Executive Compensation in a Changing Environment: Cautious Investors and Sticky Contracts

Thomas F. Cosimano†, Adam Speight‡, and Hayong Yun§

November 2011
Executive Compensation in a Changing Environment:
Cautious Investors and Sticky Contracts

Abstract

We study dynamic incentive contracts in a continuous-time agency model with productivity switching between two unobserved states, about which an investor may learn by deviating from the myopically optimal action. The optimal contract balances short-run profits from myopic actions and the long-run benefits from obtaining information on the current state. We endogenously characterize the stickiness of contracts and show that the optimal contract offered by the investor can substantially deviate from the myopic solution when strategic collection of information is possible. The results provide a rational explanation for the commonly observed low pay-for-performance sensitivity and sluggish adjustment of executive compensation.

Journal of Economic Literature Classification Numbers: C61, D82, D83, E32, G34
Key Words: Executive compensation, optimal contracts, moral hazard, strategic experimentation.
This paper explores equilibrium explanations for the commonly observed passive behavior of investors, e.g., low pay-performance sensitivity and slow adjustment of incentive contracts. We consider dynamic incentive contracts in a continuous-time moral hazard framework with parameter uncertainties regarding underlying economic conditions. In this model, a representative investor needs to hire a manager, who possesses unique skills for production, and cannot base the contract on the manager's costly effort that improves the firm's profit. The optimal level of effort depends on the current state of the economy (e.g., boom or bust, investor sentiment, consumer preferences), which switches randomly and is observed by neither the principal nor the agent. However, the investor can improve his knowledge about current economic conditions by experimentation, that is, by deviating from a myopically optimal contract that just maximizes the investor’s current payoff. As a result, the optimal incentive contract reflects tradeoffs between the short-run maximization of current payoffs and the long-run benefits of gaining information on model uncertainty.

While incentive contracts are central to mitigating conflicts of interest between managers and shareholders, and business conditions are continually subject to change (e.g., business and product cycles), studies on optimal incentive contracts that adapt to changing business conditions are rare.\textsuperscript{1} We are interested in a number of issues related to the role of changes in economic conditions and the strategic behavior of agents on optimal incentive contracts. For example, how does an optimal incentive contract under experimentation differ from a myopic one? Why do they differ? And when are these differences large?

Our model has two main implications for optimal executive compensation policies: cautious behavior by investors and the persistence of contracts. First, when there is a value to

\textsuperscript{1} A notable exception is Eisfeldt and Rampini (2008); they consider the relationship between managerial incentives and capital allocation across business cycles. However, the parameters of their model are known to agents. Hence, they do not consider the impact of learning or experimentation.
learning about model uncertainty, i.e., the current state of business in this model, investors may exhibit cautious behavior. That is, the optimal contract offered by investors is less sensitive to the current outcome than a myopic contract when parameter uncertainties are present. Second, the model endogenously characterizes the persistence of contracts based on a factor, the signal-to-noise ratio. This factor is proportional to the intensity of state switching and the precision (inverse of volatility) of the underlying stochastic process.

Based on this contract inertia, we identify two distinct experimentation regimes. In a moderate experimentation regime, when contract inertia is large, the incremental benefit of new information is moderate and investors set incentive contracts to be closer to purely myopic contracts. In an extreme experimentation regime, when signal-to-noise ratio is large (or contract inertia is small), the incremental benefit of new information is very large and investors switch incentive contracts between corner solutions (extreme values implied by full belief in boom and bust). High contract inertia (e.g., consistent contract) is most frequently encountered in developed countries where economic uncertainties are moderate, and hence contracts are persistent. The model also predicts the possibility of extreme adjustments of optimal contracts when the marginal value of learning is very large due to great uncertainty in underlying economic conditions, e.g., a financial crisis or emerging economies.

Increasingly, the literature questions the appropriateness of current executive compensation practices. For example, Jensen and Murphy (1990) note that a typical CEO in their sample earns only $3.25 per $1000 increase in the shareholders’ wealth. Bebchuk and Fried (2004) argue that executive compensation is not closely tied to executives' performances. Many of these works suggest that a poorly designed corporate governance system is mainly responsible for this low pay-performance sensitivity, and call for reform of the current system. However, as
Weisbach (2007) notes, strong managerial control over dispersed shareholders (and the resulting low pay-performance sensitivity) is one of the corporate governance elements that has persisted for a long time, and the modern dispersed corporation has been responsible for the vast growth of this economy throughout modern history. In line with Weisbach's argument, we propose a rational explanation for the cautious behavior of shareholders, which leads to lower pay-performance sensitivity even though the manager cannot manipulate shareholders' knowledge.

From a broader perspective, our model provides a possible explanation for the frequently-observed lax activism of investors. Fundamentally, we address the question of why investors maintain the current institutional form, e.g., a passive board and low pay-performance sensitivity, rather than seeking tighter control of management, e.g., an active board with frequent monitoring and high pay-performance sensitivity, despite its imperfections in investor protection. We find that in a world with a changing economic environment, investors face a tradeoff between controlling managers' opportunistic behavior and learning about how their actions impact the firm. We also find that when uncertainties are present, contracts become stickier and are not responsive to new information (e.g., current firm performance). As a result, the optimal action by the investor may often be a delayed or mitigated response to managerial misbehavior, which may appear as inefficient and lax activism from a perfectly-identified model optimization viewpoint.

There is a growing literature on dynamic optimal contracting models, including work by Holmstrom and Milgrom (1987), Green (1987), Spear and Srivastava (1987), Phelan and Townsend (1991), and Atkeson (1991), among others. In recent years, a number of papers have focused on the discrete-time dynamic moral hazard models of firms such as Alberquerque and Hopenhayn (2004), Quadrini (2004), Tchistyi (2005), Clementi and Hopenhayn (2006), and DeMarzo and Fishman (2007a, 2007b). These discrete-time moral hazard models have been
extended to a continuous-time framework by Sannikov (2008), Williams (2008) and He (2009). Williams (2008) provides a general discussion of moral hazard models with hidden actions and/or states. Especially, Sannikov (2008) provides a general framework for continuous-time agency problems in the context of executive compensation. Sannikov shows that an agent's continuation value is the natural choice of a state variable, which avoids keeping track of past realizations of outcomes. He (2009) extends Sannikov's (2008) Brownian motion cash flow to geometric Brownian motion, and shows how both the agent's continuation value and the cash flow rate emerge as natural choice of state variables. Prat and Jovanovic (2010) consider optimal long-term contracts in which the prior for the investor and manager can deviate, so that it is possible for the manager to obtain a contract less sensitive to pay for performance by manipulating the shareholder's information. In their work, the productivity of the manager is constant but unknown and productivity is dependent on the cumulative effort of the agent. As a result, cumulative effort influences the manager’s prior about the unknown productivity. In this case, the unknown parameters are eventually revealed, but agents can influence the learning process by changing their hidden actions. In addition, the effort of the manager is a deterministic function of the precision of the estimate of the company’s revenue.

Our work differs from Prat and Jovanovic (2010) in that the unknown parameters, which depend on an unobserved switching regime, are never revealed. In addition, the revenue of the firm is influenced by the current actions of the managers, but neither the shareholders nor the manager know what influence the manager’s actions will have, since this effect is determined by the unknown parameters. Thus, the current actions of the manager have an uncertain influence on both the drift and volatility of the shareholders' prior beliefs about the state of business through observed revenue. As a result, our model is able to endogenously characterize two qualitatively
distinct regimes based on the extent to which optimal contracts incorporate new information, i.e., extreme and moderate experimentation regimes. Whereas Prat and Jovanovic’s (2010) setting can be suitable for long-term employment contracts, where the CEO’s ability is eventually revealed even without learning, our setting is appropriate for studying short-term incentive contracts that adapt to frequent changes in economic conditions, e.g., the business cycle or the popularity of the firm’s product.

Our work contributes to the literature by introducing strategic experimentation in dynamic incentive contracts. While strategic experimentation models have widely been studied in both macroeconomic policy and non-strategic microeconomic problems, e.g., Aghion, Bolton, Harris, and Jullien (1991), Balvers and Cosimano (1990, 1993, 1994), Bolton and Harris (1999), Cosimano (2008), Easley and Kiefer (1988), Kiefer and Nyarko (1989), Keller and Rady (1999), and Wieland (2000), they have not yet been studied in a moral hazard setting. A frequent finding in the strategic experimentation literature is the cautious behavior of the agent or social planner, who foregoes short-run gains to learn about uncertainties, obtaining information that enhances future long-run gains. In a moral hazard setting, we find cautious behavior from the principal, who sets less aggressive contracts, e.g., ones with low pay-performance sensitivity, to gain information on the model uncertainty, i.e., the state of business conditions.

The organization of the paper is as follows. Section I explains the role of learning in the contracting problem through a 2-period example. Section II describes the setup of our model and derives a Hamilton-Jacobi-Bellman equation for the optimal contract of the strategic

---

2 Notable exceptions are Jain, Jeitschko, and Mirman (2002) and Manso (2011), both of whom study the impact of experimentation in a two-period agency setting. Jain, Jeitschko, and Mirman (2002) study financial contracting between a lender and a borrower, whose profits are affected by the number of competitors in an oligopolistic market. In this setting, they show that the lender can experiment with a short term (1-period) contract that impacts the probability of the entry of a competitor in the second period. Manso (2011) shows forgiving contracts can benefit firms by providing managers incentives to take more innovative business strategies. DeMarzo and Sannikov (2008) consider a steady state learning problem in a dynamic contracting framework. The closest work to ours is Prat and Jovanovic (2010); the significant differences between our work and theirs are described above.
experimentation problem under moral hazard. Section III discusses local properties of the compensation problem. Section IV discusses the implications of our model. Section V concludes.

I. The Role of Strategic Learning in Incentive Contracts: A Two-Period Analogy

In this section, we motivate the key findings of this paper – the role of learning in a moral hazard setting – with a simple example before introducing the formal dynamic contracting model. While endogenous characterization of the strategic use of learning requires a dynamic model, we explicitly assume a tradeoff between learning and myopic (short-run) profit maximization in a two-period setting. With this simple model, we focus on the increased returns from learning leading to a convex portion of the principal’s objective function, which leads to qualitative changes in the strategic use of learning through incentive contracts.

Consider a principal who hires an agent engaged in the production of a single indivisible good $\bar{X}$ in two periods, $t \in \{0,1\}$. The project in period 0 is long-term and matures in period 1, and the project in period 1 is short-term and matures in the same period. Assume that $\bar{X}$ takes on only two values bad (0) or good ($\theta > 0$) so that $\bar{\theta} \in \{0, \theta\}$. (See the decision tree in Figure 1.) At the beginning of each project, the manager supplies costly (unverifiable) effort, $e \in [0,1]$, according to the cost function, $C(e) = \frac{1}{2} e^2$. Effort affects output by increasing the probability of a good outcome, $\bar{X} = \bar{\theta}$, i.e., $\Pr(\bar{X} = \bar{\theta}) = e$. The products from both periods mature and are collected at $t = 1$. Initially, both investor and manager believe that high ($\bar{\theta} = \theta$) and low ($\bar{\theta} = 0$) states are equally likely, i.e., $\Pr(\bar{\theta} = 0) = \Pr(\bar{\theta} = \theta) = \frac{1}{2}$.

\footnote{Note that this is a simplifying assumption to prevent agents from perfectly observing the true $\bar{\theta}$ at $t = 0$ when $\bar{\theta} = \theta$. The dynamic incentive contracting model in the following section relaxes this, as well as other restrictive assumptions used in this section.}
After the effort allocation decision is made for the long-term project at $t = 0$, a verifiable signal ($\bar{s} \in \{s_B, s_G\}$) about the true state of the world is received, and its information content is inversely related to the effort spent at $t = 0$ (Figure 1). That is,

$$\Pr(\bar{s} = s_G|\bar{\theta} = \theta) = \Pr(\bar{s} = s_B|\bar{\theta} = 0) = 1 - e$$
$$\Pr(\bar{s} = s_G|\bar{\theta} = 0) = \Pr(\bar{s} = s_B|\bar{\theta} = \theta) = e.$$ 

The signal is an abstraction of the reality that the investor and manager both learn about the quality of the project ($\bar{\theta}$) through experiences obtained by conducting the long-term project at $t = 0$. This information structure causes the tradeoff between achieving immediate higher output and obtaining more accurate information that can be used in the next period’s production. The manager is rewarded with a wage based on the realized quality of the output, $w = b\bar{X}$, where $b$, the contract offered by the investor in each period, can be considered as pay-performance sensitivity. The discount factor for the second period output is $\delta$.

In this setting, we need to find the optimal contract, $\{b_0, b_G, b_B\}$, and the corresponding optimal effort $\{e_0, e_G, e_B\}$. $b_0, b_G, b_B$ are contracts offered by the investor at $t = 0$, when $\bar{s} = s_G$, and when $\bar{s} = s_B$, respectively. $e_0, e_G, e_B$ are the corresponding efforts provided by the manager at these states.

Using backward induction of the investor’s and manager’s profits, it is straightforward to show that the optimal contract and the manager’s optimal effort at $t = 0$ are

$$\{b_G, b_B\} = \left\{\frac{1}{2}, \frac{1}{2}\right\}$$
$$\{e_0, e_G, e_B\} = \left\{\frac{\theta (4b_0 - \delta \theta)}{2(4 - \delta \theta^2)}, \frac{(1 - e_0)\theta}{2}, \frac{e_0\theta}{2}\right\}. $$
(Detailed proofs for these solutions are provided in Appendix I.) The investor’s expected profit, which is defined as the expected value of goods less contractual payments to the manager, at \( t = 0 \) is

\[
\Pi_0(b_0) = \frac{\theta^2(4b_0 - \delta\theta)}{4(4 - \delta\theta^2)}(1 - b_0) + \frac{\delta\theta^2}{8}(1 - e_0)^2 + e_0^2,
\]

(1)

which is a convex function in \( b_0 \) when \( \delta\theta^2 > 2 \), and concave if \( \delta\theta^2 \leq 2 \). Intuitively, when the future profit (\( t = 1 \)) is less discounted (i.e., \( \delta \) is large) or the value of good outcome (\( \theta \)) is large, the role of learning becomes important relative to maximizing current (\( t = 0 \)) profit. As shown by Easley and Kiefer (1988), learning has increasing returns to scale, and, hence, the investor’s objective function becomes convex when learning becomes important (i.e., when \( \delta\theta^2 > 2 \)).

Figure 2 illustrates how learning affects optimal contracts for a fixed \( \delta \) and varying \( \theta \). When the impact of learning is very small (i.e., for very small values of \( \theta < \sqrt{4/\delta} \), as shown at Point A in Panel (a)), the principal’s profit function, equation (1), is concave in \( b_0 \), with a unique internal solution for maximizing the principal’s profit (Panel B). This result follows from the decreasing marginal cost of equity. The future benefit of learning always increases effort as measured by \( \delta\theta^2 \). As \( \theta \) increases, learning becomes more important and the principal’s profit function becomes convex in \( b_0 \), as shown in Panels (d) for \( \theta = 1.46 > \sqrt{4/\delta} \), and (e) for \( \theta = 1.50 > \sqrt{4/\delta} \), and corner solutions become the optimal contracts (Points C and D in panels d and e, respectively.) As Panel (a) shows, when \( \theta \) is located near Points C or D, a small change in \( \theta \) can lead to a large change in the optimal contract (\( b_0 \)). In contrast, when \( \theta \) is located near Point A, changes in \( \theta \) will have a small impact on optimal contracts. The former case (denoted as the extreme experimentation regime) is qualitatively distinct from the latter (the moderate experimentation regime), and is caused by the increased influence of learning, which is always
convex in the effort of the manager. When there is an increasing return to effort and the optimal decision is at a corner, the shareholder compares the value of the investment at points C and D. Subsequently, she chooses the value which maximizes the investor’s profits. At some critical value, within the interval $1.46 < \theta < 1.50$, the investor’s profits at C and D are identical. Thus, for a minor change in the value of the investment in the good state there is a dramatic change in the contract, which maximizes the value function, so that the investor switches from the moderate to the extreme experimentation regime. Hence, the optimal contracts response is very sensitive to changes in the underlying parameters related to learning ($\theta$). In the following sections, we develop a dynamic model of optimal contracts under parameter uncertainty, and endogenously derive these two distinct regimes as well as other properties caused by strategic learning.

II. Dynamic Incentive Contracts under Parameter Uncertainty

The manager operates a company that can have either a good or a bad business state. This state could be a reflection of the overall economy or the popularity of the firm’s product. The firm is owned by a risk-neutral investor who does not know the state of the business. The investor would like to design a contract to compensate the manager such that the investor’s profits are maximized. This contract is designed to provide sufficient compensation to minimize opportunistic behavior by the manager. A key aspect of this contract is the degree to which the manager’s compensation responds to the observed revenue generated by the firm. However, the firm faces a tradeoff, since the manager is risk averse and does not want to be subject to too much fluctuation in her compensation.
In this section, we formalize this manager compensation problem and derive the Bellman equation that governs the dynamics of the optimal contract. In the derivation of the Bellman equation, the equation of motions for the investor’s belief and the manager’s continuation value are determined. These equations of motion are then substituted into the investor’s objective function, discounted by flow of revenue, to derive the Bellman equation expressed in terms of the investor’s belief and the manager’s continuation value.

II.A. Model Setup

Consider a continuous-time agency model of a firm, owned by a risk-neutral representative investor (principal) and operated by a manager (agent) with production skills. The revenue, \( R_t \), depends on the manager’s effort, \( A_t \), the current state of business conditions, \( k \), and a shock driven by standard Brownian motion, \( Z_t \):

\[
dR_t = (\alpha_k + \beta_k A_t)dt + \sigma dZ_t
\]

where \( \alpha_k \) and \( \beta_k \) are positive constants that depend on each state, and the volatility, \( \sigma \), determines the magnitude of the Brownian motion.

Following Hamilton (1989), we use a two-state regime-switching model to represent a change in the contracting environment, e.g., the business cycle or customer trend. There are two possible states, \( k \in \{0,1\} \), where \( 0 < \alpha_1 < \alpha_0 \) and \( 0 < \beta_0 < \beta_1 \). Consequently, the marginal value of the manager’s effort to the shareholder is larger when \( k=1 \). The state evolves following a Poisson switching process with a transition density:

\[
f(k_{t+dt} = j | k_t = i) \equiv \lambda_{i,j} = \begin{pmatrix} 1 - \lambda_0 dt & \lambda_0 dt \\ \lambda_1 dt & 1 - \lambda_1 dt \end{pmatrix}.
\]
We assume that both the investor and the manager know the parameter values, $\alpha_k$, $\beta_k$, and $\lambda_k$, $k = 0, 1$, but not the current state $k_t$. The current state must be inferred using observation of the firm’s revenue.

II.B. Beliefs about the State of the Business

The investor has an initial belief about the state that is a function of the prior probability, $\pi_0$, that $k_0 = 1$. The Brownian noise in the revenue equation (2) prevents both the investor and the manager from directly inferring the true state of business. Since the current state, $k_t$, is unobserved, denote $\pi_t \in [0,1]$ to be the best estimate of $k_t$ given the history of observations of revenues.

As a result, the expected intercept and slope for the firm’s revenue is

$$\alpha(\pi_t) = (1 - \pi_t)\alpha_0 + \pi_t \alpha_1 = \alpha_0 + \Delta \alpha \pi_t,$$

$$\beta(\pi_t) = (1 - \pi_t)\beta_0 + \pi_t \beta_1 = \beta_0 + \Delta \beta \pi_t. \quad (4)$$

Here, $\Delta \alpha = \alpha_1 - \alpha_0$, $\Delta \beta = \beta_1 - \beta_0$, $\bar{A} = -\frac{\Delta \alpha}{\Delta \beta}$ is the effort of the manager, such that the investor’s belief about the state of the business does not influence the value of the stock, since

$$dR = (\alpha_0 + \beta_0 A_t + (A_t - \bar{A}) \Delta \beta k) dt + \sigma dZ_t.$$

This result occurs, since the two revenue curves intersect here. Following Keller and Rady (1999), we call this the confounding effort.

The new information available to the investor, given the current state of business, is

$$dZ^I_t = \frac{1}{\sigma}(dR_t - E_{\pi_t}(dR_t)) = \frac{1}{\sigma}((\alpha_k + \beta_k A_t) dt + \sigma dZ_t - (\alpha(\pi_t) + \beta(\pi_t) A_t) dt).$$

$$= \frac{1}{\sigma}((A_t - \bar{A}) \Delta \beta (k_t - \pi_t) + \sigma dZ_t). \quad (5)$$
As a result, the investor’s information is a weighted average of the forecast error of the state of the business, \( k_t - \pi_t \), and the transitory shock, \( dZ_t \). This information consists of only white noise under confounding beliefs, when the manager’s effort is \( \bar{A} = -\frac{\Delta \alpha}{\Delta \beta} \). As the manager’s effort moves away from \( \bar{A} \), the signal contains additional information about the state of business.

Applying Lipster and Shiryaev’s (1977, p.355) Theorem 9.1, we can show the dynamics of the probability of good business conditions, \( \pi_t \), to be

\[
d\pi_t = \lambda(\pi_t)dt + \Sigma(\pi_t, A_t)dZ_t = -\Lambda(\pi_t - \bar{\pi})dt + \Sigma(\pi_t, A_t)dZ_t
\]

(6)

where

\[
\lambda(\pi_t) = (1 - \pi_t)\lambda_0 - \pi_t\lambda_1 = -\Lambda(\pi_t - \bar{\pi})
\]

\[
\Sigma(\pi_t, A_t) = \tau(\pi_t)(A_t - \bar{A}), \quad \tau(\pi_t) = \frac{(1-\pi_t)\pi_t\Delta \beta}{\sigma}
\]

\[
\bar{\pi} = \frac{\lambda_0}{\lambda_0 + \lambda_1} \quad \text{and} \quad \Lambda = \lambda_0 + \lambda_1.
\]

\( \Sigma(\pi_t, A_t) \) is the percentage of the investor’s information that is apportioned to the signal about the state of business rather than the noise. The information has no value when the investor thinks the state of business is at either extreme, \( \pi_t = 0 \) or \( \pi_t = 1 \), or when the manager’s effort is at \( \bar{A} = -\frac{\Delta \alpha}{\Delta \beta} \) (two possible revenue curves cross at this point). In these cases, the investor has no incentive to learn about the state of business. Consequently, \( \Sigma(\pi_t, A_t) \) is called the value of learning and \( \tau(\pi_t) \) the marginal value of effort to learning.\(^4\)

II.C. Incentive Compatibility and the Manager’s Continuation Value

\(^4\)This corresponds to the impact of effort on the probability of a good or bad signal in the static model of section I.
By exerting effort, $A_t$, the manager experiences a cost of effort, $h(A_t) \geq 0$, which is null at zero, smooth, increasing and convex in the manager’s effort. The revenue is observable by both the investor and the manager. The manager’s effort is assumed to be known to the investor, but the contract cannot be based on this effort. Following Grossman and Hart (1986), this assumption reflects a real-life situation where managerial actions may be observed but the implications for a firm’s output are not clear to outsiders. Hence, contracts based on these observed but unverifiable actions cannot be enforced by a third party (e.g., the court). An example of such an action is working capital liquidation. The decision not to replace liquid assets as they are drawn down is made during the routine course of business. While this action is observable, it is difficult to verify what the optimal level of working capital should be.\(^5\) Since manager’s actions are not contractible, the investor must use observations of revenue to determine the manager’s incentives to make costly effort.\(^6\) Specifically, before the manager starts working for the firm, the investor commits to a contract that offers the manager a flow of consumption, $c_t \geq 0$, based on the observed revenue in each period. The manager derives utility from consumption, described by a utility function such that $u(c_t) \geq 0$ with $u(0) = 0$. Utility is assumed to be a smooth increasing, and concave function. The manager discounts the flow of net utility (i.e., utility subtracted by the cost of effort) at a discount rate, $r$.

Consider an arbitrary consumption process, $C = \{c_t\}$, and the manager’s strategy for effort, $A = \{A_t\}$, which may or may not be optimal for the manager. Let $C$ be bounded by

\(^5\) For example, when L.A. Gear liquidated its working capital to finance various comeback strategies, its lead lender, Bank of America, did not put into place covenants restricting the liquidation of working capital because of their informational disadvantage (relative to the firm's management) in determining the optimal level of working capital and because of the difficulties of constraining the use of working capital in a productive way (DeAngelo, DeAngelo, and Wruck (2002)).

\(^6\) Consequently, we rule out hidden information, as discussed in Gayle and Miller (2009), to further the analysis. If information were hidden then we would have to keep track of the managerial effort recommended by the contract and actual effort, as in Prat and Jovanovic (2010). Hidden information is an interesting complication that would require a verification constraint on the manager’s action.
[0, c_{max}] and A bounded by \( \mathcal{A} = [A_{min}, A_{max}] \). A contract is defined as a consumption flow that is adapted to the filtration \( \mathcal{F}_t^M \) generated by observing the returns process (2). The utility the manager derives from consumption, \( c \), under the contract given an effort process, \( A \), is

\[
W_t(c, A) = E_t^M \left[ \int_0^\infty e^{-rt} (u(c_t) - h(A_t)) \, dt \right].
\]

(7)

Here \( E_t^M \) refers to the manager’s expectation under her strategy, \( A \), conditional on, \( \mathcal{F}_t^M \), the filtration generated by the business’s revenue (2). This manager is willing to accept a contract that prescribes effort \( A \) only when the initial value of her net consumption is greater than the manager’s reservation value \( \tilde{W} > 0 \). That is, the contract requires a participation constraint \( \tilde{W} \leq W_0(c, A) \).

Since contracts must depend on effort only through observed returns any “recommended effort” process must satisfy an incentive compatibility condition. Following Sannikov (2008), incentive-compatible contracts are defined in terms of the key pay for performance ratio

\[
\gamma(\pi, a) = \min \{ y \in [0, \infty) : a \in \arg \max_{a \in \mathcal{A}} \, a(\pi) + \beta(a') - h(a') \}.
\]

(8)

The Appendix II shows that a contract is incentive-compatible contract with recommended effort \( A \) if it admits the representation with \( Y = \gamma(\pi, \omega) \).

\[
dW_t = \left( rW_t - u(C_t) + h(A_t) \right) \, dt + \sigma Y_t \, dZ_t,
\]

when \( W_t > 0 \).

(9)

The investor is risk neutral and derives a profit from offering a particular incentive compatible contract \( (c, A) \)

\[
F(\pi, W; c, A) = E_t \left[ \int_0^\omega e^{-rt} (dR_t - c_t \, dt) \, | \pi_0 = \pi, W_0 = W \right]
\]

The investor’s optimization problem consists of choosing a recommended effort level and compensation process \( (c, A) \) that are (i) progressively measurable with respect to the filtration process, \( \mathcal{F}_t^I \), generated by observing the revenue process, \( \{R_s| s \in [0, t]\} \), (ii) stays within the
bounds $\mathcal{A} \times [0, \infty)$, and (iii) lead to square integrable processes $\pi$ and $W$ given by (6) and (9).

This leads to the value function

$$F(\pi, W) = \sup_{(c, A)} F(\pi, W; c, A)$$

Standard arguments from dynamic programming can he used to derive the Hamilton-Jacobi-Bellman (HJB) equation over the domain $(\pi, W) \in [0,1] \times (0, \infty)$,

$$rF(\pi, W) = \max_{(c, A) \in Q = [0, e_{\max}] \times \mathcal{A}} \left\{ -\Lambda (\pi - \tilde{\pi}) \frac{\partial F}{\partial \pi} + \frac{1}{2} \Sigma (\pi, A)^2 \frac{\partial^2 F}{\partial \pi^2} + \Sigma (\pi, A) \sigma \frac{\partial^2 F}{\partial \pi \partial W} ight.$$  

$$+ (\alpha_0 + \beta_0 A + (A - \bar{A}) \Delta \beta \pi - c_I) + (rW - u(c) + h(A)) \frac{\partial F}{\partial W} + \frac{1}{2} \gamma (A) \sigma^2 \frac{\partial^2 F}{\partial W^2} \right\}$$  \hspace{1cm} (10)

To derive equation (10), we apply Ito’s lemma to $dF(\pi, W)$, which requires the first and second derivatives of $F(\pi, W)$ to exist. Based on Balvers and Cosimano (1993), we know this is false at critical points where the value function, $F(\pi, W)$, switches between a concave and convex function (and the optimal solution switches corners).\(^7\) Otherwise the solution will be smooth. Consequently, the Hamilton-Jacobi-Bellman (HJB) equation provides the local property of the solution to the investor’s compensation problem but captures global properties only in a formal sense.\(^8\) Consequently, we cannot use the classical approach to verify the uniqueness of the value function, since it is not smooth enough. Thus, we rely on the theory of viscosity solutions to establish the existence and uniqueness of the value function which is the solution to the HJB equation (10).\(^9\)

The left hand side of the HJB represents the incremental increase in the investor’s value at the risk-free rate. The first term on the right hand side ($-\Lambda (\pi - \tilde{\pi}) \frac{\partial F}{\partial \pi}$) is the change in value due to belief switching. $\Lambda (\pi - \tilde{\pi}) = (1 - \pi) \lambda_0 - \pi \lambda_1$ is the expected drift of belief. If belief of

---

\(^7\) The function has a left and right derivative at these points but they are not necessarily equal.  
\(^8\) The numerical scheme for solving the contract is designed to be flexible enough that these points of discontinuity in the derivative of the value function are accounted for, so that the global solution can be approximated.  
\(^9\) See Fleming and Soner (2006, Ch. VII) and Pham (2009, Ch. IV).
the bad state of business is $1 - \pi$, then $(1 - \pi)\lambda_t dt$ is the probability of switching into the good state, while $\pi\lambda_t dt$ is the chance of switching from the good state to the bad state. As a result, $\Lambda(\pi - \tilde{\pi}) dt$ is the change in the chance that belief will be in the good state, while $\frac{\partial F}{\partial \pi}$ is the shadow price for such a change in belief. The second term, $\frac{1}{2} \Sigma(\pi, A) \frac{\partial^2 F}{\partial \pi^2}$, is the value of information acquired. The value of learning, $\Sigma(\pi, A)$, measures the signal-to-noise ratio from the revenue signal, which determines the percentage of the signal that is attributed to the beliefs of the manager and investor. $\frac{\partial^2 F}{\partial \pi^2}$ is shown to be positive below, so that there are increasing returns to learning about the probability of the good state. The third term, $\Sigma(\pi, A)\gamma(A)\sigma \frac{\partial^2 F}{\partial \pi \partial W}$, deals with the correlation between beliefs about the state of business and the manager’s continuation value. $\gamma(\pi, A)\sigma$ is the standard deviation of the manager’s continuation value, which is correlated with the shock to belief based on the value of learning $\Sigma(\pi, A)$. $\frac{\partial^2 F}{\partial \pi \partial W}$ captures the effect of changes in the manager’s continuation value on the shadow price for a change in belief, which results from this correlation. If this term is positive it reinforces the value of actively acquired information so that the manager’s continuation value is a complement to the value of information, $\Sigma(\pi, A)^2 \frac{\partial^2 F}{\partial \pi^2}$. The last set of terms in the second line represents the myopic payoff from the current period under a moral hazard setting. Prior work by Sannikov (2008) shows that these terms are concave in the manager’s effort, $A$.

There is no retirement in the model. Rather, it is feasible for the investor to choose $A = 0$ and $u(c) = rW$ in order to freeze the state of promised utility and effectively provide an annuity to the agent. By convention, define the value function associated with this policy as $F^{Ret}(\pi, W) = F(\pi, W; u^{-1}(rW), 0)$. At $W = 0$, the pay for performance ratio must be zero, so

$^{10}$ Condition A below assures us that this is true.
that only \( c = 0 \) and \( A = 0 \) is the only incentive compatible contract. Therefore, the boundary condition at \( W = 0 \) is \( F(\pi, 0) = F^{\text{Ret}}(\pi, 0) \).

**Proposition 2.1.** The investor’s optimization problem has a finite value function \( F \geq F^{\text{Ret}} \) with \( F(\pi, W)/F^{\text{Ret}}(\pi, W) \downarrow 1 \) as \( W \uparrow \infty \). Furthermore, the value function is the unique viscosity solution to the HJB (10) equation with the boundary condition \( F(\pi, 0) = F^{\text{Ret}}(\pi, 0) \) and the optimal pair \((c, A)\) given by the policy functions associated with this HJB equation. Retirement occurs the first time \( W_t \) hits zero and is implemented by choosing \( c = 0 \) and \( A = 0 \).

Proof: See Appendix II for proof, which extends Sannikov (2008) analysis to the viscosity approach which must be used in the present case in which there are two state variables. These two state variables are driven by a single Brownian shock with nonlinearity in both the quadratic variation and correlation between the increments of the state variables.

**II.D. Additional Contractual Possibilities**

The model described in the previous subsection gives a solution such that the promised utility tends to become either zero or else drift to a high level where the manager essentially retires, since zero effort is optimal. For time series simulations discussed below, it is desirable to have a model in which there are no absorbing states associated with the manager’s career path. Following Sannikov (2008) and Spear and Wang (2005), we extend the model by allowing the investor to retire and replace a manager at any time. This is accomplished by purchasing an annuity sufficient to provide the promised utility at cost, \( c^{\text{Ret}}(W) = \frac{u^{-1}(\pi W)}{W} \), and making a contract with a young agent that provides the market reservation utility \( \tilde{W} \).

To assure that the value function is decreasing in promised utility, we also allow the
investor to renegotiate the contract in favor of the manager. Here, the investor may always move instantaneously from a state $W \geq 0$ to any state $W' \geq W$. Modifications to the HJB equation (10) that reflect retirement and replacement as well as renegotiation are given in the Appendix III. The numerical results presented below use these modifications.

III. Local Properties of the Compensation Problem

In this section, we examine the optimal policy functions given the uniqueness of the value function. First, examine the optimal consumption decision, following Sannikov (2008). To be specific, suppose the utility of the manager is a constant relative risk aversion (CRRA) utility:

$$u(c) = \frac{c^{1-\chi}}{1-\chi} \text{ such that } \chi \in (0,1).$$

It is straightforward to show that the optimal compensation rule is

$$c = \left(-\frac{\partial F}{\partial W}\right)^{-\frac{1}{\chi}}.$$

As a result, the HJB equation (10) under optimal compensation is

$$rF(\pi, W) = \max_{A \in \mathcal{A}} \left( \alpha_0 + \beta_0 A + (A - \bar{A})\Delta \beta \pi - \left(-\frac{\partial F}{\partial W}\right)^{-\frac{1}{\chi}} \left( \frac{\chi}{1-\chi} \right) \right)$$

$$+ \left( rW(c_t, A_t) + \frac{1}{1-\chi} + h(A_t) \right) \frac{\partial F}{\partial W}$$

$$+ \Lambda(\bar{\pi} - \bar{\pi}) \frac{\partial F}{\partial \pi} + \frac{1}{2} \left\{ \Sigma(\pi_t, A_t)^2 \frac{\partial^2 F}{\partial \pi^2} + 2 \Sigma(\pi_t, A_t) \gamma(A_t) \sigma^2 \frac{\partial^2 F}{\partial \pi \partial W} + \gamma(A_t)^2 \sigma^2 \frac{\partial^2 F}{\partial W^2} \right\}. \quad (11)$$

The HJB equation (11) will not in general have second derivatives defined everywhere, since there are circumstances in which it is optimal for the manager to undertake extreme levels of effort. However, under moderate circumstances the properties of the solution arise from the necessary and sufficient conditions for the maximum problem within the HJB equation (11).
Consequently, a local solution to the manager’s compensation problem, when the state of business is hidden, is the solution to the nonlinear partial differential equation (PDE) (11). This solution gives the value of the firm to the investors when the optimal contract for the manager is implemented. While an explicit solution to this PDE is unknown, we can identify local properties of this solution that will aid us in the interpretation of the numerical global solution to the manager’s compensation problem.

In this section, the properties of the stock price and the optimal effort of the manager are discussed given a global solution, $F(\pi, W)$, to the stock price. In the next section, a robust numerical scheme is used to find the global solution. We can identify these properties by recognizing that the HJB equation is a combination of two elements: first, Sannikov’s (2008) compensation problem for a given belief about the demand, and second, the impact of learning on the optimal contract, as in Keller and Rady (1999). When one understands each part of the complete problem, one can also understand the properties of the optimal managerial compensation contract.

III.A. Myopic Behavior ($\pi$ Is Fixed)

Here, the moral hazard problem without learning is described to establish a baseline for comparison with the effect of learning on the optimal manager’s compensation contract. Suppose that the beliefs are given, so that the investor and manager do not recognize the impact of their actions on their beliefs. This case is called the myopic contract. What would be the optimal choice of the manager’s behavior? Under the myopic contract, the last line of equation (11), which involves partial derivatives with respect to beliefs, $\pi$, is ignored.
To be specific, let’s assume, as in Sannikov (2008), the following function for the cost of the manager’s effort:

\[ h(A) = H_1 A + \frac{1}{2} H_2 A^2. \]

where \( H_1 > 0 \) and \( H_2 > 0 \). Solving the optimal incentive contract following equation (8), we have

\[ a' = \beta(\pi) \gamma - \frac{H_1}{H_2} > 0 \text{ so that } \gamma(a) = \frac{H_2 a + H_1}{\beta(\pi)}. \]

For the lowest effort \( a=0 \), \( \gamma(a) \) is the smallest multiplier which induces zero effort.

By appropriate definitions of constants, both the cost of effort, \( h(A) \), and the volatility of the manager’s contract, \( \gamma(A_t)^2 \), are quadratic functions of \( (A - \bar{A}) \).\(^{11}\) In particular,

\[ h(A - \bar{A}) = K_0 + K_1 (A - \bar{A}) + \frac{1}{2} K_2 (A - \bar{A})^2, \quad \text{and} \]

\[ \gamma(A - \bar{A})^2 = \left( \frac{K_2 (A - \bar{A})^2 + 2 K_1 K_2 (A - \bar{A}) + K}{\beta(\pi)^2} \right). \]

From now on, equations (12) and (13) will represent the cost of effort and the sensitivity of the contract to new information. This allows one to write the myopic problem in terms of \( A - \bar{A} \), which can be compared with the effort under the myopic contract, \( A_t^m - \bar{A} \). Subsequently, the solution can be related to the action \( A - \bar{A} = 0 \), in which no learning takes place.\(^{12}\) As a result, the shareholder and manager follow the myopic contract.

As a result, the HJB equation (11) under equations (12) and (13) may be cut into two parts, when it holds locally. First, the value function under the myopic contract satisfies

\[ m(\pi, W) \equiv \max_{A_t^m - \bar{A}} \{ \mu_R(\pi; W; A - \bar{A}) \}. \]

\(^{11}\) Details of these transformations are shown in Appendix V.

\(^{12}\) Recall that there exists an effort \( \bar{A} = -\frac{\Delta \pi}{\Delta \beta} \) such that the signal is not informative. This result follows from the correlation between the signal and beliefs being zero, i.e., \( \Sigma(\pi_t, A) = 0 \).
which can be solved using a numerical procedure such as the one undertaken by Sannikov (2008).

The first order condition for an optimal myopic effort by the manager is

\[ A_t^m - \bar{A} = \frac{\beta(\pi) + \frac{K_2}{(\pi)^2} \gamma^2 \frac{\partial^2 F}{\partial W^2}}{H(z_1 + z_2 \gamma^2 \frac{\partial^2 F}{\partial W^2})}. \] (15)

Figure 3 draws the myopic effort of the manager for \( A \in [0,1] \). When managerial effort is within this interval in Figure 3, equation (15) is true, when the second order condition

\[ K_2 \left( \frac{\partial F}{\partial W} + \frac{K_2}{(\pi)^2} \gamma^2 \frac{\partial^2 F}{\partial W^2} \right) < 0 \] (16)

is satisfied. In this case, the manager’s effort is decreasing in the manager’s continuation value for any given belief, as in Sannikov (2008, Figure 1). This result is replicated in Figure 3 for the more complicated contracting environment. As in Sannikov (2008), the value of the firm is concave in the manager’s continuation value for given beliefs under normal circumstances. Substituting this effort into equation (13) yields the sensitivity of the manager’s contract under the myopic condition to information about the revenue of the firm. By equation (9), the pay for performance influences the manager’s continuation value. In addition, Figure 3 demonstrates that

---

13 The \( F(\pi, W) \) in this equation is the solution to the complete problem equation (11). This part and the subsequent learning part are separated out to illustrate how the complete solution is determined.

14 Figure 7 illustrates how the second order condition behaves across a simulation of the state of business.

15 See Appendix A4 of Sannikov. Proving concavity in general is problematic because of the behavior of the value function under extreme experimentation.
the myopic effort of the manager increases in the probability of the good state of business. Thus, the manager’s contract is more sensitive to the firm’s revenue when the state of business is expected to improve.

At the corner solutions $A=\{0,1\}$, equation (15) is violated. For example, at $A=0$, 

$$A_t^m - \tilde{A} < 0 - \tilde{A}$$

so that the derivative of equation (14) with respect to $A$ is less than zero, so that it is optimal to choose the corner solution $A=0$. As a result, pay for performance is found by substituting the corner solution into the optimal contract condition (13). At these corners, the manager’s contract is most sensitive to the firm’s revenue.

**III.B. Optimal Learning for a Given Continuation Value.**

The last line of the HJB equation (14) captures the effect of learning on the manager’s optimal contract:

$$\mathcal{L}(A_t - \tilde{A}; \pi; W) = -\Lambda(\pi - \tilde{\pi}) \frac{\partial F}{\partial \pi} + \frac{1}{2} \left\{ \tau(\pi)^2 \frac{\partial^2 F}{\partial \pi^2} + 2\tau(\pi) \frac{H_2}{\beta(\pi)} \sigma^2 \frac{\partial^2 F}{\partial \pi \partial W} \right\} (A - \tilde{A})^2$$

$$+ \tau(\pi) \frac{(H_1 + H_2)(A - \tilde{A})}{\beta(\pi)} \sigma^2 \frac{\partial^2 F}{\partial \pi \partial W}$$

(17)

Here, the standard deviation of beliefs is replaced by its expression in terms of the manager’s effort from equation (6) and the marginal value of effort to learning. This effect of learning is convex in the probability of the state of business $\pi$ (see lemma 3.4, below). The increasing return from learning, represented by the convexity of $\mathcal{L}(A_t - \tilde{A}; \pi; W)$ in $\pi$, through additional managerial effort is the key component for understanding the properties of the optimal contract for managerial compensation.

Effort has no effect on learning when the possible revenue curves cross $A = \tilde{A}$ or the beliefs of the investor and manager are at extreme values $\pi = \{0,1\}$, since there is no standard
deviation in beliefs. Consequently, belief would tend to its stationary value, \( \pi^* \), by equation (6).

Thus, the optimal and myopic effort levels agree, as one sees in Figure 5. From this minimal benefit from learning on the HJB (14), we want to examine how increases or decreases in effort and subsequent increases in the value of learning \( \tau(\pi)\left(A_e - \bar{A}\right) \) affect the HJB equation through equation (17). The following lemma characterizes the value function of the HJB equation.

**Lemma 3.1.** The value function \( F(\pi, W) \) is continuous and convex in the manager’s and investor’s beliefs, \( \pi \).

Proof: See Appendix V for proof following Keller and Rady (1999).16

One can observe this in Figure 4. The firm’s stock price under the optimal contract is convex in the probability of a good state of business around the maximum marginal value of learning, \( \tau(\pi) \), at \( \pi = 1/2 \).

Given this convexity of the firm’s value function in the shareholders’ and manager’s beliefs about the state of business, we can examine how managerial effort will influence the HJB (14) through learning. First, \( \frac{1}{2} \tau(\pi)^2 \frac{\partial^2 F}{\partial \pi^2} \left(A - \bar{A}\right)^2 > 0 \) following the convexity of the value function of the firm. This endogenous term in the HJB equation corresponds to the ad hoc convex impact of learning illustrated in section 2. However, this effect of learning is a consequence of the positive effect of effort on the value of learning on the HJB equation (14) and the convexity of the value function for the firm. The other terms dealing with the effect of learning on the HJB equation (14) could lead to an ambiguous effect of a change in the

---

16 The convexity of the value function with respect to beliefs was first recognized by Easley and Kiefer (1988).
manager’s effort, since the state of business and the shadow price of the manager’s continuation value, \( \frac{\partial^2 F}{\partial \pi \partial W} \), are unknown. To focus on the effect of learning, the analysis here is on the circumstances in which

**Condition A**: \( A < \bar{A} \) and \( \frac{\partial L(A-\bar{A};\pi;W)}{\partial A} < 0 \),

so that the benefit from learning increases as the manager’s effort moves away from \( \bar{A} \). This would be the case as long as the increasing returns for effort of learning in condition A dominate the indirect effects from \( \frac{\partial^2 F}{\partial \pi \partial W} \).

The first part of this condition is chosen since it corresponds to cases in which the shareholder must sacrifice myopic revenue in order to learn about underlying demand for the company’s product. In Figure 5, there is a region, behind the peak, where this condition is violated. In this region, where there is a high value of the good state of business and a high manager’s continuation value, the shareholders do not have to sacrifice revenue to learn. Thus, condition A means that we are not concerned in this work about that case.

The second part of Condition A assures us that the manager and the representative investor have marginal gains from learning when the manager’s effort is reduced. The second condition guarantees that pure learning effects, terms related to \( \frac{\partial^2 F}{\partial \pi^2} \) in the HJB equation (17), are larger than substitution effects between learning and moral hazard (terms related to \( \frac{\partial^2 F}{\partial \pi \partial W} \) in the HJB equation). Figure 5 demonstrates that Condition A is true for \( A < \bar{A} \), since the second order condition associated with the learning effect is positive and there are increasing returns to learning.17 This behavior can also be seen in Figure 4, in that the firm’s stock price is convex in

---

17 The second order condition under the optimal policy is separated into the myopic and learning components to illustrate the tradeoff between the optimal contract and the increasing return from learning.
the probability of the good state of business.

As a result, the HJB equation (14) is given by

$$rF(\pi, W) = \max_{A_t - \tilde{A} \in \mathcal{A}} \{ \mu_r(A_t - \tilde{A}; \pi; W) + \mathcal{L}(A_t - \tilde{A}; \pi; W) \}. \quad (18)$$

Since this form of the HJB equation is quadratic in the manager’s effort, $A - \tilde{A}$, the first order condition may be written as

$$(A - \tilde{A}) = A^m - \tilde{A} - \frac{\tau(\pi)\left(\frac{\tau(\pi)(A - \tilde{A})}{\pi^2} + \left(\frac{2H_2(A - \tilde{A}) + H_1 + H_2\tilde{A}}{\beta(\pi)}\sigma^2 \frac{\partial^2 F}{\partial \pi \partial W}\right)\right)}{\left(\frac{K^2 \partial F}{\beta(\pi)^2 \sigma^2 \partial W^2}\right)}, \quad (19)$$

when the optimal solution is in the interval $(0,1)$. As a result, optimal effort by the manager can be ranked relative to the myopic contract. In Figure 6, the difference $A^m - A$ is graphed. For a given continuation value $W$, there is a probability of the good state which maximizes this difference (see the peak of the curve in Figure 6 for each given $W$). As the continuation value increases, this critical probability of a good state varies. The source of this tradeoff is the dual effect of the probability of the good state of business in equation (19). First, the marginal value of learning, $\tau(\pi)$, is largest at $\pi = 1/2$ and increasing when $\pi < 1/2$. On the other hand, $\beta(\pi)$ increases in the probability of the good state, so that $\left(\frac{2H_2(A - \tilde{A}) + H_1 + H_2\tilde{A}}{\beta(\pi)}\sigma^2 \frac{\partial^2 F}{\partial \pi \partial W}\right)$ decreases. Thus, the biggest impact on the optimal effort of the manager, the peak in Figure 6, depends on the relative magnitudes of the increasing return to learning, $\frac{\partial^2 F}{\partial \pi^2}$, and the impact on the probability of the good state on the marginal value of the manager’s continuation value on the stock price $\frac{\partial F}{\partial W}$.

The myopic contract has an internal solution, but that does not mean that the optimal contract has one, since the increasing returns from learning form lemma 3.1 may dominate the optimization problem in the HJB equation. In this case, equation (19) is a global minimum rather
than a maximum.\textsuperscript{18} As a result, it is optimal for the contract to elicit a corner solution either $A = A_{\text{min}} = 0$ or $A = A_{\text{max}} = 1$ depending on whether $\mu_r (A_{\text{min}} - \tilde{A}; \pi; W) + \mathcal{L}(A_{\text{min}} - \tilde{A}; \pi; W) \geq \mu_r (A_{\text{max}} - \tilde{A}; \pi; W) + \mathcal{L}(A_{\text{max}} - \tilde{A}; \pi; W)$, respectively. In the simple model of Section 2, this was illustrated by Figures 2 (d) and (e).

This behavior is illustrated in Figures 4, 7, and 8. In this case, the concavity of the myopic problem is reduced by making the manager risk-neutral and reducing the impact of the manager’s effort on its marginal cost. The net effect of these changes is to minimize the moral hazard problem and to emphasize the increasing returns from learning. In Figure 4, the firm’s stock price is convex in the probability of the good state of business. Under myopic behavior, the individual chooses the dotted effort. However, the value of learning is highest near $\pi = 0.45$, so that the increasing returns from learning cause the managerial effort to be at either extreme, $A = \{10, 40\}$. At a slightly higher probability of the good state of business, the firm’s value is highest, so that the highest level of effort, $A = 40$, is chosen, while at a slightly lower probability the minimal level of effort is chosen. This behavior is highlighted in the simulation across time in Figure 7. When the good state is high for a while and switches to low, eventually the probability of the good state of business crosses the singularity at $\pi = 0.45$. At this point, the optimal effort switches from its highest value to its lowest value. The opposite occurs when the underlying state switches from the bad to the good state.\textsuperscript{19} In Figure 8, the second order condition is graphed under the same simulation. Each time the manager undertakes extreme behavior, the second order condition becomes positive, so that the increasing returns from

\textsuperscript{18} See Theorem 17.8 of Simon and Blume (1994).

\textsuperscript{19} This corresponds to the illustration in panels d and e of Figure 2 in the first section. This mechanism was used by Balvers and Cosimano (1994) to explain why central banks in hyperinflation countries would tend to stabilize inflation quickly rather than use a more gradual strategy, as in moderate inflation countries. Also keep in mind that this probability of a good state can also cross the critical value when there is a sufficiently large shock to the firm’s revenue.
learning dominate. Under moderate learning, the second order condition never goes positive, so extreme experimentation is not optimal. Consequently, it is possible to have extreme experimentation in the optimal contract for managerial compensation in which the increasing return from learning from the manager’s effort dominates the effort under the myopic contract. This extreme experimentation would only occur when the marginal value of learning, \( \tau(\pi) \), is large and managerial effort is far away from the effort under no learning, \( \bar{A} \). In both cases, the value of learning is large. Thus, the investor and manager prefer to maximize the impact of learning rather than the myopic profits.

IV. Cautious Investors and Sticky Contracts

Closed form expression for the solution of the HJB equation is not available, and, hence, we rely on numerical simulation. While the model of this paper is complex and can generate various phenomena, we focus on two issues of optimal contracts related to shareholder activism: the cautious behavior of investors and the extent to which optimal contracts incorporate new information.

The numerical procedure, described in Appendix VII, is designed to account for the complexity of the optimal learning problem, in which internal and corner solutions can be chosen. The procedure uses value function iteration, in which the three possible choices are examined at each possible node in the state space.

The parameter values are given by
\( \alpha_0 = 1, \beta_0 = 0.75, \Delta \beta = 1.5, \bar{A} = 0.65, \bar{\pi} = 0.45, \Lambda = 0.01, H_1 = 0.5, H_2 = 0.95, r = 0.1, \sigma = 2, W = 4, W \in [0, 10] \) and \( A \in [0, 1] \). \( \text{(20)} \)

These parameters are chosen to capture the properties in both Sannikov (2008) and Keller and Rady (1999). In addition, Condition A is satisfied so that there is a tradeoff between learning and myopic profits in the HJB equation (14) and it is subject to increasing returns from learning.

Figures 9, 10, and 11 portray the optimal effort, the firm’s stock values, and the manager’s consumption for these parameters.

The following proposition states that the optimal compensation contract exhibits “cautious” behavior. That is, the optimal contract with strategic learning is less sensitive to the current period’s output than a myopic one that only accounts for the manager’s moral hazard.

**Proposition 4.1. Cautious Investors:** If the myopic objective \( \mu_r(A - \bar{A}; \pi; W) \) has an interior maximum \( 0 < A^m < \bar{A} \), and Condition A holds, then the optimal solution is smaller than the myopic solution in the neighborhood of the myopic solution. Thus, the manager’s compensation is less sensitive to random fluctuations in information. \( \text{21} \)

Proof: See Appendix VI.

In the compensation problem, the investor and the manager can learn more information by deviating from the myopic contract, and can use this information to implement a more efficient contract in future periods (see Figures 6 and 9). The marginal gains from obtaining new

---

20 The solution takes 210 seconds on a standard PC. Because of the large dimension of the state space the looping over the space was done in C++, so that the solution was found in reasonable time. The calculations of the functions and their derivatives were undertaken in Mathematica, while the evaluations of these functions at the nodes in the state space were done in C++.

21 To keep the argument clear, we use the case \( \chi = \frac{1}{2} \). If we use another value, then the Taylor polynomial of this expression would be needed. The conclusions of this analysis remain the same for a general case when \( \chi \) takes a different value from \( \frac{1}{2} \).
information may outweigh the marginal costs of giving up some current period revenue by deviating from the myopic contract. Since Condition A restricts the compensation problem to the case where less effort induces more learning ($A < \bar{A}$), investors will prefer less effort than a myopic case to increase the long-term gains of learning. The magnitude of this difference is largest when the marginal value of learning is highest at $\pi = 1/2$. This result can be seen in the simulation of the state of business in Figures 12 and 13. The probability of the good state of business crosses this threshold when the underlying state shifts from a good to a bad state or vice versa. As the probability moves across this threshold the difference between the myopic and the optimal managerial contracts is maximized.

Consequently, the investor will voluntarily write a forgiving compensation contract that is less sensitive to the random fluctuation of new information. That is, the investor will exhibit cautious behavior when writing an optimal compensation contract. Such a contract will induce a lower level of effort by the agent, which leads to a reduction in short-term revenue due to the increased costs of moral hazard, but leads to increases in long-term gains from obtaining more information on the current state ($k = \{0,1\}$). Cautious behavior by investors is often observed in practice. For example, a prior study notes that a typical CEO in their sample earns only $3.25 per $1000 increase in shareholders' wealth (Jensen and Murphy, 1990).

For the parameter values in equation (20), a plot of effort under the myopic contract relative to the optimal contract $A^m - A$ in Figure 6 reveals that the effort is lower under the optimal contract for low continuation values and beliefs. Eventually, for higher continuation values, $6 < W < 9$, and sufficiently high beliefs about the possibility of a good state of business, $\pi > 0.5$, an effort which exceeds the myopic solution is optimal. (Look behind the hill in Figures
Yet, in reality, cautious behavior is the most probable solution to the manager’s compensation problem, which is consistent with the typical compensation of a CEO.

Consequently, given that investors face a lot of uncertainty about the current state of business, this study provides a rationale for this low pay-performance sensitivity based on strategic learning motives by investors. The managerial compensation problem also sheds light on changes in optimal compensation contracts when the level of uncertainty of the underlying economic environment changes. For example, Proposition 4.1 suggests that optimal compensation contracts tend to have lower pay-for-performance sensitivity in times of great economic uncertainty, such as financial crises, because gains from learning by deviating from myopic moral-hazard-based contract can be very large.

This result can be seen in Figures 5 and 6. In Figure 6, the most cautious behavior occurs along a ridge that trades off the state of beliefs and the manager’s continuation value. In the same figure, the optimal effort increases with beliefs about a good state of business until the probability of a good state is about one half. At this point, the uncertainty about the state of business is largest. Thus, during bad times, in which belief decreases and continuation values are smaller, the optimal effort is smaller, so the contract is less sensitive to current information from revenue.

The second feature of the model that we focus on is how much new information is incorporated into optimal contracts. The model predicts that the extent to which optimal contracts incorporate new information depends on the signal-to-noise ratio, $\tau(\pi)$ and $(A - \bar{A})$. This signal-to-noise ratio is highest for given $\Delta\beta$ and $\sigma$, when the prior beliefs are diffuse, $\pi = 1/2$, and the optimal contract yields the minimum level of effort. In Figure 12, a simulation of the managerial compensation problem for 300 periods shows that the optimal managerial
contract does visit these situations frequently. These situations arise as the state of business
switches from good (bad) to bad (good) states of business, so that the expected probability of
good business migrates from high (low) to low (high) values. The number of times this occurs is
associated with the persistence of the good state, $\Lambda = 0.1$, which gives a half life of $\ln(2)/0.1 = 7$
periods for the shock. The results are summarized in the following proposition:

**Proposition 4.2. Sticky Contracts:** If the myopic objective $\mu_R(A - \tilde{A}; \pi; W)$ has an interior
maximum $A^m < \tilde{A}$, and Condition A holds, then the optimal solution is an interior solution on
the compact interval $[0, A_{max}]$ and the adjustment of the contract to new information is gradual,
when the concavity of the myopic solution *dominates* the increasing returns from learning.
When the concavity of the myopic solution *is dominated by* the increasing returns from learning,
then the optima occurs at either $A_t = 0$ or $A_t = A_{max}$ based on the comparison $\mu_R(0 -
\tilde{A}; \pi; W) + \mathcal{L}(0 - \tilde{A}; \pi; W)$ and $\mu_R(A_{max} - \tilde{A}; \pi; W) + \mathcal{L}(A_{max} - \tilde{A}; \pi; W)$.

Proof: See Appendix VI.

Proposition 4.2 states that there are two different types of optimal contracts based on how new
information is incorporated into these contracts. When the signal-to-noise ratio, $\tau(\pi)$, is small,
the moral hazard problem, i.e., $\mu_R(A - \tilde{A}; \pi; W)$, is more important than the learning effect, i.e.,
$\mathcal{L}(A - \tilde{A}; \pi; W)$. Since the moral hazard problem is concave in effort whereas the learning
problem is convex in effort, the total revenue from the myopic and the learning problem,
$\mu_R(A - \tilde{A}; \pi; W) + \mathcal{L}(A - \tilde{A}; \pi; W)$, is concave in effort, $A$. Figure 4 shows the second order
conditions from the moral hazard and the learning problem by subtracting the second order
component from myopic behavior $\mu_R(A - \tilde{A}; \pi; W)$ from the total second order condition. As
shown in Figure 4, the net effect is positive but the resulting total revenue is concave in effort. In analogy to the two period model in section I, Figures 5, 6, and 9 correspond to the case of Point A in Figure 2, where the effect of moral hazard dominated the learning effect and the objective function (revenue) is concave in effort. In this case, the optimal contract leads to an internal solution, i.e., the recommended effort induced by the optimal contract lies between 0 and \( A_{max} = 1 \). In this case, small changes due to the arrival of new information lead to corresponding small changes in the optimal contract (pay-for-performance sensitivity, \( \gamma \), and induced effort, \( A \)). Figures 12 and 13 show the corresponding time-series plot for a simulated random revenue path. As shown in the Figure, optimal contracts adjust slowly with respect to changes in switching states. We denote this mode as the moderate experimentation regime.

When the signal-to-noise ratio, \( \tau(\pi) \) and \((A - \bar{A})\), is large, the learning problem becomes more important than the moral hazard problem. Since the learning problem has increasing returns to scale (i.e., convex in effort), the total revenue from learning and moral hazard becomes convex in effort. To obtain this result, it is assumed that the manager is risk-neutral and the marginal cost of the manager’s effort is less sensitive to her effort. As a result, the optimal effort induced by the optimal contract leads to a corner solution in either 0 or \( A_{max} \) when the signal-to-noise ratio is highest, around \( \pi = 0.45 \). In this case, a small change in belief about the current state of the business, \( \pi \), can lead to a large jump between two extreme values, \( A_{min} = 10 \) or \( A_{max} = 60 \). Figure 4 shows the second order condition of the moral hazard and the learning effect. As shown in this Figure, the learning effect, which is convex, dominates the moral hazard effect when the probability of the good state is near \( \pi = 0.45 \). In analogy to a two-period model, this situation corresponds to Point C or D in Figure 2. Figures 7 and 8 show a time-series plot for a simulated random revenue path for this case. As shown in the Figure, recommended effort
induced by the optimal contract jumps between $A_{min} = 10$ or $A_{max} = 60$ as belief about the state of business changes due to switching states over time. This switch occurs when the underlying state of business changes from good (bad) to bad (good), so that the estimated probability of the good state crosses the critical value $\pi = 0.45$. In this case, the effort goes from 60 to 10 (10 to 60) for a small change in the probability of the good state. We denote this mode as the extreme experimentation regime.

Results from Proposition 4.2 show that the extent to which optimal contracts incorporate new information can be endogenously determined by the signal-to-noise ratio, i.e., the larger the signal-to-noise ratio, the larger the benefit of learning relative to the moral hazard problem. Hence, information will be more actively incorporated in optimal contracts when the signal-to-noise ratio is large. The proposition also suggests that the way new information is incorporated in optimal contracts can occur in two qualitatively different ways: in a moderate experimentation regime, where the signal-to-noise ratio is very small, new information is gradually incorporated in optimal contracts. In contrast, when the signal-to-noise ratio is very large, optimal contracts lead to bang-bang solutions and small changes in new information can lead to sudden jumps between two extreme corner values.

Predictions from Proposition 4.2 suggest that extremely volatile adjustment of optimal contracts can occur when the marginal value of learning is very large and the moral hazard effect is smaller due to a large amount of uncertainty in underlying economic conditions. For example, in times of financial crisis or in transition economies where uncertainty about the underlying economic condition is large, the value of gaining additional information can be very large. In these circumstances, optimal compensation contracts can experience very large changes. In contrast, in developed countries where economic uncertainties are moderate, optimal contracts
are expected to be more persistent.

V. Conclusions

In this paper, we developed a dynamic incentive contracts model with parameter uncertainties. In this model, investors' choice of optimal contracts trade off short-run profits from myopic actions with long-run benefits from gaining information on the current state of business. We endogenously characterize the stickiness of contracts and identify two distinct effects based on the long-term benefits of learning relative to short run profits. We also show that the optimal contract offered by investors can substantially deviate from the myopic solution (e.g., less sensitive to the current outcome) when the strategic collection of information is possible and valuable. Results of this study can provide a rational explanation for the commonly observed low pay-for-performance sensitivity and sluggish adjustment of executive compensation.

Our problem focuses on the impact of strategic learning on moral hazard problems. An important extension would be to consider situations where both moral hazard and adverse selection (e.g., unobservable managerial action) problems are present. We leave them for future studies.
References


Green, E., 1987, Lending and the Smoothing of Uninsurable Income, in Edward Prescott and Neil Wallace eds.: *Contractual Arrangements for Intertemporal Trade*, University of Minnesota Press, Minneapolis, MN.


Appendix I. Proof of the Two-Period Model in Section I

Since the outcome is realized and the game ends at $t = 1$, we use backward induction to obtain optimal contracts.

(1) Step 1: The agent's utility at $t = 1$

Let’s denote the initial beliefs about high and low states at $t = 0$ as $\Pr(\tilde{\theta} = 0) = 1 - \omega$ and $\Pr(\tilde{\theta} = \theta) = \omega$. To simplify the associated algebra for this example, we assume $\omega = \frac{1}{2}$.

Both the principal and the agent observe the signal after the initial effort decision ($e_0$) is made. Hence, the optimal effort and optimal contract at $t = 0$ depends on the signal received.

i) If $\tilde{s} = s_G$, the principal offers $b_G$, and the agent exerts effort $e_G$

$$u_{1G}(b_G, e_G) = \omega(1 - e_0) \left\{ e_G b_G \theta - \frac{1}{2} e_G^2 \right\} + (1 - \omega) e_0 \left\{ e_G b_G 0 - \frac{1}{2} e_G^2 \right\}. $$

From the first order condition for $e_G$, we find that the optimal effort when $\tilde{s} = s_G$ is

$$e_G = \frac{\omega(1 - e_0) b_G \theta}{\omega(1 - e_0) + (1 - \omega) e_0}. $$

ii) If $\tilde{s} = s_B$, the principal offers $b_B$, and the agent exerts effort $e_B$,

$$u_{1B}(b_B, e_B) = \omega e_0 \left\{ e_B b_B \theta - \frac{1}{2} e_B^2 \right\} + (1 - \omega)(1 - e_0) \left\{ e_B b_B 0 - \frac{1}{2} e_B^2 \right\}. $$

From the first order condition for $e_B$, we find that the optimal effort when $\tilde{s} = s_B$ is

$$e_B = \frac{\omega e_0 b_B \theta}{\omega e_0 + (1 - \omega)(1 - e_0)}. $$

(2) Step 2: The principal's profit at $t = 1$

If $\tilde{s} = s_G$, the principal offers $b_G$,

$$\Pi_{1G}(b_G) = \frac{\omega^2 \theta^2 (1 - e_0)^2}{\omega(1 - e_0) + (1 - \omega) e_0} b_G (1 - b_G). $$
From the first order condition for optimizing $b_G$, we find that the optimal contract ($b_G$) when $\bar{s} = s_G$ is

$$b_G = \frac{1}{2}.$$ 

Similarly, we find that the optimal contract ($b_B$) when $\bar{s} = s_B$ is

$$b_B = \frac{1}{2}.$$ 

(3) Step 3: The agent's utility at $t = 0$

Since $b_G = b_B = \frac{1}{2}$ and $\omega = \frac{1}{2}$,

$$u_0(b_0, e_0) = \frac{1}{2} e_0 b_0 \theta - \frac{1}{2} e_0^2 + \frac{\delta \theta^2}{16} ((1 - e_0)^2 + e_0^2).$$

From the first order condition for $e_0$, we find that the initial effort is

$$e_0 = \frac{\theta (4b_0 - \delta \theta)}{2(4 - \delta \theta^2)}.$$ 

(4) Step 4: The principal’s profit at $t = 0$

$$\Pi_0(b_0) = \frac{\theta^2 (4b_0 - \delta \theta)}{4(4 - \delta \theta^2)} (1 - b_0) + \frac{\delta \theta^2}{8} ((1 - e_0)^2 + e_0^2).$$

Hence, the second derivative of the expected principal’s profit is

$$\frac{d^2 \Pi_0}{db_0^2} = \frac{4\theta^2(2 + \delta \theta^2)}{(-4 + \delta \theta^2)^2},$$

which is positive (convex) when $\delta \theta^2 > 2$.

**Appendix II.** The HJB Equation and Verification (Proof of Proposition 2.1)

The evolution of the manager’s continuation value and the incentive compatibility condition are found using the same argument as Sannikov (2008) in his sections A1, A2, and A3. One just has to keep track of the expected state of business. These calculations may be obtained
The two state variables \((\pi, W)\) are driven by a single Brownian motion and the controls modulate the local variance and correlation of the state variables. This causes the HJB equation to be degenerate elliptic and presents some difficulties, both theoretical and numerical. In particular, the investor's value function does not possess the necessary smoothness to be a classical solution to the HJB equation. Also, the optimal controls cannot be expressed in feedback form at points of insufficient smoothness. The theory of viscosity solutions is an indispensable tool for this case. See Fleming and Soner (2006, Ch. VII) or Pham (2009, Ch.4) for a systematic exposition.

We begin by relaxing the investor's optimization problem by adding a small amount \(\varepsilon > 0\) of noise to the dynamics of the promised utility in the manager's contract in a way that does not destroy the incentive compatibility of the contract. After showing that the problem is well-posed for each \(\varepsilon > 0\), the method of vanishing viscosity (Fleming and Soner (2006, VII.9)) immediately gives uniform convergence of the value functions (and policy functions) to the unique viscosity solution of the HJB equation with \(\varepsilon = 0\).

The investor's problem then becomes to maximize the firm value

\[
F^\varepsilon(\pi, W; c, A) = \mathbb{E}^I \left[ \int_0^\infty e^{-r\tau} (dR_t - c_t dt) \mid \pi_0 = \pi, W_0 = W \right]
\]

over admissible effort and control pairs \((c, A)\). In light of our standing assumption that the action is essentially observable to the investor for purposes of updating his beliefs, though not for contracting, we adopt a “weak” formulation of the SDEs governing the state variables \((\pi, w)\). That is, we consider only the joint law of motion, not the path-by-path relations to unobserved quantities in the filtering problem. Together with the \(\varepsilon\) relaxation, this leads to dynamics in terms of independent Brownian motions \((Z^I_t, Z^\varepsilon_t)\) given by
\[
\begin{align*}
\text{d}R &= (\alpha_0 + \beta_0 A_t + (A_t - \bar{A})\Delta \beta k) \text{d}t + \sigma \text{d}Z_t^1 \\
\text{d}\pi_t &= \lambda(\pi_t) \text{d}t + \Sigma(\pi_t, A_t) \text{d}Z_t^2 \\
\text{d}W_t &= (r W_t - u(C_t) + h(A_t)) \text{d}t + \sigma \gamma(\pi_t, A_t) \text{d}Z_t^1 + \varepsilon \text{d}Z_t^3
\end{align*}
\]

The \( \varepsilon \) relaxed value function is

\[F^\varepsilon(\pi, W) = \sup_{(c,A)} F^\varepsilon(\pi, W; c, A),\]

where the supremum is taken over the progressively measurable processes \((c, A) \in [0, \infty) \times \mathcal{A}\) for which the state variables \((\pi, W)\) are well-defined and square integrable for all \(t \geq 0\).

This control problem is classical and one can expect a smooth value function for every \(\varepsilon > 0\), since the covariance matrix for \((d\pi, dW)\) positive definite. In particular, there is a \(K > 0\) such that for all \(\pi \in [0,1]\), \(A \in \mathcal{A}\) and for all real \(x, y\)

\[\Sigma^2(\pi, A)x^2 + 2 \sigma \gamma(\pi, A) \Sigma(\pi, A) xy + (\sigma^2 \gamma^2(\pi, A) + \varepsilon^2)y^2 \geq K(x^2 + y^2)\]

The HJB equation associated with this problem is the same as in Section II.C. equation (10) with the addition of the term \(\varepsilon^2 \frac{\partial^2 F^\varepsilon}{\partial W^2} \).

It is evident that the \(\varepsilon\)-relaxed problem satisfies the conditions of Theorem IV 5.1 in Fleming and Soner (2006). From this classical verification theorem, we infer that the \(\varepsilon\)-relaxed value function \(F^\varepsilon(\pi, W)\) is identified with the unique viscosity solution of the relaxed HJB equation and the policy functions characterize optimal controls in feedback form \((c^\varepsilon(\pi, W), A^\varepsilon(\pi, W))\).

Having established the validity of the relaxed model, it remains to show that the uniform limit of \(F^\varepsilon(\pi, W)\) is the unique viscosity solution for the un-relaxed HJB equation and similarly for the optimal controls. This is assured because the HJB equation satisfies a comparison principle required for using the Barles and Perthame procedure outlined in Section VII.3 of Fleming and Soner (2009) and proved in their subsequent sections.
Appendix III. Additional Contractual Possibilities

The state space for the system is $\pi \in [0,1]$ and $W \in [0, W_{\text{max}}]$, where $W_{\text{max}}$ is the bliss point utility, consumption is maximal and effort is zero. We consider two additional contractual possibilities. First, from state $(\pi, W)$ the shareholder may instantaneously renegotiate the contract to any state $(\pi', W')$, where $W' \geq W$ makes the manager better off. Typically, this option is not needed by the principal, but this behavior specifies boundary conditions for the HJB equation at $W_{\text{min}} = 0$, as seen below. Second, the shareholder may retire the manager and hire a new one. To retire the manager, the shareholder must purchase an annuity for the manager that delivers the promised utility, $W$. The cost to retire the agent is

$$C^{\text{ret}}(W) = \frac{u^{-1}(rw)}{r}.$$ 

Immediately after retiring the manager, a new one is hired at the market price given by the reservation utility for new managers, $\hat{W} > W_{\text{min}}$. This causes the state to move from $(\pi, W)$ to $(\pi, \hat{W})$.

Letting $dt > 0$ be a vanishingly small time interval, the Bellman equation for the firm’s value $F(\pi, W)$ is given by the maximum over three possibilities of continuing or renegotiating the contract or retiring the manager.

$$\max_{A_c} E[dR - cd + e^{-\gamma dt}F(\pi + d\pi, W + dW)],$$

$$\max_{W' \geq W} F(\pi, W'),$$

and

$$F(\pi, W) - C^{\text{ret}}(W).$$

Applying Ito’s calculus, the Bellman equation can be reduced to a system of HJB variational inequalities, where, at each point in the state space $\pi \in [0,1]$ and $W \in (0, W_{\text{max}})$, the following are expected to hold, and at least one of the inequalities is an equality.
Here, the infinitesimal generator of \((\pi, W)\) is given by Ito’s lemma applied to \(F(\pi, W)\) using the stochastic processes (6) and (9).

If the promised utility ever reaches the level \(W_{\text{max}}\), the manager must retire, so the boundary condition holds:
\[
F(\pi, W_{\text{max}}) = F(\pi, W_{\text{max}}) - C^{ret}(W_{\text{max}}).
\]

At \(W_{\text{min}}\), incentives can no longer be provided to the agent by the usual mechanism in (9). Therefore, renegotiation is required and the following condition holds:
\[
F(\pi, W_{\text{min}}) = \max_{W' \geq W_{\text{min}}} F(\pi, W').
\]

A further simplification is possible if the stock price is strictly decreasing and concave in \(W\) when \(W \geq W_{\text{min}}\). In this case, the renegotiation problem is trivial: \(\max_{W' \geq W} F(\pi, W') = F(\pi, W)\), and renegotiation can only be optimal at \(W_{\text{min}}\). Then, we have a simpler system of HJB variational inequalities
\[
0 \geq \max_{A, c} E[\alpha(\pi) + \beta(\pi)A - c - rF(\pi, W) + dF(\pi, W)], \quad \text{and}
\]
\[
0 \geq F(\pi, \bar{W}) - C^{ret}(\bar{W}) - F(\pi, W).
\]
\[
F(\pi, W_{\text{min}}) \quad \text{and} \quad F(\pi, W_{\text{max}}) = F(\pi, W_{\text{max}}) - C^{ret}(W_{\text{max}}).
\]

Since we do not expect \(F\) to be sufficiently smooth to possess all derivatives in this variational inequality, it is to be understood in the sense of viscosity solutions. See Appendix II for a discussion of this issue. Thus, we can proceed with the analysis of the property of this solution.

Appendix IV. Transformation of Effort and Pay-for-Performance Sensitivity
\[ h(A - \bar{A}) = H_1 A + \frac{1}{2} H_2 A^2 = H_1 \bar{A} - \frac{1}{2} H_2 \bar{A}^2 + H_2 \bar{A}^2 + (H_1 + H_2 A)(A - \bar{A}) + \frac{1}{2} H_2 (A - \bar{A})^2 \]

\[ = K_0 + K_1 (A - \bar{A}) + \frac{1}{2} K_2 (A - \bar{A})^2 \]

and

\[ \gamma(A - \bar{A})^2 = \left( \frac{H_2 A + H_1}{\beta(\pi)} \right)^2 \]

\[ = \left( \frac{H_2^2 (A - \bar{A})^2 + 2H_2 (H_1 + H_2 A)(A - \bar{A}) + H_2 (H_1 + H_2 A)\bar{A} - H_2^2 \bar{A}^2 + H_1}{\beta(\pi)^2} \right) \]

\[ = \left( \frac{K_2^2 (A - \bar{A})^2 + 2K_2 (K_1 + K_2 \bar{A})(A - \bar{A}) + K}{\beta(\pi)^2} \right). \]

**Appendix V.** Convexity and Continuity of the Value Function in \( \pi \) for a Given \( W \)

**Lemma 3.4.** The value function \( \eta F(\pi, W) \) is continuous and convex in \( \pi \).

Proof: This argument follows Proposition B.1. of Keller and Rady’s (1999) working paper.

Recall (4) and (5) for fixed \( A - \bar{A} \in \mathcal{A} \) and \( W \). It is linear in \( \pi \).

\[ u(\pi, W) = \pi E_{k=1} \left[ \int_0^\infty e^{-rt}((\alpha_k + \beta_k A_t) - c_t)dt \right] \]

\[ + (1 - \pi) E_{k=0} \left[ \int_0^\infty e^{-rt}((\alpha_k + \beta_k A_t) - c_t)dt \right] \]

Subject to providing an initial value of at least \( \widehat{W} \),

\[ \pi E_{k=1} \left[ \int_0^\infty e^{-rt}(u(c_t) - h(A_t))dt \right] + (1 - \pi) E_{k=0} \left[ \int_0^\infty e^{-rt}(u(c_t) - h(A_t))dt \right] \geq \widehat{W}. \]

Notice that we are using the law of iterative expectations to reverse the way we treat \( \pi \) relative to the rest of the analysis. This works as long as we can use Fubini’s Theorem. For this work, we presume that all the expectations or integrals are well defined, which is the subject of analysis of stochastic integrals. Essentially, you cannot take on any policy which causes the integrals to blow up.
For \( \pi = \eta \pi_1 + (1 - \eta)\pi_2 \) with \( 0 \leq \eta \leq 1 \), it follows that

\[
u(\pi, W) = \eta \nu(\pi_1, W) + (1 - \eta)\nu(\pi_2, W) \leq \eta F(\pi_1, W) + (1 - \eta)F(\pi_2, W),
\]

where \( F \) refers to \( u(\pi, W) \) after optimization over \( A - \bar{A} \in \mathcal{A} \). The inequality comes from the definition of an optimal policy.

Apply supremum to yield

\[
F(\eta \pi_1 + (1 - \eta)\pi_2, W) \leq \eta F(\pi_1, W) + (1 - \eta)F(\pi_2, W),
\]

because the right hand side is already the supremum by the definition of the value function. Thus, the value function is convex in \( \pi \).

The logic for continuity in \( \pi \) is identical to Keller and Rady’s (1999) Proposition B.1. and is omitted here.

**Appendix VI.** Proofs on Strategic Experimentation

**Proposition 4.1. Cautious Investors:** If the myopic objective \( \mu_r(A; \pi; W) \) has an interior maximum \( 0 < A^m < \bar{A} \), and Condition A holds, then the optimal solution is smaller than the myopic solution in the neighborhood of the myopic solution. Thus the manager’s compensation is less sensitive to random fluctuations in information.

Proof: To prove the proposition in the neighborhood of the myopic solution, the perturbation method (see Bender and Orzag, Chapter 7, 1999) is used. Introduce a perturbation parameter \( \epsilon > 0 \), so that the signal-to-noise ratio is zero when \( \epsilon = 0 \). As a result, the HJB equation is

\[
rF(\pi, W) = -\Lambda(\pi - \bar{n}) \frac{\partial F}{\partial \pi} + \max_{A_i \in \mathcal{A}, \bar{A}} \left\{ \left( \epsilon \tau(\pi) (A - \bar{A}) \right)^2 \frac{\partial^2 F}{\partial \pi^2} + \epsilon \tau(\pi) \left( \frac{H_2(A - \bar{A})^2 + (H_1 + H_2 \bar{A})(A - \bar{A})}{\beta(\pi)} \right) \sigma^2 \frac{\partial^2 F}{\partial \pi \partial W} \right\}
\]
Suppose the solution to the HJB equation has the form
\[ F(\pi, W) = \sum_{j=0}^{\infty} \varepsilon^j F_j(\pi, W). \] (A1)

Consequently, the HJB becomes
\[
\begin{align*}
&= r \sum_{j=0}^{\infty} \varepsilon^j F_j(\pi, W) = -\Lambda(\pi - \bar{\pi}) \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial F_j}{\partial \pi} + \max_{A_1 - \bar{A} \in \mathcal{A}} \left\{ \left( \varepsilon \tau(\pi)(A - \bar{A}) \right)^2 \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial \pi^2} 
\right.
\end{align*}
\]
\[
+ \varepsilon \tau(\pi) \left( \frac{H_2 (A - \bar{A})^2 + (H_1 + H_2 \bar{A})(A - \bar{A})}{\beta(\pi)} \right) \sigma^2 \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial \pi \partial W}
\]
\[
+ \left( \alpha_0 - \beta_0 \bar{A} + \beta_0 (A - \bar{A}) + (A - \bar{A}) \Delta \pi - \sum_{j=0}^{\infty} \varepsilon^j \sum_{k=0}^{j} \frac{\partial F_k}{\partial W} \frac{\partial F_{j-k}}{\partial W} \right)
\]
\[
+ \left( rW + \frac{1}{1 - \chi} + K_0 + K_1 (A - \bar{A}) + \frac{1}{2} K_2 (A - \bar{A})^2 \right) \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial F_j}{\partial W}
\]
\[
+ \frac{1}{2} \left( \frac{K_2^2 (A - \bar{A})^2 + 2K_2 (K_1 + K_2 \bar{A})(A - \bar{A}) + K}{\beta(\pi)^2} \right) \sigma^2 \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial W^2}. \] (A2)

The necessary condition for a local maximum is
\[
(A - \bar{A}) = A^m - \bar{A} - \frac{\tau(\pi) \left( \tau(\pi)(A - \bar{A}) \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial \pi^2} + \left( \frac{2H_2 (A - \bar{A}) + H_1 + H_2 \bar{A}}{\beta(\pi)} \right) \sigma^2 \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial \pi \partial W} \right)}{\left( K_2 \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial F_j}{\partial W} + \frac{K_2}{\beta(\pi)} \sigma^2 \sum_{j=0}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial W^2} \right)}. \] (A3)

where
\[ A^m - \bar{A} = -\left\{ \frac{\beta(\pi) + K_1 \sum_{j=0}^{\infty} e^j \frac{\partial F_j}{\partial \pi} + (K_2(K_1 + K_2 \bar{A})) \sigma^2 \sum_{j=0}^{\infty} e^j \frac{\partial^2 F_j}{\partial \pi^2}}{K_2 \sum_{j=0}^{\infty} e^j \frac{\partial F_j}{\partial \pi} + \frac{K_2^2 K_1 + K_2^2 (K_1 + K_2 \bar{A})}{\beta(\pi)^2} \sigma^2 \sum_{j=0}^{\infty} e^j \frac{\partial^2 F_j}{\partial \pi^2}} \right\}. \]  

(A4)

(A4) is true since \( \mu_r(\pi; W) \) has an interior maximum, \( 0 < A^m < \bar{A} \). In addition, the denominator in (A3) is negative by the second order condition of the myopic solution. The numerator is also negative by Condition A. Thus, \( A - \bar{A} < A^m - \bar{A} \) when there is an interior solution to the general problem.

The sufficient condition for such a local maximum is

\[ K_2 \sum_{j=0}^{\infty} e^j \frac{\partial F_j}{\partial \pi} + \frac{H^2}{\beta(\pi)^2} \sigma^2 \sum_{j=0}^{\infty} e^j \frac{\partial^2 F_j}{\partial \pi^2} + \tau(\pi) \left\{ \left( \tau(\pi) (A - \bar{A}) \right)^2 - \frac{\partial^2 F_{j-1}}{\partial \pi^2} \right\} < 0. \]

The next steps show that as \( \epsilon \to 0 \), the optimal solution tends toward the myopic solution. Suppose for each \( \epsilon^j \) for \( j = 0, \cdots, \infty \) the associated HJB is

\[ r F_j(\pi, W) = -\Lambda(\pi - \bar{\pi}) \frac{\partial F_j}{\partial \pi} - \sum_{k=1}^{\infty} \frac{\partial F_k}{\partial W} \frac{\partial F_{j-k}}{\partial W} + \max_{A_1, A_2, A_3} \frac{1}{2} \left\{ \left( \tau(\pi) (A - \bar{A}) \right)^2 - \frac{\partial^2 F_{j-1}}{\partial \pi^2} \right\} \]

\[ + \tau(\pi) \left( \frac{H_2 (A - \bar{A})^2 + (H_1 + H_2 \bar{A}) (A - \bar{A})}{\beta(\pi)} \right) \sigma^2 \frac{\partial^2 F_{j-1}}{\partial \pi \partial W} \]

\[ + (\alpha_0 - \beta_0 \bar{A} + \beta_0 (A - \bar{A}) + (A - \bar{A}) \Delta \beta \pi) \]

\[ + \left( r W + \frac{1}{1 - \chi} + K_0 + K_1 (A - \bar{A}) + \frac{1}{2} K_2 (A - \bar{A})^2 - 2 \frac{\partial F_0}{\partial W} \frac{\partial F_j}{\partial W} \right) \]

\[ + \frac{1}{2} \left( \frac{K_1^2 (A - \bar{A})^2 + 2 K_2 (K_1 + K_2 \bar{A}) (A - \bar{A}) + K_2 \beta(\pi)^2}{\beta(\pi)^2} \right) \sigma^2 \frac{\partial^2 F_j}{\partial \pi^2} \]  

(A5)

for \( j = 0, \cdots, \infty \), where each partial derivative with a negative subscript is set to zero. Let \( F_j(\pi, W) \) solve the HJB (A5). Thus, we take \( F_j(\pi, W) \) for \( j = 0, \cdots, \infty \) to be the functions in the proposed solution (A1). Consequently, (A1) is the solution to (A2).

The initial problem for \( j = 0 \) has an additional term, \( -\Lambda(\pi - \bar{\pi}) \frac{\partial F_j}{\partial \pi} \), relative to the myopic problem. However, in this case, the signal-to-noise ratio is zero, so that both agents
know $\pi \to \tilde{\pi}$ as $t \to \infty$. Consequently, a rational agent would anticipate this and set this term to zero. As a result, the initial problem is the same as the myopic (pure moral hazard) problem.

For subsequent problems, $\sum_{k=1}^{l} \frac{\partial F_k}{\partial w} \frac{\partial F_{j-k}}{\partial w^k}, \frac{\partial F_{j-1}}{\partial \pi^2},$ and $\frac{\partial F_{j-2}}{\partial \pi^2}$ were calculated in the previous step. Thus, each subsequent problem is a combination of a myopic problem and a quadratic learning function in $(A - \tilde{A})$. Therefore, each subproblem is comparable to Keller and Rady (1999), so their logic can be used to approximate the problem.

The first order condition for a local solution to the subsequent problems (A5) is

$$\left( A_j - \tilde{A} \right) = A_j^m - \tilde{A} - \frac{\tau(\pi)\left( \alpha_2^2 F_{j-2} + \frac{2H_2(A_j - \tilde{A}) + H_1 + H_2}{\beta(\pi)} \alpha_2^2 F_{j-1} \right)}{\left( K_2 \frac{\partial \beta}{\partial \pi^2} + K_2 \frac{\partial \alpha}{\partial \pi^2} \right) \alpha_2^2 F_j}$$

(A6)

where

$$A_j^m - \tilde{A} = -\left\{ \frac{\beta(\pi) + K_1 \frac{\partial \alpha}{\partial w} + \left( \frac{K_2(K_2 + K_4)}{\beta(\pi)^2} \right) \alpha_2^2 F_j}{K_2 \frac{\partial \beta}{\partial \pi^2} + \frac{\partial ^2 \alpha}{\partial \pi^2} \alpha_2^2 F_j} \right\}.$$

If the myopic solution $R(A; \pi; W)$ is true for the subproblem $F_j(\pi, W)$, then the denominator is negative in (A6). In addition, Condition A applied to the subproblem $F_j(\pi, W)$ implies that the numerator is negative. Thus, the condition of the proposition holds for the subproblem.

The second order condition for a local maximum is

$$K_2 \frac{\partial F_j}{\partial w} + H_2 \frac{\partial^2 F_j}{\partial w^2} + \tau(\pi) \left\{ \alpha_2^2 F_j \frac{\partial \beta}{\partial \pi^2} + \frac{H_2}{\beta(\pi)^2} \right\} < 0.$$

Finally, the optimal action of the agent (A3) is a weighted average of the subsequent problems:

$$A - \tilde{A} = \sum_{j=0}^{\infty} e^j \omega_j \left( A_j - \tilde{A} \right)$$

where
\[ 0 < \omega_j = \frac{K_2 \frac{\partial F_j}{\partial w} + \frac{2}{\rho(\pi^2)} \sigma^2 \frac{\partial^2 F_j}{\partial w^2}}{K_2 \sum_{j=1}^{\infty} \varepsilon^j \frac{\partial F_j}{\partial w} + \frac{2}{\rho(\pi^2)} \sigma^2 \sum_{j=1}^{\infty} \varepsilon^j \frac{\partial^2 F_j}{\partial w^2}} < 1 \]

as \( \varepsilon \to 0, A - \tilde{A} \to A_0^m - \tilde{A}, \) since \( A_0 - \tilde{A} = A_0^m - \tilde{A} \).

**Proposition 4.2. Sticky Contracts:** If the myopic objective \( \mu_R(A; \pi; W) \) has an interior maximum \( A_m < \tilde{A} \), and Condition A holds, then the optimal solution is an interior solution on the compact interval \([0, A_{max}]\) and the adjustment of the contract to new information is gradual, when the concavity of the myopic solution dominates the increasing returns from learning.

When the concavity of the myopic solution is dominated by the increasing returns from learning, then the optima occurs at either \( A_t = 0 \) or \( A_t = A_{max} \) based on the comparison \( \mu_R(0 - \tilde{A}; \pi; W) + L(0 - \tilde{A}; \pi; W) \) and \( \mu_R(A_{max} - \tilde{A}; \pi; W) + L(A_{max} - \tilde{A}; \pi; W) \).

Proof: The proof is an application of the sufficient conditions (see, Chow, 1997, pp.25–27) for a maximum under Kuhn-Tucker conditions (see Chow, 1997, pp.554-550) as applied to stochastic control problems (see Chow, 1997, pp.142-145).

For each subproblem in Proposition 4.1, the second order condition is true when

\[ K_2 \frac{\partial F_j}{\partial w} + \frac{2}{\rho(\pi^2)} \sigma^2 \frac{\partial^2 F_j}{\partial w^2} < -\tau(\pi) \left\{ \psi(\pi) \frac{\partial^2 F_j}{\partial w^2} + \frac{2}{\rho(\pi^2)} \sigma^2 \frac{\partial^2 F_j}{\partial w^2} \right\} \]  

(A7)

If this is true for each subproblem (A5), then it is true for the original HJB problem. Now, if (A6) does not hold in the interior of the interval \([0, \tilde{A}]\), then a corner solution occurs at \( A - \tilde{A} = -\tilde{A} \), when

\[ \frac{\partial (\mu_R(0 - \tilde{A}; \pi; W) + L(0 - \tilde{A}; \pi; W))}{\partial (A - \tilde{A})} < 0, \]  

as long as \( \mu_R(0 - \tilde{A}; \pi; W) + L(0 - \tilde{A}; \pi; W) \geq \mu_R(A_{max} - \tilde{A}; \pi; W) + L(A_{max} - \tilde{A}; \pi; W) \), by the Kuhn-Tucker conditions. This means that the second order condition (A7) is violated at \( A - \tilde{A} = 0 \), so the objective for the subproblem (A5) is convex here.
At the other corner, if (A5) does not hold in the interior of the interval \([\bar{A}, A_{max}]\), then a corner solution occurs at \(A - \bar{A} = A_{max} - \bar{A}\), when \(\frac{\partial \mu(R(A_{max} - \bar{A}; \pi; W)) + \mu(L(A_{max} - \bar{A}; \pi; W))}{\partial (A - \bar{A})} > 0\), as long as \(\mu_R(0 - \bar{A}; \pi; W) + L(0 - \bar{A}; \pi; W) \leq \mu_R(A_{max} - \bar{A}; \pi; W) + L(A_{max} - \bar{A}; \pi; W)\), by Kuhn-Tucker conditions. This means that the second order condition (A6) is violated at \(A - \bar{A} = A_{max} - \bar{A}\), so the objective for the subproblem (A5) is convex here.

Thus, the conclusions of the proposition follow, given the properties in Proposition 4.1.

**Appendix VII. Understanding Simulations**

The numerical scheme is based on the formal expression of the Bellman equation in Appendix II, where the dynamics is given by

\[
d\pi = \lambda(\pi)dt + \Sigma(\pi, A)dZ^M.
\]

\[
dW = \eta(W, A, c)dt + \sigma_Y(\pi, A)dZ^M \quad \text{where} \quad \eta(W, A, C) = rW + h(A) - u(c). \tag{A8}
\]

\[
dR^{net} = \mu(\pi, A, c)dt + \sigma dZ^M \quad \text{where} \quad \mu(\pi, A, c) = \alpha(\pi) + \beta(\pi)A - c.
\]

For the numerical algorithm, the maximization over \((A, c) \in [A_{min}, A_{max}] \times [0, c_{max}]\) is done by replacing the rectangle with a uniform grid with \(N_A\) and \(N_c\) points in each dimension and performing a brute-force search. \(N_A\) and \(N_c\) are chosen large enough that using a grid with \(2N_A\) and \(2N_c\) gives economically negligible differences in the value and policy functions. This is done for the sake of simplifying the programming while maintaining control over accuracy and ensuring a robust algorithm. More efficient optimization methods are certainly available and would result in dramatic speedup.

Now, introduce a discrete time step \(\Delta t > 0\) and approximate the increment \(dZ^M\) by a two-state random variable taking values \(\pm \sqrt{\Delta t}\) with equal probability. Making these replacements, the discrete time dynamics of the stochastic variables are given by
an explicit Euler approximation to (A8)

\[ d\pi^\pm = \lambda(\pi)\Delta t \pm \Sigma(\pi, A)\sqrt{\Delta t}. \]

\[ dW^\pm = \eta(W, A, c)\Delta t \pm \sigma y(\pi, A)\sqrt{\Delta t} \]  \hspace{1cm} (A9)

\[ dR^{\text{net}} = \mu(\pi, A)\Delta t \pm \sigma d\sqrt{\Delta t}. \]

Using this discrete-time approximation, the value function iteration algorithm of Judd (1998) or Kushner and Dupuis (2001) produces as sequence of value functions \( F^n \) defined at each point \((\pi, W) \in [0,1] \times (0, W_{\text{max}})\).\(^{22}\)

\[ F^{n+1}(\pi, W) = \max \{ \max_{A,C} F^{\text{continue},n}(\pi, W, A, c), \max_{W \leq \bar{W}} F^n(\pi, W'), F^n(\pi, \bar{W}) - C^{\text{ret}}(W) \}, \]  \hspace{1cm} (A10)

where \( F^{\text{continue},n}(\pi, W, A, c) = \mu(\pi, A, c)\Delta t + e^{-\Delta t} \left( \frac{1}{2} F^n(\pi + d\pi^+, W + dW^+) + \frac{1}{2} F^n(\pi + d\pi-, W - dW-) \right) \).

Separate boundary conditions at \( \pi \in \{0,1\} \) are not required since \( \Sigma(\pi, A) = 0 \) and the drift \( \lambda(\pi) \) points toward the center of the domain. At the lower boundary, \( W=0 \), the “renegotiate” option is chosen. The maximal continuation utility, \( W_{\text{max}} \), is taken to be \( C^{\text{ret}}(W) = \frac{u^{-1}(rW)}{r} \), where it is assumed that the agent retires.

To implement the algorithm in (A10), the discrete-time Bellman equation is further discredited in the “spatial” variables \((\pi, W)\). Value functions \( F^n \) are stored on a uniform grid approximating \([0,1] \times [0, W_{\text{max}}]\). The uniform grid has spacing \( h_\pi \) and \( h_W \) so that grid points are defined by \( \pi_i = ih_\pi \) and \( W_j = jh_W \) for \( i = 0, ..., N_\pi \) and \( j = 0, ..., N_W \). Let \( F^n_{i,j} = F^n(\pi_i, W_j) \)

\(^{22}\) It is straightforward to adapt this algorithm to use policy function algorithm and would result in dramatic speedup. Value function iteration is used here to simplify the programming somewhat.
denote the grid function. The $\max_{W \geq W} \text{ in the "renegotiate" component of (A10)}$ is computed by maximizing over $W' \geq W$ restricted to the grid points.

For points $(\pi, W)$ that are not grid points, $F^n(\pi, W)$ is defined by bilinear interpolation. This is consistent with the Markov chain construction of Kushner and Dupuis (2001). For example, let a point $(x, y)$ be bracketed inside grid points $(i, j)$ and $(i + 1, j + 1)$ so that $\pi_i \leq x < \pi_{i+1}$ and $W_j \leq y < W_{j+1}$. Then the bilinear interpolation formula can be written as an expected value of the neighboring grid points with appropriate weights:

$$F^n(x, y) = \left(1 - \frac{x}{h_\pi} - \frac{y}{h_W} + \frac{x}{h_\pi} \frac{y}{h_W}\right) F^n_{i,j} + \left((1 - \frac{y}{h_\pi}) \frac{x}{h_W}\right) F^n_{i+1,j} + \left((1 - \frac{x}{h_\pi}) \frac{y}{h_W}\right) F^n_{i,j+1}$$

$$+ \left(\frac{x}{h_\pi} \frac{y}{h_W}\right) F^n_{i+1,j+1}.$$

We therefore understand the evaluations of $F^n$ in (A10) to be the sum of conditional expectations given the realization of the shock is $\pm \sqrt{\Delta t}$. With this insight, the discrete time process (A9) can be interpreted as a Markov chain on the set of grid points, in which it is possible to transition from each state to up to eight states. Transitions may include fewer than eight neighbors if the points $\pi \pm d\pi, W \pm dW$ share neighboring grid points.

Having constructed an approximating Markov chain for the state variables, the algorithm in (A9) can be shown to converge to a unique solution of the discrete time problem for a given $\Delta t > 0$ as $n \uparrow 0$. This result is based on the contraction mapping theorem and is standard in discrete time/discrete state dynamic programming. Ensuring that the solution to the discrete time/discrete state problem converges to the solution of the continuous time problem as $\Delta t$ and $(h_\pi, h_W)$ go to zero is more delicate. Kushner and Dupuis (2001) have shown that the value and
policy functions for the discrete time/discrete state problem are shown to converge uniformly under the following “local consistency” conditions that should hold as $\Delta t \downarrow 0$:

$$E_t^{A,c}[d\pi] = \lambda(\pi)\Delta t + o(\Delta t),$$

$$E_t^{A,c}[dW] = \eta(W, A, c)\Delta t + o(\Delta t),$$

$$Var_t^{A,c}[d\pi] = \Sigma(\pi, A)^2\Delta t + o(\Delta t),$$

$$Var_t^{A,c}[dW] = \sigma^2 \gamma(\pi, A)^2\Delta t + o(\Delta t),$$

and

$$Cov_t^{A,c}[d\pi, dW] = \sigma\gamma(\pi, A)\Sigma(\pi, A)\Delta t + o(\Delta t).$$

To implement this method, it remains to specify how the time step $\Delta t$ is to be chosen in relation to the grid spacing $(h_\pi, h_W)$ so that the local consistency conditions hold. Following Kushner and Dupuis (2001), we choose a time step that depends on the grid spacing $(h_\pi, h_W)$, the state $(\pi, W)$ an the controls $(A, c)$. Our choice of time step is driven by noting that the dynamics of $\pi$ are determined by a precise rule derived from Bayesian updating, while those of $W$ are derived from an incentive-compatibility constraint and a promise keeping condition. In this light, we give priority to maintain close fidelity to the dynamics of $\pi$ and interpret deviations from local consistency in the dynamics of $W$ as being an optimal choice from a restricted set of contracts resulting from the necessity to use lotteries to keep continuation utilities on grid points while maintaining a discrete analog of the incentive compatibility condition.

Ideally, the time step changes the rate of information flow so that the Markov chain for $\pi$ is a random walk, with zero probability of a state transitioning to itself:

$$\Delta t^{A}_\pi = \frac{h_\pi^2}{\Sigma(\pi, A)^2}. \quad (A11)$$

This implies marginal transition probabilities of $\pi \rightarrow \pi \pm h_\pi$ with probability

$$\frac{1}{2} \left( 1 \pm \frac{h_\pi|\lambda(\pi)|}{\Sigma(\pi, A)^2} \right).$$

These probabilities are sensible provided the drift does not dominate the volatility.
relative to the grid. That is, when $h_\pi |\lambda(\pi)| < \Sigma(\pi, A)^2$ Since $\Sigma(\pi, A)$ is proportional to $\pi(1 - \pi)(A - \tilde{A})$, there are two conditions under which this choice is inappropriate. First, if $\pi$ is close to zero or one, the dynamics of $\pi$ are drift dominated and the interaction between learning and incentive provision is very weak. In this case, we use a time step derived from $dW$. An appropriate choice is $\Delta t^{A,c,K}_\pi$ with $K = 1$, where

$$\Delta t^{A,c,K}_\pi = \frac{K^2 h_W^2}{\sigma^2 (\gamma(\pi, A))^2 + K h_W |\eta(W, A, c)|}.$$  

In general, using $\Delta t \leq \Delta t^{A,c,K}_\pi$ ensures $W$ transitions inside a window within $K \geq 1$ grid points. Second, if the control is chosen so that $(A - \tilde{A})$ is very small, the implied time step $\Delta t^{A}_\pi$ blows up. This would invalidate any approximations made for small $\Delta t$. To exclude this possibility, we employ two strategies. First, restrict the set of controls to exclude controls where $|A - \tilde{A}|$ is above a given tolerance $\varepsilon_A$, which goes to zero along with the grid spacing. This choice maintains the local consistency conditions, but rules out some choices of controls that may be economically relevant. Another choice is to modify the time step, taking $\Delta t^{A,e}_\pi = h_\pi^2 / \Sigma(\pi, A)^2 + h_\pi \varepsilon_\pi$. This ensures that $\Delta t^{A,e}_\pi \leq h_\pi / \varepsilon_\pi$ uniformly across all states and controls, but taking $\varepsilon_\pi$ too large introduces significant randomization into the dynamics of $\pi$, which manifests itself in “numerical noise” in the results. Combining the two approaches, we can bound the time step uniformly in $A$ for a given state $\pi$ by $\Delta t^{A,e}_\pi \leq h_\pi^2 / (\tau(\pi)^2 \varepsilon_A^2 + h_\pi \varepsilon_\pi)$.  

We have defined a candidate time step for cases in which $\pi$ is drift dominated or not. The former operates for $\pi$ near the boundary and when when $A \approx \tilde{A}$ and the latter for $\pi$ in the deep interior of the grid. For grid points with $w_j$ close to the upper or lower boundaries, there remains the possibility that one or both of the $W + dw^\pm$ may be outside the grid. We introduce a
parameter $K>1$ and insist that for a given $(\pi, W, A, c)$ the time step be small enough so that $W + dw^\pm$ stays inside the grid and even inside a window of width $K$ so that $-Kh_W \leq dw^\pm \leq Kh_W$. We therefore take the step size $\Delta t^{W,A,c,e}_n$ to be the smaller of the candidate time step and a fallback time step defined to be the largest $\Delta t$ such that $W + dw^\pm \in [\max(W_{\min}, W - Kh_W), \min(W_{\max}, W + Kh_W)]$. Except near the $W \in \{0, W_{\max}\}$ boundaries, the fallback time step is unnecessary for sufficiently large $K$.

Except at points near the boundary of $W$, time steps chosen this way result in locally consistent transition probabilities as $(h_{\pi}, h_W) \downarrow 0$ and $K \uparrow \infty$ and $(\varepsilon_A, \varepsilon_W) \downarrow 0$. In practice, these parameters should be carefully chosen to ensure that the time step in operation is as close as possible to (A11) without compromising the stability of the algorithm.
Figure 1. Game Tree of a Two-Period Moral Hazard Model with Learning.
Figure 2. Value of Learning and Optimal Contracts for a Two-Period Moral Hazard Model with Learning. (a) Optimal contract, \( b_0 \) vs. value of high outcome, \( \theta \) (b) Shareholder value, \( \Pi_0 \), vs. initial contract, \( b_0 \), at point A (\( \theta = 1.08 \)) (c) Shareholder value, \( \Pi_0 \), vs. initial contract, \( b_0 \), at point B (\( \theta = 1.43 \)) (d) Shareholder value, \( \Pi_0 \), vs. initial contract, \( b_0 \), at point C (\( \theta = 1.46 \)) (e) Shareholder value, \( \Pi_0 \), vs. initial contract, \( b_0 \), at point D (\( \theta = 1.50 \)).
Figure 3. Myopic Managerial Effort.
The parameters are $\alpha = 1$, $\beta=0.75$, $\bar{A} = 0.65$, $\Lambda=0.1$, $\pi_0 = 0.45$, $H_1 = 0.5$, $H_2 = 0.95$, $b=0.1$, $r = 0.1$, $\sigma = 2$.

Figure 4. Extreme and Moderate Experimentation Relative to Myopic Actions When Manager Is Risk Neutral.
The parameters are $\alpha = 40$, $\beta=0.66$, $\bar{A} = 24$, $\Lambda=0.1$, $\pi_0 = 0.5$, $H_1 = 0.1$, $H_2 = 0.035$, $b=0.05$, $r = 0.05$, $\sigma = 10$. $\bar{A} \in [0,60]$. 
**Figure 5.** The Difference between the Second Order Conditions Associated with the Myopic Effort and the Optimal Effort.

The parameters are $\alpha = 1$, $\beta = 0.75$, $\bar{A} = 0.65$, $\Lambda = 0.1$, $\pi_0 = 0.45$, $H_1 = 0.5$, $H_2 = 0.95$, $b = 0.1$, $r = 0.1$, $\sigma = 2$.

**Figure 6.** Difference between Myopic and Optimal Effort.

The parameters are $\alpha = 1$, $\beta = 0.75$, $\bar{A} = 0.65$, $\Lambda = 0.1$, $\pi_0 = 0.45$, $H_1 = 0.5$, $H_2 = 0.95$, $b = 0.1$, $r = 0.1$, $\sigma = 2$. 

$A^{\text{myopic}}(\pi, w) - A(\pi, w)$
Figure 7. Simulation of Extreme Experimentation When Manager Is Risk Neutral. The parameters are $\alpha = 40$, $\beta = 0.66$, $A = 24$, $\Lambda = 0.1, \pi_0 = 0.5$, $H_1 = 0.1$, $H_2 = 0.035$, $b = 0.05$, $r = 0.05$, $\sigma = 10$. $A \in [0,60]$.

Figure 8. Second Order Conditions For Extreme and Moderate Experimentation under Risk Neutral Manager. The parameters are $\alpha = 40$, $\beta = 0.66$, $A = 24$, $\Lambda = 0.1, \pi_0 = 0.5$, $H_1 = 0.1$, $H_2 = 0.035$, $b = 0.05$, $r = 0.05$, $\sigma = 10$. $A \in [0,60]$.
Figure 9. Optimal Managerial Effort.
The parameters are $\alpha = 1$, $\beta = 0.75$, $\bar{A} = 0.65$, $\Lambda = 0.1$, $\pi_0 = 0.45$, $H_1 = 0.5$, $H_2 = 0.95$, $b = 0.1$, $r = 0.1$, $\sigma = 2$.

Figure 10. Firm’s Value under Optimal Contract.
The parameters are $\alpha = 1$, $\beta = 0.75$, $\bar{A} = 0.65$, $\Lambda = 0.1$, $\pi_0 = 0.45$, $H_1 = 0.5$, $H_2 = 0.95$, $b = 0.1$, $r = 0.1$, $\sigma = 2$. 
Figure 11. Manager’s Consumption Under Optimal Contract.
The parameters are $\alpha = 1, \beta=0.75, \bar{A}=0.65, \Lambda=0.1, \pi_0 = 0.45, H_1 = 0.5, H_2 = 0.95, b=0.1, r = 0.1, \sigma = 2$.

Figure 12. Simulation of Moderate Experimentation When Manager Is Risk Neutral.
The current state of business is plotted in the first graph along with the conditional expectation of the state of business. The optimal and myopic effort is in the lower left plot as the state of business changes. The other two graphs provide the promised utility to the manager as well as her consumption. The parameters are $\alpha = 1, \beta=0.75, \bar{A}=0.65, \Lambda=0.1, \pi_0 = 0.45, H_1 = 0.5, H_2 = 0.95, b=0.1, r = 0.1, \sigma = 2$. 
Figure 13. The Second Order Conditions for Moderate Experimentation.
The second order conditions under both optimal and myopic policy are plotted for the same simulation of the state of business. The parameters are $\alpha = 1$, $\beta = 0.75$, $\bar{A} = 0.65$, $\Lambda = 0.1$, $\pi_0 = 0.45$, $H_1 = 0.5$, $H_2 = 0.95$, $b = 0.1$, $r = 0.1$, $\sigma = 2$. 
Guide for Referee

This guide matches up the analysis in sections A1, A2, A3 of Sannikov (2008). We tried to keep the notation as close as possible to Sannikov. There is no new analysis, so that we do not expect to publish this part.

Proposition R1. There exists a progressively measurable process $Y_t = \{Y_t, F_t^M; 0 \leq t < \infty\}$ is in $\mathcal{L}^2$ such that

$$dW_t = \left(r W_t - u(C_t) + h(A_s)\right)dt + \sigma Y_t dZ_t^M, \text{ when } W_t > 0.$$  \hspace{1cm} (R1)

Proof: The agent’s total expected payoff under strategy $(C, A)$ given the information at time $t$ is

$$V_t = \int_0^t e^{-rs}(u(C_s) + h(A_s)) ds + e^{-rt}W_t(C, A)$$  \hspace{1cm} (R2)

which is a $Q^M$ martingale, as it is based on the manager’s information and her beliefs about the hidden state variable. Assuming that the filtration satisfies the appropriate policies, the $Q^M$ martingale $V_t$ must be right continuous with the left limit constraints given by Theorem 1.3.13 of Karatzas and Shreve (1991). By the martingale representation theorem, Karatzas and Shreve (1991, p.182) show that there exists a progressively measured process $Y_t = \{Y_t, F_t^M; 0 \leq t < \infty\}$ such that

$$E \int_0^T Y_t^2 dt < \infty \text{ for every } T \in (0, \infty),$$

and

$$V_t = V_0 + \int_0^t e^{-rs} Y_s dZ_t^M, \ t \in [0, \infty).$$  \hspace{1cm} (R3)

Here

$$Z_t^M = \frac{1}{\sigma} \int_0^t \left((\alpha(\pi_s, W_s) - \bar{A}) \Delta \beta(k_t - \pi_t) ds + \sigma dZ_s\right)$$  \hspace{1cm} (R4)

is Brownian motion in $Q^M$. (R4) implies (5).

Also $\pi_t$ is given by
\[
\pi_t = \int_0^t \left( \frac{(1-\pi_s)\pi_s (\Delta \beta)^2}{\sigma^2} (A_s - \bar{A})^2 (k_s - \pi_s) - \Lambda (\pi_s - \bar{\pi}) \right) ds + \int_0^t \Sigma (\pi_s, A_s) dZ_s.
\]

Thus, (6) is true.

Differentiation of (R2) and (R3) with respect to time \( t \) yields

\[
dV_t = e^{-rt} Y_t \sigma dZ_t^M = e^{-rt} \left( u(c_t) - h(A_t) \right) dt + d \left( e^{-rt} W_t (C_t, A_t) \right)
\]
\[
= e^{-rt} \left( u(c_t) - h(A_t) \right) dt - r e^{-rt} W_t (C_t, A_t) dt + e^{-rt} dW_t (C_t, A_t).
\]

Consequently, we have

\[
dW_t (C_t, A_t) = (r W_t (C_t, A_t) - u(c_t) + h(A_t)) dt + Y_t \sigma dZ_t^M, \tag{R5}
\]

which is (R1).

**Proposition R2.** For a given \( A_t \), let \( Y_t \) be the process from Proposition R1 that represents the continuation \( W_t (C_t, A_t) \). The pay for performance process \( Y_t \) satisfies the incentive compatibility condition

\[
\alpha(\pi_t) + \beta(\pi_t) A_t - h(A_t) \geq \alpha(\pi_t) + \beta(\pi_t) a - h(a), \forall a \in \mathcal{A}, t \in [t_0, T_{Ret}] \tag{R6}
\]

almost everywhere. Here \( \mathcal{A} \) is the compact set of the manager’s possible efforts.

Proof: Consider an arbitrary alternative strategy \( A^* \) defined by

\[
\bar{V}_t = V_0 + \int_0^t e^{-rs} (u(c_s) - h(A_s^*)) ds + e^{-rt} W_t (C_t, A_t).
\]

The expectation of the agent’s total payoff (at time \( t \)) is the sum of the manager’s cost of effort from the strategy \( A^* \) until time \( t \), and plans to follow the strategy \( A_t \) after time \( A_t \). The drift of the process \( \bar{V}_t \) under the probability measure \( Q^{M^*} \) for the alternative strategy is

\[
d\bar{V}_t = e^{-rt} \left( u(c_t) - h(A_t^*) \right) dt + d \left( e^{-rt} W_t (C_t, A_t) \right)
\]
\[
= e^{-rt} \left( -h(A_t^*) + h(A_t) \right) + \Delta \beta \pi Y_t (A_t^* - A_t) dt + e^{-rt} Y_t \sigma dZ_t^M. \tag{R7}
\]

In the second step, we use (R5). In the last step the Brownian motion under \( Q^{M^*} \) and \( Q^M \) are
related by
\[ \sigma Z_t^M = \sigma Z_t^{M^*} + \int_0^t \Delta \beta \pi (A_s^* - A_s) \, ds \quad (R8) \]
following which the revenue must satisfy (1) conditional on the estimated probability of good revenue. Notice that the manager cannot use \( k \), since it is not known by either the manager or the shareholder.

Assume that (R1) does not hold on a set of positive measure. Now, choose an \( A_t^* \) that maximizes \( \Delta \beta \pi Y_t A_t^* - h(A_t^*) \) for all \( t > 0 \). The drift of \( d\tilde{V}_t \) in (R7) under \( Q^{M^*} \) is non-negative and positive on a set of positive measure. Consequently, there exists a time \( t > 0 \) such that
\[ E^{M^*}(\tilde{V}_t) > \tilde{V}_0 = W_0(C, A). \quad (R9) \]
This follows by integrating (R7) from 0 to \( t \) and using the positive drift. Because the manager gets utility if the manager follows \( A^* \) until time \( t \) and then switches to \( A \), the strategy \( A \) is suboptimal. Thus, we have a contradiction, so that (R1) holds.

Now prove the opposite. Suppose (R1) holds for the strategy \( A \). Then \( \tilde{V}_t \) is a \( Q^{M^*} \) supermartingale for any alternative strategy \( A^* \) by (R9). Moreover, since \( W(C, A) \) is bounded from below by zero, we can add
\[ \tilde{V}_\infty = \lim_{t \to \infty} \int_0^t e^{-rs}(u(c_s) - h(A_s^*)) \, ds + \lim_{t \to \infty} e^{-rt}W_t(C_t, A_t) \quad (R10) \]
as the last element of the supermartingale \( \tilde{V} \). (See Karatzas and Shreve (1991), problem 3.16 page 18.) Therefore,
\[ W_0(C, A) = \tilde{V}_0 \geq E^{M^*}(\tilde{V}_t) = W_0(C, A^*). \quad (R11) \]
So, the strategy \( A \) is at least as good as \( A^* \).

Reference