation of Evolution of p_t in Section 4.2.1

Let

$$P_t \equiv v_0\mathbb{E}_t\left[\int_t^\infty \phi\beta_s e^{-(\phi+r)s}\exp\left(-\int_0^s ar\beta_u\sigma dB_u - \int_0^s \frac{1}{2}\left(ar\beta_u\sigma\right)^2 du\right)ds\right],$$

then $p_t = e^{(\phi+r)t}\frac{P_t}{v_t}$. According to martingale representation theorem, there exists some progressively measurable process Φ_t^P so that

$$dP_t = -\phi e^{-(\phi+r)t}\beta_t v_t dt + \Phi_t^P dB_t.$$

Recall that $dv_t = -ar\beta_t\sigma v_t dB_t$; then we have $d\left(\frac{1}{v_t}\right) = \frac{1}{v_t}\left(ar\beta_t\sigma dB_t + (ar\beta_t\sigma)^2 dt\right)$. Hence, according to Ito's lemma,

$$\begin{aligned} d\left(\frac{P_t}{v_t}\right) &= P_t d\left(\frac{1}{v_t}\right) + \frac{1}{v_t}dP_t + d\left(\frac{1}{v_t}, P_t\right) \\ &= \frac{P_t}{v_t}\left(ar\beta_t\sigma dB_t + (ar\beta_t\sigma)^2 dt\right) - \phi e^{-(\phi+r)t}\beta_t dt + \frac{\Phi_t^P}{v_t}dB_t + \frac{ar\beta_t\sigma}{v_t}\Phi_t^P dt. \end{aligned}$$

Therefore, we have

$$\begin{aligned} dp_t &= e^{(\phi+r)t}d\left(\frac{P_t}{v_t}\right) + d\left(e^{(\phi+r)t}\right)\frac{P_t}{v_t} \\ &= e^{(\phi+r)t}\left[(\phi+r)\frac{P_t}{v_t}dt + \frac{P_t}{v_t}\left(ar\beta_t\sigma dB_t + (ar\beta_t\sigma)^2 dt\right) - \phi e^{-(\phi+r)t}\beta_t dt + \frac{\Phi_t^P}{v_t}dB_t + \frac{ar\beta_t\sigma}{v_t}\Phi_t^P dt\right] \\ &= (\phi+r)p_t dt + p_t\left(ar\beta_t\sigma dB_t + (ar\beta_t\sigma)^2 dt\right) - \phi\beta_t dt + e^{(\phi+r)t}\frac{\Phi_t^P}{v_t}dB_t + ar\beta_t\sigma e^{(\phi+r)t}\frac{\Phi_t^P}{v_t}dt \\ &= \left[(\phi+r)p_t - \phi\beta_t + ar\beta_t\sigma\underbrace{\left(ar\beta_t\sigma p_t + e^{(\phi+r)t}\frac{\Phi_t^P}{v_t}\right)}_{\text{defined as } \sigma^p}\right]dt + \underbrace{\left(ar\beta_t\sigma p_t + e^{(\phi+r)t}\frac{\Phi_t^P}{v_t}\right)}_{\text{defined as } \sigma^p}dB_t \\ &= \left[(\phi+r)p_t + \beta_t\left(ar\sigma\sigma^p - \phi\right)\right]dt + \sigma^p dB_t, \end{aligned}$$

which is equation (38).

A.7 Proof for Proposition 2

Suppose along the equilibrium path the agent's continuation payoff is v_t. We would like to show similar to equation (21) we have

$$dv_t = -arv_t\beta_t\left(dY_t - \mu_t - m_t dt\right),$$

where β_t is the short-term incentive slope offered along the equilibrium path. Because the agent's future rents are always zero (principals have all the bargaining power), it is easy to show that under the optimal saving policy the private saving balance follows

$$S_t = -\frac{1}{ar}\ln\left(-arv_t\right),$$

with consumption policy

$$c_t = g\left(\mu_t\right) - \frac{1}{a}\ln\left(-arv_t\right).$$

Because the principal has all the bargaining power, the fixed wage α_t satisfies

$$\alpha_t = g\left(\widehat{\mu}_t\right) + \frac{1}{2}ar\sigma^2\beta_t^2. \tag{62}$$

Intuitively, the principal reimburses the agent's effort cost $\frac{\widehat{\mu}_t^2}{2}$, and compensates the risk premium $\frac{1}{2}ar\sigma^2\beta_t^2$ borne by the agent; they are just enough to convince the agent to take the offer. Then, in equilibrium,

$\mu_t = \widehat{\mu}_t$ and the agent's budget constraint reads

$$\begin{aligned} dS_t &= rS_t dt - c_t dt + \alpha_t dt + \beta_t \left(dY_t - \mu_t - m_t dt\right) \\ &= \frac{1}{2} ar\sigma^2 \beta_t^2 dt + \beta_t \left(dY_t - \mu_t - m_t dt\right). \end{aligned}$$

Since $v_t = -\frac{1}{ar} \exp\left(-arS_t\right)$, using Ito's lemma we have

$$\begin{aligned} dv_t &= \exp\left(-arS_t\right) dS_t + \frac{ar}{2} \exp\left(-arS_t\right) dS_t^2 \\ &= -arv_t \left[\frac{1}{2} ar\sigma^2 \beta_t^2 dt + \beta_t \left(dY_t - \mu_t - m_t dt\right)\right] - \frac{ar}{2} arv_t \sigma^2 \beta_t^2 dt \\ &= -arv_t \beta_t \left(dY_t - \mu_t - m_t dt\right). \end{aligned}$$

Thus, the agent's continuation value process is identical to equation (21). This also verifies that α_t in equation (62) is the minimum fixed wage needed to attract the agent.

Proposition 1 implies that the agent's incentive compatibility constraint satisfies

$$\begin{aligned} \mu_t &= \beta_t - \mathbb{E}_t \left[\int_t^\infty \phi \beta_s e^{-(\phi+r)(s-t)} \exp\left(-\int_t^s ar\beta_u \sigma dB_u - \frac{1}{2} \int_t^s a^2 r^2 \beta_u^2 \sigma^2 du\right) ds\right] \\ &= \beta_t - p_t. \end{aligned}$$

Importantly, because the principal t takes future β_{t+s} as given, the principal t is taking p_t as given and choosing β_t to maximize

$$\begin{aligned} & \mathbb{E}_t \left[dY_t\right]/dt - \mathbb{E}_t \left[\alpha_t dt - \beta_t \left(dY_t - \mu_t dt - m_t dt\right)\right]/dt \\ = \; & \mu_t + m_t - \alpha_t = \left(\beta_t - p_t\right) + m_t - g\left(\beta_t - p_t\right) - \frac{1}{2} ar\sigma^2 \beta_t^2. \end{aligned}$$

Hence, the first-order condition for the optimal incentive β_t is $1 - \left(\beta_t - p_t\right) - ar\sigma^2 \beta_t = 0$, which implies that

$$\beta_t = \frac{1 + p_t}{1 + ar\sigma^2}.$$

We will see this optimality condition does not hold in the long-term contracting case.

Conjecture that $\beta_t = \beta^{ST}$ and $p_t = p^{ST}$ are constants (we will verifying this property shortly.) Then, since $p^{ST} = \frac{\phi}{\phi+r} \beta^{ST}$,

$$\beta^{ST} = \frac{\phi + r}{r + ar\sigma^2 \left(\phi + r\right)}, p^{ST} = \frac{\phi}{r + ar\sigma^2 \left(\phi + r\right)},$$

and the equilibrium effort is

$$\mu^{ST} = \beta^{ST} - p^{ST} = \frac{r}{r + ar\sigma^2 \left(\phi + r\right)}.$$

Now let us rule out the case of timing varying β. Recall that the allowable set of β_t is bounded by say $[-M, M]$. Define $\overline{\beta} \equiv \sup\left\{\beta_t\right\} \in [-M, M]$. The optimality of short-term incentive implies that there exists t so that $p_t = \left(1 + ar\sigma^2\right) \overline{\beta} - 1 - \varepsilon$ for some sufficiently small ε. On the other hand, similar to the argument in Lemma 3, $p_t \leq \frac{\phi}{\phi+r} \overline{\beta}$, which implies

$$\left(1 + ar\sigma^2\right) \overline{\beta} - 1 - \varepsilon \leq \frac{\phi}{\phi + r} \overline{\beta} \Rightarrow \left(\frac{r}{\phi + r} + ar\sigma^2\right) \overline{\beta} \leq 1 + \varepsilon. \tag{63}$$

Similarly, define $\underline{\beta} \equiv \inf\left\{\beta_t\right\} \in [-M, M]$ we will have

$$\left(1 + ar\sigma^2\right) \underline{\beta} - 1 + \varepsilon \geq \frac{\phi}{\phi + r} \underline{\beta} \Rightarrow -\left(\frac{r}{\phi + r} + ar\sigma^2\right) \underline{\beta} \leq -1 + \varepsilon. \tag{64}$$

Summing equation (63) and (64), we have

$$\overline{\beta} - \underline{\beta} \leq \frac{2\varepsilon}{\frac{r}{\phi+r} + ar\sigma^2}.$$

Since ε is arbitrarily small, it must be $\overline{\beta} = \underline{\beta}$, and β_t is constant. QED.

A.8 Proof for Proposition 3

We first conjecture that the value function for the deterministic policy, $V^d(p)$, has the following quadratic form

$$V^d(p) = -\frac{1}{2}A^d p^2 + B^d p + C^d.$$

Plugging the above conjecture into the following ODE for the deterministic value function:

$$rV^d(p) = \frac{1}{2}\frac{\left(1 + p - \phi V^d(p)\right)^2}{1 + ar\sigma^2} - p - \frac{1}{2}p^2 + V_p^d(p)(\phi + r)p,$$

we can easily show that $B^d = \frac{1}{\phi}$, $C^d = 0$, and A^d satisfies

$$-\frac{1}{2}rA^d p^2 = \frac{1}{2}\frac{\left(1 + \phi A^d\right)^2 p^2}{1 + ar\sigma^2} - \frac{1}{2}p^2 - A^d(\phi + r)p^2.$$

Rearranging the above equation, we have

$$\phi^2\left(A^d\right)^2 - A^d\phi\left[\frac{r}{\phi}\left(1 + ar\sigma^2\right) + 2ar\sigma^2\right] - ar\sigma^2 = 0,$$

which gives the solution for A^d,

$$A^d = \frac{1}{2\phi}\left[\frac{r}{\phi}\left(1 + ar\sigma^2\right) + 2ar\sigma^2 + \sqrt{\left[\frac{r}{\phi}\left(1 + ar\sigma^2\right) + 2ar\sigma^2\right]^2 + 4ar\sigma^2}\right],$$

which is the same as that in the main text. Then, the optimal initial $p_0^d = \frac{B^d}{A^d}$ follows easily from the first order equation.

The incentive slope, as a function of information rent p_t, is

$$\beta_t^d = \frac{1 + p_t - V_p^d\phi}{1 + ar\sigma^2} = \frac{1 + A^d\phi}{1 + ar\sigma^2}p_t.$$

Using equation (38), we can derive the evolution of information rent p_t to be

$$\frac{dp_t^d}{p_t^d} = (\phi + r)\,dt - \frac{\beta_t^d}{p_t}\phi dt = \left(\phi + r - \frac{1 + A^d\phi}{1 + ar\sigma^2}\phi\right)dt \equiv -\lambda dt.$$

To show that $\lambda = \frac{1+A^d\phi}{1+ar\sigma^2}\phi - (\phi + r) > 0$, it is equivalent to show that $A^d > \left(1 + \frac{r}{\phi}\right)\frac{ar\sigma^2}{\phi} + \frac{r}{\phi^2}$ which always holds by following equation (72) in Lemma 4 (to be shown shortly). Finally, the optimal effort can be calculated as $\mu_t^d = \beta_t^d - p_t^d$.

A.9 Proof for Proposition 5

Before we solve the relaxed problem and derive the properties of the value function, we first define two useful functions, which prove to be useful in saving notations.

$$\begin{aligned} H_1(p) &\equiv \frac{1}{\phi}p - \frac{1}{2}\frac{r}{\phi^2}p^2 - \frac{a\sigma^2(\phi+r)^2}{2\phi^2}p^2, \\ H_2(p) &\equiv \frac{1}{r}\left[\frac{1}{2}\frac{(1+p)^2}{1+ar\sigma^2} - p - \frac{1}{2}p^2\right]. \end{aligned}$$

We state assumptions required for the proof below.

Assumption 1: The parameters satisfy the following condition:

$$a\left[2\left(\frac{\phi}{r}+1\right)^3 - \phi\sigma^2\right] < \frac{\phi}{r}. \tag{65}$$

Assumption 2: The feasible policy space for incentives is bounded given state p, i.e., $\beta(p) < \infty$ (though $\beta(p)$ may exceed M given M).

To solve the relaxed problem, we first focus on the following key ODE below:

$$\begin{aligned} rV(p) &= \frac{1}{2}\frac{(1+p-\phi V_p(p))^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{(V_p(p))^2}{V_{pp}}} - p - \frac{1}{2}p^2 + V_p(p)(\phi+r)p, \\ V(\pm M_p) &= H_1(\pm M_p). \end{aligned} \tag{66}$$

We proceed our proof with the following four steps.

Step 1: We first focus on the ODE in equation (66). We show that at $p=0$, we have $V(0)=0$ and $V_p(0)=\frac{1}{\phi}$.

Step 2: Under assumption 2, this ODE in equation (66) satisfies concavity and positivity of the denominator condition.

(a), Prove the concavity and positivity of the denominator at $p=0$.

(b), Prove the concavity and positivity of the denominator for general $p>0$.

(c), Prove the concavity and positivity of the denominator for $p<0$.

(d), Prove that the lower boundary 0 is absorbing and the upper boundary $\bar{p}$ is entrance-no-exit.

Step 3: From Steps 2.a-2.d, it follows from a standard verification theorem that the solution to the ODE in equation (66) is the value function of the principal's relaxed problem. Furthermore, it is never optimal to run outside the region $[0,\overline{p}]$ for the relaxed problem.

Step 4: Show that the value function for the relaxed problem is also the value function for the principal's original problem and $\bar{p}$ is independent of M for sufficiently large M.

A.9.1 Step 1: $V(0)=0$ and $V_p(0)=\frac{1}{\phi}$.

We proceed to show $V(0)=0$ and $V_p(0)=\frac{1}{\phi}$ by following the next three-step scheme.

Step 1.a: We first show that there must exist a solution to the following equation:

$$T(p) \equiv 1+p-\phi V_p = 0.$$

We know that $V(0)\geq 0$ which is the value of deterministic policy. We also know that

$$V(-M_p) = H_1(-M_p) < 0 \text{ and } V(M_p) = H_1(M_p) < 0.$$

Then according to the intermediate value theorem, there exists $p_1 > 0$ so that

$$T(p_1) = 1 + p_1 - \phi V_p(p_1) = 1 + p_1 - \phi \frac{V(M_p) - V(0)}{M_p} > 1,$$

and there also exists $p_2 < 0$ such that

$$T(p_2) = 1 + p_2 - \phi V_p(p_2) = 1 + p_2 - \phi \frac{V(0) - V(-M_p)}{M_p} < 0,$$

for sufficiently high M_p. Therefore, we can find a point $\underline{p}$ such that

$$1 + \underline{p} - \phi V_p\left(\underline{p}\right) = 0. \tag{67}$$

Step 1.b: Suppose that $1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\underline{p})}{V_{pp}(\underline{p})} \neq 0$. We aim to show that $\underline{p} = 0$ so that equations (67) and (66) imply $V(0) = 0$ and $V_p(0) = \frac{1}{\phi}$. Differentiating the HJB in equation (66) with respect to p at $\underline{p}$, we have

$$rV_p\left(\underline{p}\right) = -1 - \underline{p} + V_{pp}\left(\underline{p}\right)(\phi + r)\underline{p} + V_p\left(\underline{p}\right)(\phi + r),$$

which, together with equation (67), imply that

$$V_{pp}\left(\underline{p}\right)\underline{p} = 0.$$

Therefore, either $\underline{p} = 0$ or $V_{pp}\left(\underline{p}\right) = 0$. We first rule out the case of $V_{pp}\left(\underline{p}\right) = 0$. Note that this case implies the denominator of the first term in the right hand side of equation (66) is infinite, so that V must satisfy

$$rV\left(\underline{p}\right) = -\underline{p} - \frac{1}{2}\underline{p}^2 + \left(1 + \underline{p}\right)\left(1 + \frac{r}{\phi}\right)\underline{p}, \tag{68}$$

Further, it follows from Taylor expansion that

$$\begin{aligned} V\left(\underline{p} + \epsilon\right) &= V\left(\underline{p}\right) + \frac{1}{\phi}\left(1 + \underline{p}\right)\epsilon + o\left(\epsilon^2\right) \\ V_p\left(\underline{p} + \epsilon\right) &= \frac{1}{\phi}\left(1 + \underline{p}\right) + o\left(\epsilon\right). \end{aligned} \tag{69}$$

Thus, evaluating the HJB equation (66) at $\underline{p} + \epsilon$, we have

$$\begin{aligned} rV\left(\underline{p} + \epsilon\right) &= \frac{\frac{1}{2}\left(\epsilon - o(\epsilon)\right)^2}{1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\underline{p}+\epsilon)}{V_{pp}(\underline{p}+\epsilon)}} - \underline{p} - \epsilon - \frac{\left(\underline{p} + \epsilon\right)^2}{2} + \frac{1}{\phi}\left(1 + \underline{p}\right)(\phi + r)\left(\underline{p} + \epsilon\right) + o\left(\epsilon^2\right) \\ &= -\underline{p} - \frac{1}{2}\underline{p}^2 + \left(1 + \underline{p}\right)\left(1 + \frac{r}{\phi}\right)\underline{p} - \epsilon - \underline{p}\epsilon - \frac{1}{2}\epsilon^2 + \left(1 + \underline{p}\right)\left(1 + \frac{r}{\phi}\right)\epsilon + o\left(\epsilon^2\right) \\ &= rV\left(\underline{p}\right) + \frac{r}{\phi}\left(1 + \underline{p}\right)\epsilon - \frac{1}{2}\epsilon^2 + o\left(\epsilon^2\right), \end{aligned}$$

where the second equality uses the fact $\frac{V_p^2(\underline{p}+\epsilon)}{V_{pp}(\underline{p}+\epsilon)}$ goes to infinity as ϵ goes to zero because the continuity of $V_{pp}\left(\underline{p}\right)$ implies that $V_{pp}\left(\underline{p} + \epsilon\right)$ is at the order of $o(1)$, and the last equality follows from equation (68). But this is in contradiction with equation (69) since they do not match at the second order ϵ^2. As a result, $\underline{p} = 0$ and thus $V_p(0) = \frac{1}{\phi}$.

Step 1.c: Now suppose that $1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2(\underline{p})}{V_{pp}(\underline{p})}=0$, i.e.,

$$V_{pp}\left(\underline{p}\right)=-\frac{a^2r^2\sigma^2V_p^2\left(\underline{p}\right)}{1+ar\sigma^2}=-\frac{a^2r^2\sigma^2}{1+ar\sigma^2}\frac{1}{\phi^2}\left(1+\underline{p}\right)^2, \tag{70}$$

which implies that we cannot ignore the term with $1+p-\phi V_p$. Due to L'Hospital's rule,

$$\frac{\left(1+p-\phi V_p\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}=\frac{2\left(1+p-\phi V_p\right)\left(1-\phi V_{pp}\right)}{a^2r^2\sigma^2\frac{2V_pV_{pp}^2-V_p^2V_{ppp}}{V_{pp}^2}}. \tag{71}$$

Differentiating the HJB equation (66), we have,

$$\begin{aligned} rV_p &= \frac{\left(1+\underline{p}-\phi V_p\right)\left(1-\phi V_{pp}\right)}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}-\frac{1}{2}\frac{\left(1+\underline{p}-\phi V_p\right)^2}{\left(1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}\right)^2}\left[a^2r^2\sigma^2\frac{2V_pV_{pp}^2-V_p^2V_{ppp}}{V_{pp}^2}\right] \\ &\quad -1-\underline{p}+V_p\left(\phi+r\right)+V_{pp}\left(\phi+r\right)\underline{p}. \end{aligned}$$

Plugging in equation (71) into the above equation, we find that the two terms in the first line cancel each other, and

$$0=-1-\underline{p}+V_p\phi+V_{pp}\left(\phi+r\right)\underline{p}=V_{pp}\left(\phi+r\right)\underline{p},$$

which is the same as before. Therefore, either we have

$$\underline{p}=0,\text{ and }V_p\left(0\right)=\frac{1}{\phi},$$

or we have $V_{pp}\left(\underline{p}\right)=0$ and $\underline{p}=-1$ due to equation (70). In the second case, we further have

$$V_{pp}\left(-1\right)=0,V_p\left(-1\right)=0,\text{ and }\frac{V_p^2\left(-1\right)}{V_{pp}\left(-1\right)}=-\frac{1+ar\sigma^2}{a^2r^2\sigma^2}$$

In Lemma 6 below,[18] we will show that for any p, if

$$V_{pp}\left(p\right)=V_p\left(p\right)=0\text{ and }\lim\frac{V_p^2\left(p\right)}{V_{pp}\left(p\right)}\to-q,$$

then $q=0$ or $\frac{1}{ar}$. Therefore, the second alternative results in contradiction and we must have $\underline{p}=0$, and $V_p\left(0\right)=\frac{1}{\phi}$.

Now we still need to show that $V\left(0\right)=0$ under the assumption $1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2(\underline{p})}{V_{pp}(\underline{p})}=0$. Suppose that $V\left(0\right)=v>0$. First, the HJB in equation (66) implies that

$$\lim_{p\to0}\frac{\left(1+p-\phi V_p\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}=2rv>0.$$

Thus, it follows from equation (42) that the policy function $\beta\left(p\right)$ at $p=0$ has value

$$\begin{aligned} \beta\left(0\right) &= \lim_{p\to0}\frac{\left(1+p-\phi V_p\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}\cdot\frac{1}{1+p-\phi V_p} \\ &= \lim_{p\to0}2rv\frac{1}{1+p-\phi V_p}=\infty. \end{aligned}$$

This contradicts with Assumption 2 that our policy space is restricted to be bounded at $p=0$.

[18] It is important to point out that the proof for Lemma 6 below does not use any results from step 1 here. Thus, there is no circular argument.

A.9.2 Step 2: Prove the concavity and positivity

Step 2.a: Concavity & Positivity of the Denominator at $p = 0$: We want to show that under assumption 1-2, the value function at $p = 0$ is concave, $V_{pp}(0) < 0$, and $1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p(0)^2}{V_{pp}(0)} > 0$. Before we proceed to the main proof, we first show the following two lemmas, which are needed for our main proof.

Lemma 4 *Denote*

$$\chi^o \equiv \frac{A^o}{1/\phi + A^o} \text{ and } \chi \equiv \frac{A^d - \frac{a^2r^2\sigma^2(1/\phi)^2}{1+ar\sigma^2}}{1/\phi + A^d}.$$

Under the condition in equation (49), for $a > 0$, the following equalities hold:

$$\begin{aligned} A^d - A^o &> \frac{ar\sigma^2}{\phi}\left(1 + \frac{r}{\phi}\right) > \frac{2a^2\sigma^2\phi}{r(1+ar\sigma^2)}\left(1+\frac{r}{\phi}\right)^4, \\ \chi^o &< \chi. \end{aligned} \tag{72}$$

Proof. Recall the definition of A^d and A^o, we have $A^o = \frac{r}{\phi^2}$, and

$$\begin{aligned} A^d &= \frac{1}{2\phi^2}\left((2\phi + r)ar\sigma^2 + r + \sqrt{(2\phi+r)^2 a^2r^2\sigma^4 + 2\sigma^2\left[(\phi+r)^2 + \phi^2\right]ar + r^2}\right) \\ &= \frac{1}{\phi}\left(\frac{r}{2\phi}\left(1+ar\sigma^2\right) + ar\sigma^2 + \sqrt{\left[\frac{r}{2\phi}(1+ar\sigma^2) + ar\sigma^2\right]^2 + ar\sigma^2}\right). \end{aligned}$$

Because $\sqrt{\left(\frac{r}{2\phi}(1+ar\sigma^2) + ar\sigma^2\right)^2 + ar\sigma^2} > \frac{r}{2\phi}\left(1+ar\sigma^2\right) + ar\sigma^2$, it follows

$$A^d - A^o > \frac{1}{\phi}\left(\frac{r}{\phi}\left(1+ar\sigma^2\right) + 2ar\sigma^2 - \frac{r}{\phi}\right) > \frac{ar\sigma^2}{\phi}\left(1+\frac{r}{\phi}\right). \tag{73}$$

Furthermore, assumption 1 (i.e., equation (49)) implies that

$$\begin{aligned} &\frac{\phi}{r} + a\phi\sigma^2 > 2a\left(\frac{\phi}{r}+1\right)^3 \Leftrightarrow \left(\frac{r}{\phi}\right)^3\left(\frac{\phi}{r} + a\sigma^2\phi\right) > 2a\left(1+\frac{r}{\phi}\right)^3 \\ \Leftrightarrow\ &\left(\frac{r}{\phi}\right)^2\left(1+ar\sigma^2\right) > 2a\left(1+\frac{r}{\phi}\right)^3 \\ \Leftrightarrow\ &\frac{ar\sigma^2}{\phi}\left(1+\frac{r}{\phi}\right) > \frac{2a^2\sigma^2\phi}{r(1+ar\sigma^2)}\left(1+\frac{r}{\phi}\right)^4. \end{aligned} \tag{74}$$

Thus, combining equation (73) and equation (74), we complete the proof for equation (72).

Lastly, we want to show $\chi^o < \chi$. First, the equality $2\left(1+\frac{r}{\phi}\right)^3 > \left(\frac{r}{\phi}\right)^3$ implies that

$$\frac{2a^2\sigma^2\phi}{r(1+ar\sigma^2)}\left(1+\frac{r}{\phi}\right)^4 > \left(1+\frac{r}{\phi}\right)\frac{a^2r^2\sigma^2(1/\phi)^2}{1+ar\sigma^2}.$$

Combining with the equality in equation (72), we have

$$A^d > A^o + \left(1+\frac{r}{\phi}\right)\frac{a^2r^2\sigma^2(1/\phi)^2}{1+ar\sigma^2}.$$

Further, notice that

$$\begin{aligned}
\chi^o - 1 &< \chi - 1 \Leftrightarrow -\frac{1}{1+\phi A^o} < \frac{-1 - \frac{a^2r^2\sigma^2(1/\phi)}{1+ar\sigma^2}}{1+\phi A^d} \\
A^d &> A^o + \frac{a^2r^2\sigma^2\left(1/\phi\right)^2}{1+ar\sigma^2}\frac{\left(1/\phi + A^o\right)}{1/\phi} \Leftrightarrow A^d > A^o + \frac{a^2r^2\sigma^2\left(1/\phi\right)^2}{1+ar\sigma^2}\left(1+\frac{r}{\phi}\right).
\end{aligned}$$

Therefore, the inequality $\chi^o < \chi$ follows immediately. ■

Lemma 5 *Under Condition equation (49) and for $a > 0$, there exists a unique root $x_0 \in (0, \chi^o)$ to the following cubic polynomial:*

$$\begin{aligned}
G(x) \equiv\ & \left(A^d - x\left(\frac{1}{\phi} + A^d\right)\right)(1-x)^2 + \frac{2x}{1+\phi A^d}\left(\left(A^d - x\left(\frac{1}{\phi}+A^d\right)\right)\left(1+ar\sigma^2\right) - a^2r^2\frac{\sigma^2}{\phi^2}\right) \\
& - \left(A^d - x\left(\frac{1}{\phi}+A^d\right)\right) + \frac{a^2r^2}{1+ar\sigma^2}\frac{\sigma^2}{\phi^2}. \qquad (75)
\end{aligned}$$

Proof. Firstly, $G(0) = \frac{a^2r^2}{1+ar\sigma^2}\frac{\sigma^2}{\phi^2} > 0$. Secondly, notice that

$$\begin{aligned}
G'(x) =\ & -\left(\frac{1}{\phi}+A^d\right)(1-x)^2 - 2\left(A^d - x\left(\frac{1}{\phi}+A^d\right)\right)(1-x) \\
& + \frac{2}{1+\phi A^d}\left(\left(A^d - x\left(\frac{1}{\phi}+A^d\right)\right)\left(1+ar\sigma^2\right) - a^2r^2\frac{\sigma^2}{\phi^2}\right) \\
& - \frac{2x}{1+\phi A^d}\left(\frac{1}{\phi}+A^d\right)\left(1+ar\sigma^2\right) + \left(\frac{1}{\phi}+A^d\right).
\end{aligned}$$

Evaluate the above equation at $p = 0$, we have

$$G'(0) = \frac{2A^d}{1+\phi A^d}\left(-\phi A^d + ar\sigma^2 - \frac{a^2r^2\sigma^2}{A^d\phi^2}\right) - \frac{\frac{1}{\phi} + 2A^d + \phi\left(A^d\right)^2}{1+\phi A^d} < 0,$$

where the last inequality is due to the following fact

$$\phi A^d = \frac{r}{2\phi}\left(1+ar\sigma^2\right) + ar\sigma^2 + \sqrt{\left[\frac{r}{2\phi}\left(1+ar\sigma^2\right) + ar\sigma^2\right]^2 + ar\sigma^2} > ar\sigma^2.$$

Thirdly, we proceed to show that $G''(x) > 0$ for $x \le \chi^o$. Note that the second derivative of $G(x)$ can be simplified to the following:

$$G''(x) = 4\left(\frac{1}{\phi}+A^d\right)\left(1 - x - \frac{1+ar\sigma^2}{1+\phi A^d}\right) + 2\left(A^d - x\left(\frac{1}{\phi}+A^d\right)\right). \qquad (76)$$

From Lemma 4, we have $\chi^o < \chi$. Note that $\chi < \frac{A^d}{1/\phi + A^d}$, it follows that for $x \le \chi^o < \chi$,

$$A^d - x\left(\frac{1}{\phi}+A^d\right) > 0.$$

Furthermore, for $x \le \chi^o$, we have

$$1 - x - \frac{1+ar\sigma^2}{1+\phi A^d} \ge 1 - \chi_0 - \frac{1+ar\sigma^2}{1+\phi A^d} = \frac{\left(A^d - A^o\right) - \frac{ar\sigma^2}{\phi}\left(1+\frac{r}{\phi}\right)}{\left(1/\phi + A^o\right)\left(1+\phi A^d\right)} > 0,$$

where the last inequality follows equation (72). Thus, both terms in equation (76) are positive, and hence $G''(x) > 0$.

Below we further show that $G(\chi^o) < 0$ holds. After tedious algebra, one can show

$$\begin{aligned} G(\chi^o) &= (1/\phi)(2\chi^o)\left\{\frac{A^d - A^o}{1/\phi + A^o}\left[-1 + \frac{\chi^o}{2} + \frac{1+ar\sigma^2}{1+\phi A^d}\right] + \frac{a^2r^2\sigma^2(1/\phi)}{(1+ar\sigma^2)}\left[\frac{1}{2\chi^o} - \frac{1+ar\sigma^2}{1+\phi A^d}\right]\right\} \\ &< (1/\phi)(2\chi^o)\left\{\frac{A^d - A^o}{1/\phi + A^o}\left[-1 + \frac{\chi^o}{2} + \frac{1+ar\sigma^2}{1+\phi A^d}\right] + \frac{a^2r^2\sigma^2(1/\phi)}{(1+ar\sigma^2)}\frac{1}{2\chi^o}\right\}. \end{aligned}$$

It follows from straightforward algebra that

$$\begin{aligned} 1 - \frac{\chi^o}{2} - \frac{1+ar\sigma^2}{1+\phi A^d} &= \frac{\frac{r}{2\phi} + \sqrt{\left[\frac{r}{2\phi} + \frac{ar\sigma^2}{(1+ar\sigma^2)}\right]^2 + \frac{ar\sigma^2}{(1+ar\sigma^2)^2}}}{\left(\frac{r}{2\phi}+1\right) + \sqrt{\left[\frac{r}{2\phi} + \frac{ar\sigma^2}{(1+ar\sigma^2)}\right]^2 + \frac{ar\sigma^2}{(1+ar\sigma^2)^2}}} - \frac{\frac{r}{2\phi}}{1+\frac{r}{\phi}} \\ &> \frac{\frac{r}{2\phi}}{1+\frac{r}{2\phi}} - \frac{\frac{r}{2\phi}}{1+\frac{r}{\phi}} > \frac{\left(\frac{r}{2}\right)^2}{(\phi+r)^2}. \end{aligned}$$

Together with equation (72), we obtain

$$\begin{aligned} G(\chi^o) &< (1/\phi)(2\chi^o)\left\{-\frac{A^d - A^o}{1/\phi + A^o}\frac{\left(\frac{r}{2\phi}\right)^2}{\left(1+\frac{r}{\phi}\right)^2} + \frac{a^2r^2\sigma^2(1/\phi)}{(1+ar\sigma^2)}\frac{1+\frac{r}{\phi}}{2\frac{r}{\phi}}\right\} \\ &= -(1/\phi)\frac{\chi^o r^2\frac{1}{\phi}}{2\left(1+\frac{r}{\phi}\right)^3}\left\{\left(A^d - A^o\right) - \frac{2a^2\sigma^2\phi}{r(1+ar\sigma^2)}\left(1+\frac{r}{\phi}\right)^4\right\} < 0. \end{aligned}$$

Now we have $G''(x) > 0$ for $x \le \chi^o$, $G'(0) < 0$, $G(0) > 0$ and $G(\chi^o) > 0$. Hence, there must exist one point $x \in (0, \chi^o)$ such that $G(x) = 0$. We still need to show the uniqueness. We consider two cases: $G'(\chi^o) \le 0$ and $G'(\chi^o) > 0$. For the first case, since $G''(x) > 0$ for $x \le \chi^o$, we have $G'(x) < 0$ for all $x \in (0, \chi^o)$, and hence $G(x)$ is monotonically decreasing in $(0, \chi^o)$. Thus, there is a unique x such that $G(x) = 0$. For the second case, since $G''(x) > 0$ for $x \le \chi^o$, there is a unique $x_1 \in (0, \chi^o)$ such that $G'(x_1) = 0$. For $x \in (0, x_1)$, $G(x)$ is strictly decreasing since $G'(x) < 0$ for $x \in (0, x_1)$. For $x \in (x_1, \chi^o)$, $G(x)$ is strictly increasing since $G'(x) > 0$ for $x \in (x_1, \chi^o)$. Thus, $G(x_1) < 0$ and there exists a unique $x \in (0, x_1) \subset (0, \chi^o)$ such that $G(x) = 0$. We complete our proof. ■

Given the above two lemmas, now we proceed to the main proof for step 2.a. Recall that the key ODE in equation (66) is

$$rV = \frac{1}{2}\frac{(1+p-\phi V_p)^2}{1+ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}} - p - \frac{1}{2}p^2 + V_p(\phi+r)p.$$

The value function in the deterministic case V^d satisfies the following ODE

$$rV^d = \frac{1}{2}\frac{\left(1+p-\phi V_p^d\right)^2}{1+ar\sigma^2} - p - \frac{1}{2}p^2 + V_p^d(\phi+r)p.$$

Thus, taking the difference of the above two questions, we have

$$V - V^d = \frac{1}{2r}\left[\frac{(1+p-\phi V_p)^2}{1+ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}} - \frac{\left(1+p-\phi V_p^d\right)^2}{1+ar\sigma^2}\right] + \left(\frac{\phi}{r}+1\right)p\left[V_p - V_p^d\right].$$

For $p > 0$, dividing both sides by p and letting p go to zero, we have

$$\lim_{p\downarrow 0}\frac{V-V^d}{p}=\lim_{p\downarrow 0}\frac{1}{2rp}\left[\frac{\left(1+p-\phi V_p\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}-\frac{\left(1+p-\phi V_p^d\right)^2}{1+ar\sigma^2}\right]+\left(\frac{\phi}{r}+1\right)\lim_{p\downarrow 0}\left(V_p-V_p^d\right) \tag{77}$$

Because $V(0)=V^d(0)=0$, we have

$$\lim_{p\downarrow 0}\frac{V-V^d}{p}=\lim_{p\downarrow 0}\frac{V-0}{p}-\lim_{p\downarrow 0}\frac{V^d-0}{p}=V_p(0)-V_p^d(0)=\lim_{p\downarrow 0}\left(V_p-V_p^d\right). \tag{78}$$

Thus, combining equation (77) and equation (78) yields

$$-\frac{\phi}{r}\lim_{p\downarrow 0}\left(V_p-V_p^d\right)=\lim_{p\downarrow 0}\frac{1}{2rp}\left[\frac{\left(1+p-\phi V_p\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}-\frac{\left(1+p-\phi V_p^d\right)^2}{1+ar\sigma^2}\right].$$

Dividing both sides again by p and letting p go to zero, we have

$$-\frac{\phi}{r}\lim_{p\downarrow 0}\frac{V_p-V_p^d}{p}=\lim_{p\downarrow 0}\frac{1}{2rp^2}\left[\frac{\left(1+p-\phi V_p\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}-\frac{\left(1+p-\phi V_p^d\right)^2}{1+ar\sigma^2}\right]. \tag{79}$$

Noting that $V_p(0)=V_p^d(0)=\frac{1}{\phi}$, it follows from L' Hospital's rule that

$$\lim_{p\downarrow 0}\frac{V_p-V_p^d}{p}=\lim_{p\downarrow 0}\frac{V_p-(1/\phi)}{p}-\lim_{p\downarrow 0}\frac{V_p^d-(1/\phi)}{p}=V_{pp}(0)-V_{pp}^d(0), \tag{80}$$

and

$$\lim_{p\downarrow 0}\frac{\left(1+p-\phi V_p\right)^2}{p^2}=\lim_{p\downarrow 0}\frac{2\left(1+p-\phi V_p\right)\left(1-\phi V_{pp}\right)}{2p}=\left(1-\phi V_{pp}(0)\right)^2. \tag{81}$$

Further, plugging the expression for $V^d(p)$, we have

$$1+p-\phi V_p^d=1+p-\phi\left(-A^dp+B^d\right)=\left(1+\phi A^d\right)p. \tag{82}$$

Substituting the above equations (80), (81), and (82) back into equation (79), we have

$$-\phi\left(V_{pp}(0)-V_{pp}^d(0)\right)=\frac{1}{2}\frac{\left(1-\phi V_{pp}(0)\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{(1/\phi)^2}{V_{pp}(0)}}-\frac{\left(1+\phi A^d\right)^2}{2\left(1+ar\sigma^2\right)}$$

Let's define constant $\hat{x}$ as

$$\hat{x}\equiv\frac{V_{pp}(0)-V_{pp}^d(0)}{\frac{1}{\phi}+A^d}.$$

Then, after tedious algebraic manipulation, one can show that $\hat{x}$ is a root of the 3-order polynomial $G(x)$, which is defined in equation (75) in Lemma 5. According to Lemma 5, there exists a unique positive root $x_0\in(0,\chi^o)$. Thus, $\hat{x}=x_0$ is such a unique root. Further, Lemma 4 implies $\chi^o<\chi<\frac{A^d}{1/\phi+A^d}$, and hence

$$0<V_{pp}(0)-V_{pp}^d(0)=x_0\left(1/\phi+A^d\right)<\chi\left(1/\phi+A^d\right)<A^d.$$

Therefore,

$$V_{pp}(0)<V_{pp}^d(0)+A^d=-A^d+A^d=0.$$

Furthermore, since $x_0<\chi$, we have $x_0\left(1/\phi+A^d\right)-A^d<\chi\left(1/\phi+A^d\right)-A^d<0$, and hence,

$$\begin{aligned}1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2(0)}{V_{pp}(0)} &= 1+ar\sigma^2+a^2r^2\sigma^2\frac{(1/\phi)^2}{x_0\left(1/\phi+A^d\right)-A^d}\\ &> 1+ar\sigma^2+a^2r^2\sigma^2\frac{(1/\phi)^2}{\chi\left(1/\phi+A^d\right)-A^d}=0.\end{aligned}$$

Thus, the denominator at $p=0$ is always positive.

Step 2.b: Concavity & Positivity of the Denominator at $p > 0$: We want to prove for any $p > 0$, the following must hold:

$$1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(p)}{V_{pp}(p)} > 0 \text{ and } V_{pp}(p) < 0.$$

We prove the above inequalities by contradiction. Suppose that the above two inequalities fail to hold at some points. Denote the smallest point $\widehat{p} > 0$ at which at least one of the two inequalities does not hold. That is,

$$1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(p)}{V_{pp}(p)} > 0 \text{ and } V_{pp}(p) < 0 \text{ for } p \in [0, \widehat{p})$$

while at $\widehat{p}$, one of the following three cases holds:

$$\text{(Case 1) } 1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = 0 \text{ and } V_{pp}(\widehat{p}) < 0$$

or

$$\text{(Case 2) } 1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = 0 \text{ and } V_{pp}(\widehat{p}) = 0$$

or

$$\text{(Case 3) } 1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} > 0 \text{ and } V_{pp}(\widehat{p}) = 0.$$

We first show that Case 1 is impossible. Since $V_{pp}(p) < 0$ for $p \in [0, \widehat{p})$, we know that

$$T(p) \equiv 1 + p - \phi V_p \tag{83}$$

is strictly increasing in p and positive. To see this,

$$T'(p) = 1 - \phi V_{pp} > 0,\ T(0) = 0.$$

Therefore, $T(\widehat{p}) > 0$ and $T'(\widehat{p}) > 0$. This implies that

$$\frac{\left(1 + \widehat{p} - \phi V_p(\widehat{p})\right)^2}{1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})}} = \infty,$$

which contradicts with the finiteness of V. Also we should use the fact that V_p is bounded due to $V \in \mathbb{C}^2$. By the same argument, Case 2 is also impossible.

We then consider Case 3. Note that Case 3 occurs only in the situation that

$$V_p(p) \to 0^+ \text{ and } V_{pp}(p) \to 0^- \text{ when } p \uparrow \widehat{p}.$$

Otherwise, a non-zero $V_p(\widehat{p})$ implies $\frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = -\infty$, a contradiction to the fact that $1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(p)}{V_{pp}(p)} > 0$ for all $p < \widehat{p}$. Therefore, we can define a positive finite number $q \geq 0$ such that there exists some subsequence $\{p_n = \hat{p} - \epsilon_n\}$ approaching $\widehat{p}$ from left so that

$$\lim_{p_n \uparrow \widehat{p}} Q(p_n) \equiv \lim_{p_n \uparrow \widehat{p}} \frac{V_p^2(p_n)}{V_{pp}(p_n)} = -q \leq 0.$$

Moreover, we must have $\lim_{p \downarrow \widehat{p}} \frac{V_p^2(p)}{V_{pp}(p)} = -q \leq 0$ because otherwise $V(p)$ will exhibit a jump at $\widehat{p}$. Finally, we must have $1 + ar\sigma^2 - a^2r^2\sigma^2 q > 0$.

Taking the difference on the key ODE in equation (66) at $p = p_n$ and $p = \hat{p}$, for any sequence $\{p_n\} \to \hat{p}$ we have

$$\begin{aligned}
& \left(1+\hat{p}-\phi V_p\left(\hat{p}\right)\right)^2-\left(1+p_n-\phi V_p\left(p_n\right)\right)^2 \\
= \; & 2\left[rV\left(\hat{p}\right)+\hat{p}+\frac{1}{2}\hat{p}^2-V_p\left(\hat{p}\right)\left(\phi+r\right)\hat{p}\right]\left(1+ar\sigma^2+a^2r^2\sigma^2\frac{\left(V_p\left(\hat{p}\right)\right)^2}{V_{pp}\left(\hat{p}\right)}\right) \\
& -2\left[rV\left(p_n\right)+p_n+\frac{1}{2}p_n^2-V_p\left(p_n\right)\left(\phi+r\right)p_n\right]\left(1+ar\sigma^2+a^2r^2\sigma^2\frac{\left(V_p\left(p_n\right)\right)^2}{V_{pp}\left(p_n\right)}\right).
\end{aligned}$$

Simplifying the above equation further, we obtain

$$\begin{aligned}
\left(1+\hat{p}\right) \; = \; & \left(1+\hat{p}\right)\left(1+ar\sigma^2+a^2r^2\sigma^2Q\left(\hat{p}\right)\right)+o_n\left(1\right) \\
& +\left[\frac{1}{2}\frac{\left(1+p_n-\phi V_p\left(p_n\right)\right)^2}{1+ar\sigma^2+a^2r^2\sigma^2\frac{V_p^2\left(p_n\right)}{V_{pp}\left(p_n\right)}}\right]\left(a^2r^2\sigma^2\frac{Q\left(p\right)-Q\left(p_n\right)}{p-p_n}\right).
\end{aligned}$$

Rearranging the above equation and noticing that $1+ar\sigma^2-a^2r^2\sigma^2q>0$, we have

$$arq-1=\frac{1}{2}\frac{1+\hat{p}}{1+ar\sigma^2-a^2r^2\sigma^2q}ar\left(\lim_{n\to\infty}\frac{Q\left(\hat{p}\right)-Q\left(p_n\right)}{\hat{p}-p_n}\right). \tag{84}$$

Thus, $\lim_{n\to\infty}\frac{Q(\hat{p})-Q(p_n)}{\hat{p}-p_n}$ must exist and be finite.

First, suppose $q \neq 0$, then we know $\lim_{n\to\infty}\frac{V_{pp}(p_n)}{V_p^2(p_n)} = -\frac{1}{q}$, which is finite. Thus, it follows from $\frac{V_{pp}(\widehat{p}-\epsilon_n)}{V_p^2(\widehat{p}-\epsilon_n)} = -\frac{1}{q}+o_n\left(1\right)$ that

$$\begin{aligned}
V_{pp}\left(p_n\right) \; & = \; -\frac{1}{q}V_p^2\left(p_n\right)+V_p^2\left(p_n\right)o_n\left(1\right) \\
& = \; -\frac{1}{q}2V_p\left(\widehat{p}_n\right)V_{pp}\left(\widehat{p}_n\right)\epsilon_n+2V_p\left(\widehat{p}\right)V_{pp}\left(\widehat{p}_n\right)\epsilon_n o_n\left(1\right),
\end{aligned}$$

where the second equality is due to the mean value theorem and $\hat{p} \geq \hat{p}_n \geq p_n$. Therefore, for any sequence $\{p_n\} \to \hat{p}$ we have

$$V_{ppp}\left(\widehat{p}\right) \equiv \lim_{n\to\infty}\frac{V_{pp}\left(\widehat{p}-\epsilon_n\right)}{\epsilon_n} = \lim_{n\to\infty}-\frac{1}{q}2V_p\left(\widehat{p}_n\right)V_{pp}\left(\widehat{p}_n\right)+2V_p\left(\widehat{p}\right)V_{pp}\left(\widehat{p}_n\right)o_n\left(1\right)=0.$$

Thus, $V_{ppp}\left(\widehat{p}\right)$ exists, and hence using $Q\left(p\right)=\frac{V_p^2(p_n)}{V_{pp}(p_n)}$ equation (84) implies that

$$1-arq=\frac{1}{2}\frac{1+\widehat{p}}{1+ar\sigma^2-a^2r^2\sigma^2q}\left[ar\frac{V_p^2\left(\widehat{p}\right)}{V_{pp}\left(\widehat{p}\right)}\frac{V_{ppp}\left(\widehat{p}\right)}{V_{pp}\left(\widehat{p}\right)}\right]. \tag{85}$$

Below we show that there is a contradiction for Case 3 by the following three lemmas.

Lemma 6 *It must be that either $q=0$ or $q=\frac{1}{ar}$.*
Proof. *Assume that $q \neq 0$. Recall q is defined so that*

$$\lim_{p\uparrow\hat{p}}\frac{V_p^2}{V_{pp}}=-q<0,$$

which implies that

$$\lim_{p\uparrow\widehat{p}} \frac{V_p^2}{V_{pp}} = \lim_{p\uparrow\widehat{p}} \frac{2V_pV_{pp}}{V_{ppp}} = \lim_{p\uparrow\widehat{p}} 2V_p \cdot \frac{V_{pp}}{V_{ppp}} = -q,$$

Hence, we have

$$\lim_{p\uparrow\widehat{p}} \frac{V_{pp}}{V_{ppp}} = -\infty, \text{ and } \lim_{p\uparrow\widehat{p}} \frac{V_{ppp}}{V_{pp}} = 0^-$$

From equation (85) and $q \neq 0$, we have

$$\lim_{p\uparrow\widehat{p}} \frac{V_p^2}{V_{pp}} \frac{V_{ppp}}{V_{pp}} = 0,$$

and hence $q = \frac{1}{ar}$. As a result, $q = 0$ or $q = \frac{1}{ar}$. ■

Lemma 7 *It is impossible to have $q = \frac{1}{ar}$.*
Proof. *Suppose $q = \frac{1}{ar}$. By plugging $V_p(\widehat{p}) = V_{pp}(\widehat{p}) = 0$ and $q = \frac{1}{ar}$ into the key ODE in equation (66), we obtain*

$$V(\widehat{p}) = \frac{1}{2r} > V^{HM} = \frac{1}{2r}\frac{1}{1+ar\sigma^2},$$

which contradicts with condition (44). ■

Lemma 8 *It is also impossible to have $q = 0$.*
Proof. *Suppose that $q = 0$. From equation (84), we have*

$$-1 = \frac{1}{2}\frac{(1+\hat{p})}{1+ar\sigma^2}\left(ar \lim_{n\to\infty} \frac{Q(\hat{p}) - Q(p_n)}{\hat{p} - p_n}\right),$$

which implies that $\lim_{n\to\infty} \frac{Q(\hat{p})-Q(p_n)}{\hat{p}-p_n} < 0$. Using this property, we can obtain a contradiction. To see this, $Q(\hat{p}) = \lim_{n\to\infty} \frac{V_p^2(p_n)}{V_{pp}(p_n)} = -q = 0$ at $\widehat{p}$. However, because $\widehat{p}$ is the first point so that $V_{pp}(\widehat{p}) = 0$, we know that $V_{pp}(\widehat{p}-) < 0$, which implies that

$$Q(\widehat{p}-) = \frac{V_p^2(\widehat{p}-)}{V_{pp}(\widehat{p}-)} < 0.$$

As a result, $\lim_{n\to\infty} \frac{Q(\hat{p})-Q(p_n)}{\hat{p}-p_n} > 0$, which is a contradiction. ■

Combining the above three lemmas, we obtain that Case 3 is impossible. Thus, we complete the proof for Step 2.b. Since $V(0) = 0$, $V_p(0) = \frac{1}{\phi} > 0$, and $V(M_p) < 0$, and $V(p) < 0$ for $p \geq 0$, it follows that there exists $\overline{p} > 0$ such that $V'(\overline{p}) = 0$. The analysis above also implies that $V_{pp}(\overline{p}) < 0$ strictly. Therefore, the function is strictly concave and with positive denominator for $p \in [0, \overline{p})$. Before we proceed to Step 2.c, we show the following lemma on the property of $\overline{p}$, which is useful in the subsequent proof.

Lemma 9 *$\overline{p} \leq \overline{p}^d$ where $\overline{p}^d \equiv \overline{p}_0^d = B^d/A^d$.*
Proof. *From the key ODE equation (66), we know that at $\overline{p}$,*

$$V(\overline{p}) = H_2(\overline{p}) = \frac{1}{r}\left[\frac{1}{2}\frac{(1+\overline{p})^2}{1+ar\sigma^2} - \overline{p} - \frac{1}{2}\overline{p}^2\right].$$

Similarly, under the deterministic policy,

$$V^d(\overline{p}^d) = H_2(\overline{p}^d) = \frac{1}{r}\left[\frac{1}{2}\frac{(1+\overline{p}^d)^2}{1+ar\sigma^2} - \overline{p}^d - \frac{1}{2}(\overline{p}^d)^2\right].$$

It is straightforward that $H_2'(p) = \frac{1}{r}\left[\frac{1+p}{1+ar\sigma^2} - 1 - p\right] < 0$, *and hence to show* $\overline{p} \leq \overline{p}^d$, *it is enough to prove that* $V(p) > V^d(p)$. *Let's define function*

$$H(p) \equiv V(p) - V^d(p),$$

then the ODE equation (66) implies

$$rH(p) - p(\phi + r) H_p(p) = \frac{1}{2}\left[\frac{(1+p-\phi V_p(p))^2}{1+ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2(p)}{V_{pp}(p)}} - \frac{\left(1+p-\phi V_p^d(p)\right)^2}{1+ar\sigma^2}\right]. \tag{86}$$

It follows from $\hat{x} = \frac{V_{pp}(0)-V_{pp}^d(0)}{1/\phi+A^d} > 0$ *that* $V_p(0+) > V_p^d(0+)$, *and hence initially* $V(p)$ *is above* $V^d(p)$, *i.e.,* $H(0+) > 0$ *and* $H(0) = 0$.

We show that $\overline{p} < \overline{p}^d$. *Suppose it is not true so that* $\overline{p} > \overline{p}^d$. *Since* V *is concave over* $[0, \overline{p}]$,

$$V_p\left(\overline{p}^d\right) > V_p(\overline{p}) = 0 = V_p^d\left(\overline{p}^d\right),$$

and hence,

$$H_p\left(\overline{p}^d\right) = V_p\left(\overline{p}^d\right) - V_p^d\left(\overline{p}^d\right) > 0.$$

In addition, if $\overline{p} > \overline{p}^d$, *then* $H\left(\overline{p}^d\right) = V\left(\overline{p}^d\right) - V^d\left(\overline{p}^d\right) < V(\overline{p}) - V^d\left(\overline{p}^d\right) = H_2(\overline{p}) - H_2\left(\overline{p}^d\right) < 0$. *Because* $H(0+) > 0$, *it must be that* V *crosses* V^d *at some* $\widehat{p}$ *before* $\overline{p}^d$, *so that*

$$H(\widehat{p}) = 0, \text{ and } H_p(\widehat{p}) < 0.$$

As a result, there exists another point $p_1 \in \left[\widehat{p},\ \overline{p}^d\right]$ *so that*

$$H(p_1) < 0,\ H_p(p_1) = 0 \text{ which implies that } V_p(p_1) = V_p^d(p_1).$$

However, it follows from equation (86) that at $p = p_1$,

$$\begin{aligned} rH(p_1) &= \frac{1}{2}\left[\frac{(1+p_1-\phi V_p(p_1))^2}{1+ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2(p_1)}{V_{pp}(p_1)}} - \frac{\left(1+p_1-\phi V_p^d(p_1)\right)^2}{1+ar\sigma^2}\right] \\ &= \frac{(1+p_1-\phi V_p(p_1))^2}{2}\left[\frac{1}{1+ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2(p_1)}{V_{pp}(p_1)}} - \frac{1}{1+ar\sigma^2}\right] > 0, \end{aligned}$$

which contradicts with $H(p_1) < 0$. ■

Step 2.c: Concavity & Positivity of the Denominator at $p < 0$ **:** Here we want to show that, for any $p < 0$,

$$1 + ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2(p)}{V_{pp}(p)} > 0 \text{ and } V_{pp}(p) < 0.$$

Again, we prove these inequalities by contradiction, with similar argument as in Step 2.b. Suppose that the above two inequalities fail to hold at some point. Denote the largest point $\widehat{p} < 0$ at which at least one of the two inequalities does not hold. That is,

$$1 + ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2(p)}{V_{pp}(p)} > 0 \text{ and } V_{pp}(p) < 0 \text{ for } p \in (\widehat{p}, 0]$$

Because $V_{pp}(p) < 0$ for $p \in (\widehat{p}, 0]$ and $V_p(0) = \frac{1}{\phi} > 0$, we have

$$\begin{aligned} V_p(p) &> V_p(0) > 0 \text{ for } p \in (\widehat{p}, 0] \\ V_p(\widehat{p}) &\geq V_p(0) > 0. \end{aligned}$$

Similar to the argument in Step 2.b, one of the three cases below holds at $\widehat{p}$:

$$\text{(Case 1) } 1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = 0 \text{ and } V_{pp}(\widehat{p}) < 0$$

or

$$\text{(Case 2) } 1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = 0 \text{ and } V_{pp}(\widehat{p}) = 0$$

or

$$\text{(Case 3) } 1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} > 0 \text{ and } V_{pp}(\widehat{p}) = 0.$$

We show that all the three cases are impossible below.

Case 1. We show this case is impossible. Suppose this is true. Then it must be true that

$$\frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = -\frac{1 + ar\sigma^2}{a^2r^2\sigma^2}.$$

Moreover, for the value function to be bounded, it must be true that

$$1 + \widehat{p} - \phi V_p(\widehat{p}) = 0.$$

Since $V_p(\widehat{p}) \geq V_p(0)$ the above equation implies

$$\widehat{p} = \phi V_p(\widehat{p}) - 1 \geq 0,$$

which contradicts with $\widehat{p} < 0$.

Case 2. This case is impossible, because $V_p(\widehat{p}) > 0$ and $V_{pp}(\widehat{p}) = 0$ imply $\lim_{p \downarrow \widehat{p}} \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = -\infty$, which contradicts with $1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = 0$.

Case 3. Note that Case 3 occurs only in the situation that

$$V_p(p) \to 0^+ \text{ and } V_{pp}(p) \to 0^- \text{ when } p \downarrow \widehat{p}.$$

Otherwise, a non-zero $V_p(\widehat{p})$ implies $\frac{V_p^2(\widehat{p})}{V_{pp}(\widehat{p})} = -\infty$, a contradiction to the fact that $1 + ar\sigma^2 + a^2r^2\sigma^2 \frac{V_p^2(p)}{V_{pp}(p)} > 0$ for all $\widehat{p} < p < 0$. On the other hand, as shown above,

$$V_p(\widehat{p}) \geq V_p(0) > 0,$$

which contradicts with the fact that $V_p(p) \to 0^+$ when $p \downarrow \widehat{p}$.

Step 2.d: Absorbing lower boundary 0 & entrance-no-exit upper boundary $\overline{p}$:

Lemma 10 *The lower bound $p = 0$ is absorbing. The upper boundary $\overline{p}$ is entry-no-exit, i.e., $\sigma_p = 0$ and $\mu_p < 0$. Therefore, $p_t \in [0, \overline{p}]$ always.*

Proof. We have shown that there exists an upper boundary $\overline{p}$ such that

$$V_p(\overline{p}) = 0.$$

This implies that $\beta(\overline{p}) = \frac{1+\overline{p}}{1+ar\sigma^2}$ and $\sigma^P(\overline{p}) = 0$, and

$$\begin{aligned}
\left.\frac{dp}{dt}\right|_{p=\overline{p}} &= (\phi + r)\, p_t + \beta_t\,(ar\sigma\sigma^p - \phi) \\
&= (\phi + r)\, p_t - \beta_t\phi = \phi\,(\overline{p} - \beta(\overline{p})) + rp_t \\
&= \left(1 - r\overline{p}\frac{1}{\phi}\right)\frac{\phi}{(V_{pp}\phi - 1)}M.
\end{aligned}$$

Since $V_{pp} < 0$, the above equation is negative if and only if $1 - r\overline{p}\frac{1}{\phi} > 0$. Hence, it follows from $A^d > r\left(\frac{1}{\phi}\right)^2$, (where $A^d > 0$ is the coefficient on the quadratic term for the deterministic value function), we have

$$1 - r\overline{p}^d\frac{1}{\phi} = 1 - r\left(\frac{1}{\phi}\right)^2\frac{1}{A^d} > 0.$$

Therefore, it follows from Lemma 9 that $\bar{p} \le \bar{p}^d$, and thus $1 - r\bar{p}\frac{1}{\phi} \ge 1 - r\bar{p}^d\frac{1}{\phi} > 0$.

By a similar argument, we can prove that $p = 0$ is absorbing, since

$$\left.\frac{dp}{dt}\right|_{p=\underline{p}} = \beta_t\,(ar\sigma\sigma^p - \phi) = 0.$$

■

This lemma is important because it implies that V is determined by the policy within the region $[0, \overline{p}]$. Because V is in C^2 and $V_{pp} < 0$ strictly, we know that the policy function

$$\beta(p) = \frac{1 + p - \phi V_p}{1 + ar\sigma^2 + a^2r^2\sigma^2\frac{V_p^2}{V_{pp}}}$$

is bounded. Therefore, we can pick some arbitrary M to bound it.

A.9.3 Step 3: Verification

Define the auxiliary gain process, as a function of the contract Π, to be

$$G_t(\Pi) = \int_0^t e^{-rs}\left((\beta_s - p_s) - \frac{1}{2}(\beta_s - p_s)^2 - \frac{1}{2}a^2r^2\sigma^2\beta_s^2\right)ds + e^{-rt}V(p).$$

Define τ as the hitting time of p reaches $\pm M_p$, which could be infinite. Obviously, $G_\tau(\Pi)$ is the actual payoff from the contract Π. For given t, it is easy to show that

$$\mathbb{E}_t\left[e^{rt}dG_t\right] = \left[\begin{array}{c} -rV(p) + (\beta_t - p_t) - \frac{1}{2}(\beta_t - p_t)^2 - \frac{ar\sigma^2}{2}\beta_t^2 \\ +V_p\left[(\phi + r)\, p_t + \beta_t\left(ar\sigma\sigma^P - \phi\right)\right] + \frac{1}{2}V_{pp}\left(\sigma_t^P\right)^2 \end{array}\right]dt + V_p\sigma_t^P dB_t.$$

Therefore,

$$dG_t = \mu_G(p)\,dt + e^{-rt}V_p\sigma_t^P dB_t.$$

Due to construction of ODE in the HJB equation, under the optimal policy Π^* we have $\mu_G(p) = 0$, while for other policies we have $\mu_G(p) \le 0$. Also, we restrict the policy $\left\{\sigma_t^P\right\}$ to be well-behaved (square integrable in the usual sense) so that $\int_0^\infty e^{-rt}V_p\sigma_t^P dB_t$ is a martingale. Therefore, under the optimal contract $\mathbb{E}\left[G_\tau(\Pi^*)\right] = G_0(\Pi^*) = V(p_0)$.

Given any $T > 0$, we have

$$\begin{aligned}\mathbb{E}\left[G_\tau(\Pi)\right] &= \mathbb{E}\left[G_{T\wedge\tau}(\Pi) + \mathbf{1}_{T\leq\tau}\left(\int_T^\tau e^{-rs}\left((\beta_s - p_s) - \frac{1}{2}(\beta_s - p_s)^2 - \frac{1}{2}a^2r^2\sigma^2\beta_s^2\right)ds + e^{-r\tau}V(p_\tau)\right)\right] \\ &\leq G_0 + e^{-rT}\mathbb{E}\left[\int_T^\tau e^{-r(s-T)}\frac{1}{2}ds\right]\end{aligned}$$

where $\mathbb{E}\left[\int_T^\tau e^{-r(s-t)}\frac{1}{2}ds\right]$ is the first-best project value. Therefore, let $T \to \infty$, we have $\mathbb{E}\left[G_\tau(\Pi)\right] \leq G_0 = V(p)$. This implies that the proposed contract solves the relaxed problem.

A.9.4 Step 4: $\bar{p}$ is independent of M

We now show that $\bar{p}$ is independent of M. Take some sufficiently large M_1, and consider the solution obtained with the upper entry-no-exit boundary $\bar{p}_1$. Note that $\bar{p}_1 < M_1$ strictly, because $V(\bar{p}_1; M_1) > 0$ while at $M_{p,1} = \left(1 + \frac{r}{\phi}\right)M_1$ the value is strictly negative. And, for $p \in [0, \bar{p}_1]$, we have

$$V(p; M_1) = \mathbb{E}\left[\int_0^\infty e^{-rt}\left((\beta_t - p_t) - \frac{(\beta_t - p_t)^2}{2} - \frac{1}{2}a^2r^2\sigma^2\beta_t^2\right)dt\middle| p_0 = p\right].$$

It is clear that given $\bar{p}_1$, this function is independent of M_1, because under the optimal policy $p \in [0, \bar{p}_1]$, and M does not affect the flow payoff per se (note that β_t is assumed to be unconstrained in the relaxed problem).

Now consider $M_2 \in (\bar{p}_1, M_1)$. The next lemma follows.

Lemma 11 *$V(p; M_1) \geq V(p; M_2)$ for $p \in [0, \bar{p}_1]$.*

Proof. *Denote the corresponding M_p's as $M_{p,i}$. Since $M_1 > M_2$ and $M_{p,1} > M_{p,2}$, the policy space for the semi-constrained problem with M_1 is strictly larger than the policy space in the problem with M_2. To see this, note that for the problem 1 (with M_1), the principal can choose $\beta_s = \left(1 + \frac{r}{\phi}p\right)$ for $s > t$ once $p_t \in [M_{p,2}, M_{p,1}]$, which is exactly the constraint for the policy space of problem 2 with M_2. As a result, $V(p; M_2) \leq V(p; M_1)$ for $p \in [0, \bar{p}_1]$.* ■

However, given M_2, consider the exact same policy under M_1 with endogenous upper entry-no-exit boundary $\bar{p}_1$, which generates the same value as $V(p; M_1)$. As a result, the policy under M_1 also solves the problem with M_2. Therefore we must have the same solution for both M_i's, and $\bar{p}_1 = \bar{p}_2$. QED.

A.10 Proof for Theorem 1

We now show that the relaxed problem solves the original problem. As explained before, our original problem in equation (39) has more stringent constraints than the relaxed problem: in the original problem we require $\beta_t \leq M$ always, while for the relaxed problem we only require that $\beta_t \leq M$ whenever p_t hits $\pm M_p$. As a result, we have $V(p) \geq V^C(p)$ always. Here, $V^C(p)$ denotes the value for the principal's original problem.

To show our theorem, it suffices to show that in the region $[0, \overline{p}]$, we have $V(p) = V^C(p)$, i.e., the relaxed problem and the original problem achieve the same value for sufficiently high M. Take the solution $V(p)$ and its corresponding incentive policy $\beta^M(\cdot)$; and define

$$B(M) \equiv \max_{0\leq p\leq\bar{p}}\left|\beta^M(p)\right|,$$

where $\beta^M(p)$ emphasizes the possibility of the dependence of the optimal policy on the parameter value M. If we can show that we can choose sufficiently high M so that $B(M) \leq M$ holds, then the additional

constraints are never binding in the original problem, and then both problems share the same solution obtained in Proposition 5.

We show that $B(M)$ is independent of M for sufficiently high M, which immediately implies our result. To show this, it is sufficient to show that both the relaxed value function $V(p)$ and $\bar{p}$ are independent of M. We have shown that $\bar{p}$ is independent of M when M is sufficiently high. Moreover, since the endogenous state p never go outside the region $[0, \bar{p}]$, we have

$$V(p) = \mathbb{E}_0 \left[\int_0^\infty e^{-rt} \left((\beta_t - p_t) - \frac{(\beta_t - p_t)^2}{2} - \frac{1}{2} a^2 r^2 \sigma^2 \beta_t^2 \right) dt \middle| p_0 = p \right]$$

to be independent of M. As a result, $\max_{0 \le p \le \bar{p}} \left| \beta^M(p) \right|$ is independent of M, and our result follows.

B Appendix B: Non-stationary Learning

B.1 Formulating the problem

Suppose that the unknown parameter θ is a constant. Define $\phi_t \equiv \Sigma_t^\theta / \sigma^2$, hence

$$dm_t = \Sigma_t^\theta \frac{dY_t - (\mu_t + m_t)\, dt}{\sigma^2} = \phi_t \sigma \frac{dY_t - (\mu_t + m_t)\, dt}{\sigma} \equiv \phi_t \sigma dB_t^\mu, \text{ and } \phi_t = \frac{\phi_0}{1 + \phi_0 t}, \tag{87}$$

where we still use ϕ_t without risk of confusion. According to Prat and Jovanovic (2012), the agent's incentive-compatibility constraint is

$$\begin{aligned} \mu_t &= \beta_t - \mathbb{E}_t \left[\int_t^T \phi_s \beta_s e^{-r(s-t)} e^{\left(- \int_t^s a r \beta_u \sigma dB_u - \frac{1}{2} \int_t^s a^2 r^2 \beta_u^2 \sigma^2 du \right)} ds \right] \\ &= \beta_t - \int_t^T \phi_s \beta_s e^{-r(s-t)} ds, \end{aligned}$$

where in the second equation we invoke the restriction that $\{\beta\}$ are deterministic. Thus, the principal's problem can be written as (where T can take a value of infinity)

$$\begin{aligned} & \max_{\{\beta_t\}} \int_0^T e^{-rt} \left(\mu_t - \frac{1}{2} \mu_t^2 - \frac{1}{2} a r \sigma^2 \beta_t^2 \right) dt \\ s.t. \quad & \mu_t = \beta_t - \int_t^T \phi_s \beta_s e^{-r(s-t)} ds. \end{aligned}$$

Define $p_t \equiv \int_t^T \phi_s \beta_s e^{-r(s-t)} ds = e^{rt} \int_t^T \phi_s \beta_s e^{-rs} ds$ so that

$$\mu_t = \beta_t - p_t,\ p_t' = -\phi_t \beta_t + r p_t \Rightarrow \beta_t = -\frac{1}{\phi_t} (p_t' - r p_t).$$

Under this transformation, the objective becomes $\max_{\{p_t\}} \int_0^T L(t, p_t, p_t')\, dt$, so that

$$L(t, p_t, p_t') \equiv e^{-rt} \left(-\frac{1}{\phi_t} (p_t' - r p_t) - p_t - \frac{1}{2} \left(\frac{1}{\phi_t} (p_t' - r p_t) + p_t \right)^2 - \frac{a r \sigma^2}{2} \left(\frac{1}{\phi_t} (p_t' - r p_t) \right)^2 \right).$$

with the constraint that $p_T = 0$.

B.2 Euler equation and its sufficiency for optimality

We have a standard problem of calculus of variation, and the Euler equation for this problem is:

$$L_p\left(s,p_s,p_s'\right)=\frac{dL_{p'}\left(s,p_s,p_s'\right)}{ds}. \tag{88}$$

Because

$$\begin{aligned}
L_p\left(s,p_s,p_s'\right) &= e^{-rs}\left(\frac{r}{\phi_s}-1+\left(\frac{1}{\phi_s}\left(p_s'-rp\right)+p_s\right)\left(\frac{r}{\phi_s}-1\right)+ar\sigma^2\left(\frac{1}{\phi_s}\left(p_t'-rp\right)\right)\frac{r}{\phi_s}\right),\\
L_{p'}\left(s,p_s,p_s'\right) &= e^{-rs}\left[-\frac{1}{\phi_s}-\frac{1}{\phi_s}\left(\frac{1}{\phi_s}\left(p_s'-rp\right)+p_s\right)-ar\sigma^2\left(\frac{1}{\phi_s}\right)^2\left(p_s'-rp_s\right)\right],
\end{aligned}$$

conducting algebraic simplifications on the Euler equation (88) yields

$$0=-\frac{1}{\phi_s}\left(2p_s'-rp+\frac{1}{\phi_s}\left(p_s''-rp_s'\right)\right)-2ar\sigma^2\left(\frac{1}{\phi_s}\right)\left(p_s'-rp_s\right)-ar\sigma^2\left(\frac{1}{\phi_s}\right)^2\left(p_s''-rp_s'\right).$$

Therefore, from the above equality, we obtain that the optimal policy must satisfy:

$$\frac{1}{\phi_s}p_s''=-2p_s'+\frac{1}{\phi_s}rp_s'+rp\left(1+\frac{ar\sigma^2}{1+ar\sigma^2}\right) \tag{89}$$

Combining with $p_T=0$, one can solve the optimal path using initial condition p_0. Then maximizing over p_0 one can find the solution to the original problem.

Before we proceed to prove the main result in the next subsection, we further show that the first-order optimality of the Euler equation is sufficient for global optimality. Notice $L_{pp}<0$ and $L_{p'p'}<0$. Furthermore, for any x and y, we aim to show that

$$L_{pp}y^2+2L_{pp'}xy+L_{p'p'}x^2\leq 0, \tag{90}$$

where the equality holds only for $x=y=0$. Notice that

$$\begin{aligned}
e^{rs}\left[L_{pp}y^2+2L_{pp'}xy+L_{p'p'}x^2\right] &= -\left(\left(\frac{r}{\phi_s}-1\right)^2+ar\sigma^2\left(\frac{r}{\phi_s}\right)^2\right)y^2\\
&+2xy\left(\left(\frac{1}{\phi_s}\left(\frac{r}{\phi_s}-1\right)+ar\sigma^2\left(\frac{r}{\phi_s}\right)^2\right)\right)-x^2\left(1+ar\sigma^2\right)\left(\frac{1}{\phi_s}\right)^2.
\end{aligned}$$

Thus, it is sufficient to show the following:

$$\left[\left(\frac{r}{\phi_s}-1\right)+ar\sigma^2\frac{r}{\phi_s}\right]^2<\left[\left(\frac{r}{\phi_s}-1\right)^2+ar\sigma^2\left(\frac{r}{\phi_s}\right)^2\right]\left(1+ar\sigma^2\right).$$

The left-hand-side is $\left(\frac{r}{\phi_s}\left(1+ar\sigma^2\right)-1\right)^2$, while the right-hand-side is

$$\begin{aligned}
&\left[\left(1+ar\sigma^2\right)\left(\frac{r}{\phi_s}\right)^2-\frac{2r}{\phi_s}+1\right]\left(1+ar\sigma^2\right)\\
=\;&\left(1+ar\sigma^2\right)^2\left(\frac{r}{\phi_s}\right)^2-\frac{2r}{\phi_s}\left(1+ar\sigma^2\right)+1+ar\sigma^2\\
=\;&\left(\frac{r}{\phi_s}\left(1+ar\sigma^2\right)-1\right)^2+ar\sigma^2>LHS.
\end{aligned}$$

Thus, equation (90) holds, and the Euler equation is both necessary and sufficient for optimality (Theorem 2.3 in Chapter one of Fleming and Rishel (1975)).

B.3 Time-decreasing effort policy

The following sequence of argument shows that the optimal initial information rent $p_0^* > 0$, and the associated effort policy μ_t^* is decreasing over time.

Step 1. From the Euler equation we show that p never changes signs.

If $p(0) = p_0 > 0$ and $p(T) = 0$, but p turns negative somewhere in between. Then there must exist some point $\widehat{t}$ so that

$$p\left(\widehat{t}\right) < 0,\ p'\left(\widehat{t}\right) = 0 \text{ but } p''\left(\widehat{t}\right) \geq 0$$

This contradicts with the Euler equation (89):

$$p''\left(\widehat{t}\right)\left(\frac{1}{\phi_{\widehat{t}}}\right) = rp\left(\widehat{t}\right)\left(1 + \frac{ar\sigma^2}{1 + ar\sigma^2}\right) < 0.$$

Similarly, we can show that if $p(0) < 0$ and $p(T) = 0$ then $p < 0$ always. Therefore p never changes sign.

Step 2. We now prove that if $p_0 > 0$ then the optimal effort policy goes down with time. From equation (89), and $\mu_t = \beta_t - p_t = -\frac{1}{\phi_t}\left(p_t' - rp_t\right) - p_t$, we have

$$\mu_t' = -\frac{1}{\phi_t}p_t'' + \frac{1}{\phi_t}rp_t' - 2p_t' + rp_t = -rp_t\frac{ar\sigma^2}{1 + ar\sigma^2} < 0,$$

since $p_t > 0$ always. Here we also used the fact that $\left(\frac{1}{\phi_t}\right)' = 1$ from equation (87). Similarly, if $p_0 < 0$, the optimal effort policy goes up with time.

Step 3. We show the optimal $p_0 > 0$. First, note that a positive p_0 strictly improve the principal's value over $p_0 = 0$. When $p_0 = 0$, it follows that $p(t) = 0$, $\forall t \in [0, T]$ satisfies the Euler equation, and hence it is optimal with zero value. Now suppose that $\beta_t = \epsilon$ always where $\epsilon > 0$ is sufficiently small, so that

$$\mu_t = \beta_t - \int_t^T \phi_s \epsilon e^{-r(s-t)} ds,\ p_t = \int_t^T \phi_s \epsilon e^{-r(s-t)} ds \text{ with } p_0 = \int_0^T \phi_s \epsilon e^{-r(s-t)} ds.$$

Then the value from this policy must be strictly positive, as the benefit $\int_0^T e^{-rt}\mu_t dt$ is in the order of ϵ while the cost $\int_0^T e^{-rt}\left(-\frac{1}{2}\mu_t^2 - \frac{1}{2}ar\sigma^2\beta_t^2\right)dt$ is in second order of ϵ^2. Because the constant policy might not be optimal, the policy that satisfies the Euler equation should be better, i.e., $V(p_0) > 0$. Finally we rule out $p_0 < 0$ being optimal. To see this, we know that when $p_0 < 0$ then the optimal effort policy $\{\mu_t\}$ is increasing over time. However,

$$\mu_T = -\frac{1}{\phi_T}\left(p_T' - rp_T\right) - p_T = -\frac{1}{\phi_T}p_T' \leq 0.$$

The last inequality follows from the fact that $p_t \leq 0$ for $t < T$ while $p_T = 0$. As a result, the optimal effort policy $\{\mu_t\}$ is negative always. Thus, if $p_0 < 0$, then the objective $\int_0^T e^{-rt}\left(\mu_t - \frac{1}{2}\mu_t^2 - \frac{1}{2}ar\sigma^2\beta_t^2\right)dt$ is negative as well.

Combining the above three steps, we complete the proof that the optimal effort policy is decreasing over time.

C Appendix C: Asymptotic Analysis

C.1 Asymptotic analysis for a risk-tolerant agent

To achieve more analytical tractability, we perform an asymptotic analysis for agents who are sufficiently risk tolerant (relatively small a). We first establish the result for a risk-neutral agent, i.e., $a = 0$. In this

case, the optimal deterministic policy obtained in Section 5.2 achieves first-best (and hence is optimal among all possible contracts).

Lemma 12 *When $a = 0$, the deterministic policy in Proposition 3 is optimal with the first-best effort level $\mu^o = 1$. The corresponding value function, denoted by $V^o(p)$, is given by*

$$V^o(p) = -\frac{1}{2}A^o p^2 + B^o p, \tag{91}$$

where $A^o \equiv r/\phi^2$ and $B^o \equiv 1/\phi$, and the optimal information rent $\overline{p}^o$ satisfies $V_p^o(\overline{p}^o) = 0$ so that

$$\overline{p}_t^o = \overline{p}^o = \frac{B^o}{A^o} = \frac{\phi}{r}. \tag{92}$$

Proof. *Because the first-best level is achieved by the proposed contract, the result is trivial.* ■

When the agent is risk averse (i.e, $a > 0$), we consider asymptotic expansions around the benchmark case $a = 0$. More specifically, denote $a_r \equiv ar$, and we solve for the expansions in the following form:

$$V(p;a) = Q_0(p) + \textstyle\sum_{i=1}^{I} a_r^i Q_i(p) + o\left(a_r^I\right), \tag{93}$$

$$\overline{p} = q_0^* - \textstyle\sum_{j=1}^{J} a_r^j q_j^* + o\left(a_r^J\right), \tag{94}$$

where the highest expansion orders I and J are positive integers to be chosen later. We have $Q_0(p) = V^o(p)$ as given in equation (91), and $q_0^* = \overline{p}^o$ as given in equation (92). We solve for $Q_i(p)$ and $q_i^*(p)$ in closed-form (i.e., they are not asymptotic expansions). Also, given $V(p,a)$ it is easy to derive the optimal control pair $\{\beta(p), \sigma^p(p)\}$ using equation (42).

The optimal deterministic contract studied in Proposition 3 is helpful in deriving Q_i's in equation (93) and q_i^*'s in equation (94). Indeed, by expanding both $V^d(p;a)$ and $\overline{p}^d = B^d/A^d$ in the forms of equation (93), we can compare the optimal deterministic policies to those optimal stochastic ones. We choose the expansion orders I and J to be the lowest orders so that these two expansions start to differ. For instance, we choose $J = 3$ because the asymptotic analysis reveals that the upper boundary $\overline{p}$ in the stochastic optimal contract starts to differ from the deterministic counterpart $\overline{p}^d$ in the third order a_r^3. The next proposition summarizes our results.

Proposition 6 *When a is relatively small, we have the following approximations:*

$$\overline{p} = \overline{p}^d - a_r^3 \frac{\phi^5\sigma^4}{r^6}\left(1 + \frac{r}{\phi}\right)^3 + o\left(a_r^3\right), \tag{95}$$

$$V(\overline{p}) = V^d\left(\overline{p}^d\right) + a_r^4\left[\frac{\phi^6\sigma^6}{r^8}\left(1 + \frac{r}{\phi}\right)^5 + \frac{\phi^5\sigma^6}{r^7}\left(1 + \frac{r}{\phi}\right)^3\right] + o\left(a_r^4\right). \tag{96}$$

Furthermore, we show that

$$\beta(p) = \beta^d(p) - a_r^2\frac{\sigma^2}{\phi}\left(1 + \frac{r}{\phi}\right)p\left(p - \overline{p}^o\right)\left[\left(p - \overline{p}^o\right) + \left(1 + \frac{r}{\phi}\right)p\right] + o\left(a_r^2\right), \tag{97}$$

$$\sigma^p(p) = a_r\sigma\left(1 + \frac{r}{\phi}\right)\left(1 + a_r\frac{\phi\sigma^2}{r}\right)p\left(\overline{p} - p\right) + o\left(a_r^2\right). \tag{98}$$

Proof. *Available upon request.* ■