The role of accounting in a dynamic model of CEO pay and turnover

Judson Caskey, George-Levi Gayle, and Chen Li

Address

December 6, 2017
The role of accounting in a dynamic model of CEO pay and turnover

Abstract

(Abstract)

Keywords:


1 Model

1.1 Setup

Our model extends DeMarzo and Sannikov (2017) by including a CEO turnover decision, an earnings process, and multiple firms. There are \( N \) firms in the economy. Each firm \( n \in \{1, 2, \ldots, N\} \) has a CEO and sets compensation to optimize its net cash flows. We denote the firm’s time \( t \) cumulative cash flows by \( x_{nt} \), and its time \( t \) flow compensation to the agent by \( c_t \). The firm’s cash flows depend on an unobservable industry-wide component \( \mu_{0t} \) and the CEO’s ability \( \mu_{nt} \). Both of these are unobservable so that firm value from the principal’s perspective depends on estimates \( \hat{\mu}_{0t} = E_t[\mu_{0t}] \) and \( \hat{\mu}_{nt} = E_t[\mu_{nt}] \). The firm incurs a cost \( k \) when replacing the CEO, at which time it hires a replacement CEO with ability normalized to zero. We denote the current CEO’s continuation value by \( w_t \).

We can write the firm’s value \( b_t = b(\hat{\mu}_{0t}, \hat{\mu}_{nt}, w_t) \) as follows where \( \tau_i \) denotes the \( i \)th stopping time to replace the CEO and the discount rate is \( r \):

\[
b(\hat{\mu}_{0t}, \hat{\mu}_{nt}, w_t) = E_t \left[ \int_t^{\tau_1} e^{-r(s-t)} \left( dx_{ns} - c_s ds \right) + \sum_{i=1}^{\infty} \left( \int_{\tau_i}^{\tau_{i+1}} e^{-r(s-t)} \left( dx_{ns} - c_s ds \right) - e^{-r(\tau_{i+1}-t)}k \right) \right]
\]

\[
= E_t \left[ \int_t^{\tau_1} e^{-r(s-t)} \left( dx_{ns} - c_s ds \right) + e^{-r(\tau_1-t)} \left( b(\hat{\mu}_{0\tau_1}, 0, w_{\tau_1}) - k \right) \right].
\]

(1)

The second line incorporates the recursive nature of the firm’s payoff.

The firm’s cash flows depend on random shocks, the CEO’s effort, the CEO’s ability, and CEO turnover decisions. Each firm \( n \) has cumulative cash flows \( x_{nt} \) at time \( t \) that evolve as follows:

\[
dx_{nt} = (\mu_{0t} + \mu_{nt} - a_{nt}) dt + \sigma_x (\beta d z_{0xt} + d z_{nxt}), \quad x_{n0} = 0,
\]

(2)
where \(a_{nt} \geq 0\) is the CEO’s unobservable ‘bad’ action (e.g., shirking), \(\mu_{0t}\) is the unobservable industry profitability, \(\mu_{nt}\) is the CEO’s unobservable ability, and \(z_{0xt}\) \(\text{and} \ z_{nxt}\) are independent standard Brownian motions that represent shocks to the respective common and idiosyncratic portions of cash flows. The parameter \(\sigma_x\) scales volatility and \(\beta\) scales the common versus idiosyncratic portions of volatility. The profitability processes evolve as follows:

\[
d\mu_{0t} = \beta \sigma_\mu d\mu_{zt}, \quad d\mu_{nt} = \sigma_\mu d\mu_{zt},
\]

(3)

where \(\sigma_\mu\) is a parameter and \(z_{0\mu t}\) \(\text{and} \ z_{n\mu t}\) are independent standard Brownian motions that represent shocks to the respective industry and CEO-specific portions of profitability. The CEO’s ability is idiosyncratic.

The firm also has a cumulative earnings processes \(e_{nt}\) that evolves as follows:

\[
d e_{nt} = \theta (x_{nt} - e_{nt}) dt + \sigma_{e} (\beta d z_{0et} + d z_{net}), \quad e_{n0} = 0,
\]

(4)

where the parameter \(\theta\) governs how quickly accruals reverse in the sense of cumulative earnings converging to cumulative cash flows, \(\sigma_{e}\) is a volatility parameter, and \(z_{0et}\) \(\text{and} \ z_{net}\) are independent standard Brownian motions that represent shocks to the respective common and idiosyncratic portions of cash flows.

We assume that investors also directly observe the industry-level components of cash flows and earnings, \(x_{0t}\) and \(e_{0t}\), with:

\[
d x_{0t} = \mu_{0t} dt + \beta \sigma_x d z_{0xt}, \quad d e_{0t} = \theta (x_{0t} - e_{0t}) dt + \sigma_{e} d z_{0et}.
\]

(5)

Directly observing the industry-level cash flows and earnings is equivalent to observing the vector of all firms’ cash flows and earnings where the number of firms \(N\) approaches infinity. This simplification allows for tractable solutions to the filtering problem to infer the firm- and industry-level profitability.
The agent obtains a benefit $\lambda a_t$ from shirking, where $\lambda \in (0, 1)$ ($\lambda \in (0, \frac{\mu}{\mu + r})$ so that the action is socially wasteful) so that shirking is socially wasteful. In addition, if the contract calls for shirking of $a_t$ and the agent engages in shirking $\hat{a}_t \neq a_t$, then this distorts the principal’s process for learning the CEO ability $\mu_{nt}$. In this case, the agent’s beliefs $\hat{\mu}_{nt}^a$ do not equal the principal’s beliefs $\hat{\mu}_{nt}$. If the agent leaves the firm when the agent’s and principal’s beliefs are $\hat{\mu}_{nt}^a$ and $\hat{\mu}_{nt}$, respectively, then the present value of his future payoffs is $\hat{R}(\hat{\mu}_{\tau}, \hat{\mu}_{\tau}^a)$. Denoting by $\tau$ the stopping time at which the agent leaves the firm, the agent’s continuation value is then the following where $E_t^a[\cdot]$ denotes expectation with respect to the agent’s beliefs:

$$w_t^a = E_t^a\left[\int_{t}^{\tau} e^{-r(s-t)}(\lambda a_s + c_s)ds + e^{-r(\tau-t)}\hat{R}(\hat{\mu}_{\tau}, \hat{\mu}_{\tau}^a)\right].$$ (6)

### 1.2 The filtering problem

The following presents the steady-state beliefs $\hat{\mu}_{0t}$ and $\hat{\mu}_{nt}$ for the case where the agent takes action $a_t \geq 0$. We later show that $a_t = 0$ in equilibrium. The appendix presents the derivations. We assume that all parties share the following priors where $\mu_{Nt} = \{\mu_{1t}, \mu_{2t}, \ldots, \mu_{Nt}\}$, $1$ denotes a vector of ones, $0$ denotes a vector of zeros, and $I$ denotes the identity matrix:

$$(\mu_{00}^\mu \mu_{N0}^\mu) \sim \mathcal{N}\left(\left(\begin{array}{c} \hat{\mu}_{00} \\ 0 \end{array}\right), \left(\begin{array}{cc} \hat{\gamma}_0 & \hat{\gamma}_n \varepsilon \hat{\gamma}_n \varepsilon \\ 0 & 0 \end{array}\right)\right).$$ (7)

In other words, all correlated profitability comes via the $\mu_{0t}$ term and the $\mu_{nt}$ terms are purely firm-specific.

---

1DeMarzo and Sannikov (2017) place the weaker restriction $\lambda \in (0, 1)$. We require the more stringent upper bound on $\lambda$ because $\lambda > \frac{\mu}{\mu + r}$ will cause the principal’s payoff to decrease in profitability for sufficiently high levels of profitability due to the costs of compensating the agent.
We denote the shock correlations as follows:

\[
\begin{align*}
\rho_{\mu x} &= \frac{1}{dt} E[dz_{0\mu}dz_{0xt}] = \frac{1}{dt} E[dz_{n\mu}dz_{nxt}], \\
\rho_{\mu e} &= \frac{1}{dt} E[dz_{0\mu}dz_{0et}] = \frac{1}{dt} E[dz_{n\mu}dz_{net}], \\
\rho_{xe} &= \frac{1}{dt} E[dz_{0xt}dz_{0et}] = \frac{1}{dt} E[dz_{nxt}dz_{net}].
\end{align*}
\]

(8)

The steady-state variances are:

\[
\begin{align*}
\hat{\gamma}_n &= \lim_{t \to \infty} \text{var}_t(\mu_{nt}) = \sigma_\mu \sigma_x \left( \sqrt{(1 - \rho_{\mu e}^2)(1 - \rho_{xe}^2)} - (\rho_{\mu e} - \rho_{\mu e} \rho_{xe}) \right) \\
&= \sigma_\mu \sigma_x \left( \sqrt{(1 - \rho_{\mu e}^2)(1 - \rho_{xe}^2)} \frac{1}{\sqrt{1 - \rho_{\mu e}^2}(1 - \rho_{xe}^2)} \right) > 0,
\end{align*}
\]

(9)

\[
\begin{align*}
\hat{\gamma}_0 &= \lim_{t \to \infty} \text{var}_t(\mu_{0t}) = \beta^2 \hat{\gamma}_n.
\end{align*}
\]

The belief processes evolve as:

\[
\begin{align*}
\text{d}\hat{\mu}_{0t} &= \beta \sigma_\mu \text{d}\hat{z}_{0\mu t}, \\
\text{d}\hat{\mu}_{nt} &= \sigma_\mu \text{d}\hat{z}_{n\mu t},
\end{align*}
\]

(10)

where:

\[
\begin{align*}
\text{d}\hat{z}_{0\mu t} &= \nu \frac{\sigma_x}{\sigma_\mu} \frac{1}{\sqrt{\sigma_x}} \left( \text{d}x_{0t} - \hat{\mu}_{0t} \text{d}t \right) + \frac{\rho_{\mu e} \sigma_x - \nu \rho_{xe} \sigma_e}{\sigma_\mu} \frac{1}{\beta \sigma_e} \left( \text{d}e_{0t} - \theta \left( x_{0t} - e_{0t} \right) \right) \\
\text{d}\hat{z}_{n\mu t} &= \nu \frac{\sigma_x}{\sigma_\mu} \frac{1}{\sqrt{\sigma_x}} \left( \text{d}x_{nt} - (\hat{\mu}_{0t} + \hat{\mu}_{nt} - a_{nt}) \text{d}t - \beta \sigma_x \text{d}\hat{z}_{0xt} \right) \\
&+ \frac{\rho_{\mu e} \sigma_x - \nu \rho_{xe} \sigma_e}{\sigma_\mu} \frac{1}{\beta \sigma_e} \left( \text{d}e_{nt} - \theta \left( x_{nt} - e_{nt} \right) \text{d}t - \beta \sigma_e \text{d}\hat{z}_{0et} \right),
\end{align*}
\]

(11)

where \( \nu = \frac{\sigma_x}{\sigma_\mu} \sqrt{\frac{1 - \rho_{\mu e}^2}{1 - \rho_{xe}^2}} \). On the equilibrium path \( \text{d}\hat{z}_{0\mu t}, \text{d}\hat{z}_{n\mu t}, \text{d}\hat{z}_{0xt}, \text{d}\hat{z}_{0xt} \) are Brownian motions and we can express the dynamics of cash flows in terms of observables:

\[
\text{d}x_{0t} = \hat{\mu}_{0t} \text{d}t + \beta \sigma_x \text{d}\hat{z}_{0xt}, \quad \text{d}x_{nt} = (\hat{\mu}_{0t} + \hat{\mu}_{nt}) \text{d}t + \sigma_x (\beta \text{d}\hat{z}_{0xt} + \text{d}\hat{z}_{nxt}).
\]

(12)

If the agent deviates from the equilibrium action \( a_{nt} \) by taking action \( \hat{a}_{nt} \), then the principal’s beliefs \( \hat{\mu}_{nt} \) continue to follow (10), while the agent’s beliefs \( \hat{\mu}_{nt}^\sigma \) follow:
\[
\begin{align*}
\text{d}\hat{\mu}_{nt}^a &= \sigma_{\mu} \left( \sqrt{1 - \rho_{nx}^2} \frac{1}{\rho_{x\sigma}} (dx_{nt} - (\hat{\mu}_{0t} + \hat{\mu}_{nt}^a - \hat{a}_{nt}) dt - (dx_{0t} - \hat{\mu}_{0t} dt)) 
+ \left( \rho_{\mu e} - \rho_{xe} \sqrt{1 - \rho_{nx}^2} \frac{1}{1-\rho_{nx}^2} \right) d\zeta_{net} \right) 
\end{align*}
\]
\[
= \text{d}\hat{\mu}_{nt} + \nu(\hat{a}_{nt} - a_{nt} - (\hat{\mu}_{nt}^a - \hat{\mu}_{nt})) dt.
\]

The above dynamics imply the belief divergence \( \hat{\mu}_{nt}^a - \hat{\mu}_t \) is:

\[
\hat{\mu}_{nt}^a - \hat{\mu}_t = \nu \int_0^t e^{-\nu(t-s)} (\hat{a}_{ns} - a_{ns}) ds.
\]

The term \( \nu \) is the rate of decay of the agent’s information advantage. In DeMarzo and Sannikov (2017), \( \nu = \frac{\sigma_{\mu}}{\sigma_x} \). In our setting, the introduction of earnings can increase or decrease \( \nu \) from that benchmark. For example, if earnings shocks are highly correlated with shocks to the profitability process (high \( \rho_{\mu e} \)), but not with the cash flow shocks (low \( \rho_{xe} \)), then earnings will be highly informative about profitability but relatively less useful for controlling agency costs (low \( \hat{\gamma}_n \) and \( \nu < \frac{\sigma_{\mu}}{\sigma_x} \)).

### 1.3 Contracting and the first-best solution

The firm pays flow compensation \( c_t \) to the CEO, which must satisfy \( c_t \geq \zeta \geq 0 \). We follow DeMarzo and Sannikov (2017) and use a linear form for the CEO’s termination payoff:\(^2\)

\[
\hat{R}(\hat{\mu}_t, \hat{\mu}_t^a) = R_0 + R_\mu \hat{\mu}_{nt} + \frac{\lambda}{\nu} (\hat{\mu}_{nt}^a - \hat{\mu}_{nt}),
\]

where the term \( \frac{\lambda}{\nu} (\hat{\mu}_{nt}^a - \hat{\mu}_{nt}) \) reflects that the CEO can earn \( \lambda \) from the deviation \( \hat{\mu}_{nt}^a - \hat{\mu}_{nt} \) given in (14) in perpetuity. Because \( R_0 \) and \( R_\mu \) are fixed, this specification omits the possibility that the CEO’s outside opportunities change during the

\(^2\)DeMarzo and Sannikov (2017) include a term for the portion of the agent’s information rents that vanish upon termination. We exclude that term both to facilitate estimation and to reflect the notion that \( \mu_{nt} \) reflects CEO ability.
contract period. In other words, the firm can anticipate any ‘poaching’ of the CEO. Furthermore, principals do not choose $R_0$ and $R_\mu$, so that we do not treat the CEO’s termination value as being paid by the firm. As in DeMarzo and Sannikov (2017), we assume that $R_\mu \in [0, \lambda \left( \frac{1}{\nu} + \frac{1}{\gamma} \right))$ or, equivalently, $\lambda > R_\mu / \left( \frac{1}{\nu} + \frac{1}{\gamma} \right)$.

With no agency costs ($a_t = 0$), the firm-specific portion of the total surplus is the following:

$$
v_n (\hat{\mu}_{nt}) = E_t \left[ \int_t^\tau e^{-r(s-t)} \hat{\mu}_{ns} ds + e^{-r(\tau-t)} (v_n (0) + R (\hat{\mu}_{nt}) - k) \right]
= \frac{1}{\tau} \hat{\mu}_{nt} + E_t \left[ e^{-r(\tau-t)} (v_n (0) - \frac{1}{\tau} \hat{\mu}_{nt} + R (\hat{\mu}_{nt}) - k) \right]
$$

(16)

and the total surplus is $\frac{1}{\tau} \hat{\mu}_{0t} + v_n (\hat{\mu}_{nt})$. The myopic policy compares keeping the current CEO in perpetuity for a value of $\frac{1}{\tau} \hat{\mu}_{nt}$ to terminating the CEO for the net benefit of $R(\hat{\mu}_{nt}) - k$ and hiring a new CEO with ability normalized to zero. This gives the myopic termination threshold:

$$
\frac{1}{\tau} \hat{\mu}_o = R(\hat{\mu}_o) - k \quad \Rightarrow \quad \hat{\mu}_o = -\frac{k - R_0}{\frac{1}{\tau} - R_\mu}.
$$

(17)

Optimal termination policies consider the value of the termination option and have lower thresholds.

The contract is a $(x_{0t}, e_{0t}, x_{nt}, e_{nt})$ measurable pair $(c_t, \tau)$ where $c_t$ denotes cash payments, and the stopping time $\tau$ denotes the time of contract termination. Given the filtering process described in expressions (10) through (12), we can write the firm’s payoff as follows where $w_0$ is the CEO’s expected value of working for the firm:

$$
b (\hat{\mu}_{0t}, \hat{\mu}_{nt}, w_t) = \frac{1}{\tau} \hat{\mu}_{0t} + E_t \left[ \int_t^\tau e^{-r(s-t)} (\hat{\mu}_{ns} - a_s - c_s) ds + e^{-r(\tau-t)} (b_n (0, w_0) - k) \right].
$$

(18)

At each termination date $\tau_i$, the firm selects a compensation process $c_t$, an action process $a_t$, and a termination policy $\tau_{i+1}$ to optimize the above payoff subject to the
following participation, and incentive compatibility:

\[
\begin{align*}
\frac{w_0}{u_0} & \leq \mathbb{E} \left[ \int_{\tau_i}^{\tau_{i+1}} e^{-r(s-\tau_i)} (\lambda a_s + c_s) \, ds + e^{-r(\tau_{i+1}-\tau_i)} \hat{R}(\hat{\mu}_{n\tau_{i+1}}) \right], \quad (19a) \\
w_t &= \mathbb{E} \left[ \int_{t}^{\tau_{i+1}} e^{-r(s-t)} (\lambda a_s + c_s) \, ds + e^{-r(\tau_{i+1}-t)} \hat{R}(\hat{\mu}_{n\tau_{i+1}}) \right] \geq R(\hat{\mu}_{nt}), \quad (19b) \\
w_{\tau_i} &\geq \mathbb{E} \left[ \int_{\tau_i}^{\tau_{i+1}} e^{-r(s-\tau_i)} (\lambda \hat{a}_s + c_s) \, ds + e^{-r(\tau_{i+1}-\tau_i)} \hat{R}(\hat{\mu}_{n\tau_{i+1}}, \hat{\mu}_{n\tau_{i+1}}) \right], \forall \hat{a}_t (19c)
\end{align*}
\]

We further restrict lump-sum payments to be nonnegative. The constraint (19a) is a participation constraint requiring that the new agents expect value of at least \( w_0 \) to accept the contract. The constraint (19b) avoids the agent leaving early under the prescribed actions \( a_t \) and termination policy. The constraint (19c) implements the required actions \( a_t \). DeMarzo and Sannikov (2017) show that the (19c) can be written as of the initial date and need not include the choice to terminate early.

With \( a_t = 0 \), the firm-specific portion of the principal’s surplus is:

\[
b_n(\hat{\mu}_{nt}, w_t) = \mathbb{E}_t \left[ \int_{t}^{\tau_{i+1}} e^{-r(s-t)} (\mu_{ns} - c_s) \, ds + e^{-r(\tau_{i+1}-t)} (b_n(0, w_0) - k) \right] \\
= \mathbb{E}_t \left[ \int_{t}^{\tau_{i+1}} e^{-r(s-t)} \mu_{ns} \, ds + e^{-r(\tau_{i+1}-t)} (b_n(0, w_0) + R(\hat{\mu}_t) - k) \right] - w_t \quad (20)
\]

The following proposition summarizes the termination rule \( \mu_{*b} \) that maximizes the surplus \( v_n \), and firm’s preferred termination policy \( \mu_{**} \) that maximizes the firm’s payoff \( b_n \) when only the participation constraint must be considered. We later use \( \mu_{*} \) to denote the firm’s choice of termination policy under the full set of constraints.

**Proposition 1.** The first-best termination policy is:

\[
\mu_{fb} = \mu_{o} - \frac{\sigma_{\mu}}{\sqrt{2\pi}} \left( 1 + \omega \left( -\sqrt{2\pi} \left( \mu_{o} - \frac{\sigma_{\mu}}{\sqrt{2\pi}} \right) \right) \right) < \mu_{o}, \quad (21)
\]

where \( \omega(\bullet) \) denotes Lambert’s W function, and \( \omega(\bullet) \) increases from \(-1\) at \( k = R_0 \) to
zero as \( k - R_0 \to \infty \). The value function under the surplus-maximizing policy is:

\[
v_n(\hat{\mu}_{nt}; \mu_{fb}) = \frac{1}{\tau} \hat{\mu}_{nt} + \frac{\sigma_{\mu}}{\sqrt{2\tau}} \left( \frac{1}{\tau} - R_{\mu} \right) e^{-\sqrt{2\tau}(\hat{\mu}_{nt} - \mu_{fb})/\sigma_{\mu}}, \tag{22}
\]

where \( e^{-\sqrt{2\tau}(\hat{\mu}_{nt} - \mu_{fb})/\sigma_{\mu}} = E_t \left[ e^{-r(\tau-t)} \right] \) reflects the discounted stopping time.

The firm’s preferred termination policy \( \mu_{**} \) (including only the participation constraint) is:

\[
\mu_{**} = \mu_{0} - \frac{1}{\tau - R_{\mu}} w_0 - \frac{\sigma_{\mu}}{\sqrt{2\tau}} \left( 1 + \omega \left( -e^{-\sqrt{2\tau} \left( \frac{\mu_{0} - \frac{1}{\tau - R_{\mu}} w_0 - \sigma_{\mu}}{\sqrt{2\tau}} \right)} \right) \right) < \mu_{fb}. \tag{23}
\]

The value of firm-specific profits dates is:

\[
b_n(\hat{\mu}_{nt}; w_0; \mu_{**}) = \frac{1}{\tau} \hat{\mu}_{nt} + \frac{\sigma_{\mu}}{\sqrt{2\tau}} \left( \frac{1}{\tau} - R_{\mu} \right) e^{-\sqrt{2\tau}(\hat{\mu}_{nt} - \mu_{**})/\sigma_{\mu}} - w_0. \tag{24}
\]

This value can be implemented with a contract that pays the agent \( c_t = r w_0 \) at all dates \( t < \tau \) and a lump sum of \( w_0 - R(\mu_{**}) \) at contract termination.

When \( k = R_0 \), a zero threshold maximizes surplus because it is costless to immediately fire the CEO given the slightest evidence that his ability \( \hat{\mu}_{nt} \) falls short of the average of zero. As the cost of firing \( k - R_0 \) becomes large, the surplus-maximizing threshold approaches \( \mu_{0} - \frac{\sigma_{\mu}}{\sqrt{2\tau}} \) as in standard termination problems without replacement. The value function in the first-best case does not depend on information quality. It is not until we introduce the potential for moral hazard \((a_t > 0)\) that information quality plays a role in payoffs.

The firm prefers a lower threshold than the surplus-maximizing threshold. This differs from prior studies where both the firm and the agent have a similar horizon and the firm prefers the surplus-maximizing threshold when there are no agency costs. In our setting, the firm faces costs from future agents that it cannot mitigate by contracting with the current agent. The firm prefers to delay these costs, and the
firm’s preferred threshold $\mu_{**}$ equals the surplus-maximizing threshold $\mu_{fb}$ only if the agent has a zero reservation value ($w_0 = 0$). In DeMarzo and Sannikov (2017), the firm deals with only one generation of agents and the firm will prefer the surplus-maximizing termination threshold in the absence of moral hazard.

### 2 Fixed termination contracts

We consider contracts that have a fixed termination threshold $\mu$ throughout the contract’s life. When the termination threshold is sufficiently high ($R(\mu) \geq \frac{1}{r\xi}$), these contracts set the agent’s continuation value equal to:

$$w(\hat{\mu}_{nt}, \mu) = R(\mu) + \lambda \left( \frac{1}{v} + \frac{1}{r} \right) (\hat{\mu}_{nt} - \mu),$$

which is the minimum required to prevent the CEO from voluntarily leaving the firm. When the termination threshold is lower ($R(\mu) < \frac{1}{r\xi}$) or, equivalently, the minimum payments $c$ are sufficiently high, the agent’s continuation value exceeds this amount and the firm must force the CEO to leave at the termination threshold.

**Proposition 2.** The following hold for contracts with a fixed termination threshold $\mu$:

1. If $w_0 < w(0, \mu)$, then the agent’s participation constraint does not bind.

2. There exists an incentive compatible contract with threshold $\mu$ with payments:

$$c_t = \max\{c, rR(\mu)\} + \lambda \frac{\mu + \epsilon}{\nu} (\hat{\mu}_{nt} - \mu) \geq 0,$$

that yields the following continuation value:

$$w_f(\hat{\mu}_{nt}, \mu) = w(\hat{\mu}_{nt}, \mu) + \left( 1 - e^{-\sqrt{2\pi}(\hat{\mu}_{nt} - \mu)/\sigma_n} \right) \max\{0, \frac{1}{r\xi} - R(\mu)\}.$$
The fastest possible payments include a lump-sum of \( \max\{0, w_0 - w_f(\mu)\} \) at contract initiation to satisfy the CEO’s participation constraint. The value of firm-specific cash flows to the principal is:

\[
b_n(\mu_{nt}) = \frac{1}{r} \mu_{nt} + e^{-\sqrt{\pi} (\mu_{nt} - \mu) / \sigma_{\mu}} \left( \frac{1}{r} - R_{\mu} \right) \left( \mu_o - \mu \right) - \max\{w_f(0, \mu), w_0\} - w_f(\mu_{nt}, \mu).
\]  

(28)

and at the initiation of each CEO:

\[
b_n(0) = \frac{e^{\sqrt{\pi} \mu / \sigma_{\mu}}}{1 - e^{\sqrt{\pi} \mu / \sigma_{\mu}}} \left( \frac{1}{r} - R_{\mu} \right) \left( \mu_o - \mu \right) - \frac{1}{1 - e^{\sqrt{\pi} \mu / \sigma_{\mu}}} \max\{w_f(0, \mu), w_0\}. 
\]  

(29)

3. The incentive compatibility constraint is everywhere binding if and only if \( R(\mu) \geq \frac{1}{r} \). In this case, the unique contract has payments given by (26).

When the condition \( R(\mu) \geq \frac{1}{r} \) holds, the CEO expects weakly more from termination than from continuing to work and receiving the minimum pay. In such cases, the CEO will voluntarily leave. When \( R(\mu) < \frac{1}{r} \), there is forced termination at \( \mu \).

The periodic compensation is:

\[
E_t \left[ \int_t^{t+1} c_s ds \right] = \max\{c, rR(\mu)\} + \frac{\lambda_{t+1}}{\nu} \left( \int_t^{t+1} \hat{\mu}_{nt} ds - \mu \right),
\]

(30)

\[
f^t_{t+1} c_s ds = \max\{c, rR(\mu)\} + \frac{\lambda_{t+1}}{\nu} \left( \hat{\mu}_{nt} - \mu \right),
\]

Earnings impacts the compensation via \( \nu = \frac{\sigma_{\mu}}{\sigma_s} \sqrt{\frac{1 - \rho_{sx}}{\rho_{sx}}} \). The more correlated earnings is with fundamentals (higher \( \rho_{sx} \)), the lower is \( \nu \) and the greater is the incentive pay.

Vice versa for the correlation with cash flows.

The density of the time-to termination \( \tau \) is:\(^3\)

\[
f \left( s, \frac{\hat{\mu}_{nt} - \mu}{\sigma_{\mu}} \right) = \frac{\hat{\mu}_{nt} - \mu}{\sigma_{\mu}} e^{-\sqrt{2\pi} (\hat{\mu}_{nt} - \mu)/\sigma_{\mu}} \left( \frac{\hat{\mu}_{nt} - \mu}{\sigma_{\mu}} \right)^2 / 2(s-t),
\]

(31)

\(^3\)This is a somewhat standard result that one can derive applying a reverse Laplace transform to \( E_t \left[ e^{-r(\tau-t)} \right] = e^{-\sqrt{2\pi} (\hat{\mu}_{nt} - \mu)/\sigma_{\mu}} \).
which implies that the probability of termination by some time \( T > t \) in the future is the following where \( \Phi(\cdot) \) denotes the standard normal distribution:

\[
P_t(T \leq T) = 2\Phi\left(-\frac{\mu_{nt} - \mu}{\sigma_n \sqrt{T-t}}\right).
\]

Sufficient conditions for market value \( b(\hat{\mu}_{0t}, \hat{\mu}_{nt}) \) to be increasing in beliefs \( \hat{\mu}_{nt} \) are that (i) the firm find it profitable to hire a CEO \( (b_n(0) > 0) \), and (ii) the CEO’s benefit from the bad action is sufficiently low relative to the speed of learning \( (\lambda < \frac{\nu}{\nu + \tau}) \).

Beliefs \( \hat{\mu}_{nt} \) as a function of market value equals the following expression, which we derive in the appendix:

\[
\hat{\mu}_{nt} = \bar{\mu} + m_b + \frac{\sigma_b}{\sqrt{2\tau}} \omega\left(-e^{-\sqrt{2\tau}m_b/\sigma_n}m_0\right),
\]

where \( \omega \) denotes Lambert’s \( W \) function and:

\[
\begin{align*}
m_0 &= \frac{\sqrt{2\tau/\sigma_n}}{1+\frac{\nu}{\tau}} \left( \frac{1}{t} - R_n(\bar{\mu}) \right) \left( \frac{1}{t} - \max\{w_f(0,\mu),w_0\} \right) + \max\{0, \frac{1}{t} \lambda - R(\bar{\mu}) \}, \\
m_b &= \frac{\frac{b+\max\{1/CR(\mu)\} - \frac{1}{t}(\hat{\mu}_{nt} - \bar{\mu})}{1+\frac{\nu}{\tau}} + \max\{0, \frac{1}{t} \lambda - R(\bar{\mu}) \}}{1 - \lambda \left( \frac{1}{\nu + \tau} \right)}.
\end{align*}
\]

3 Generating discrete data

The dynamics of the industry-level processes imply that we can express discrete changes as follows for a time increment of size \( \delta \):

\[
\begin{align*}
\mu_{0t} &= \mu_{0,t-\delta} + \delta\mu_{0t}, \\
x_{0t} &= x_{0,t-\delta} + \delta\mu_{0,t-\delta} + \delta x_{0t}, \\
e_{0t} &= \delta \left( 1 - \frac{1-e^{-\theta\delta}}{\theta\delta} \right) \mu_{0,t-\delta} + (1-e^{-\theta\delta}) x_{0,t-\delta} + e^{-\theta\delta} e_{0,t-\delta} + \delta e_{0t}, \\
\hat{\mu}_{0t} &= (1-e^{-\nu\delta}) \mu_{0,t-\delta} + e^{-\nu\delta} \hat{\mu}_{0,t-\delta} + \delta \hat{\mu}_{0t}.
\end{align*}
\]

\[\text{DeMarzo and Sannikov (2017) impose the weaker restriction } \lambda < 1.\]
where:

\[ \delta_{0t} = \beta \sigma_{\mu} \int_{t-\delta}^{t} dz_{0\mu s}, \]
\[ \delta_{0xt} = \beta \int_{t-\delta}^{t} ((t - s) \sigma_{\mu} dz_{0\mu s} + \sigma_x dz_{0xs}), \]
\[ \delta_{0et} = \delta_{0xt} - \frac{1}{\theta} \delta_{0t} + \beta \sigma_{\eta} \int_{t-\delta}^{t} e^{-\theta(t-s)} dz_{0\eta s}; \]  
\[ \delta_{0\mu t} = \beta \left( \int_{t-\delta}^{t} (1 - e^{-\nu(t-s)}) \sigma_{\mu} dz_{0\mu s} \right. \]
\[ \left. + \int_{t-\delta}^{t} e^{-\nu(t-s)} \left( \nu \sigma_x dz_{0xs} + \left( \rho_{\mu x} \sigma_{\mu} - \nu \rho_{xx} \sigma_x \right) dz_{0\eta s} \right) \right), \]

and the accrual shocks are:

\[ d_{0nt} = \frac{1}{\sigma_{\eta}} \left( \sigma_x dz_{0xt} - \sigma_x dz_{0\eta t} + \frac{1}{\theta} \sigma_{\mu} dz_{0\mu t} \right), \]
\[ \sigma_{\eta} = \sqrt{\sigma_e^2 + \sigma_x^2 + \frac{1}{\theta} \sigma_{\mu}^2 + 2 \left( \frac{1}{\theta} (\rho_{\mu x} \sigma_{\mu} - \rho_{xx} \sigma_x) \sigma_{\mu} - \rho_{xx} \sigma_x \sigma_e \right)}. \]  

When firm-specific profitability \( \hat{\mu}_{nt} \) hits the threshold \( \mu_n < 0 \), the CEO is replaced, resulting in a new draw of \( \tilde{\mu}_{nt} \) from a normal distribution with mean 0 and variance \( \tilde{\gamma}_n \). The firm-level processes imply the following:

\[ \mu_{nt} = 1_{\hat{\mu}_{nt} > \mu_n} \mu_{nt-\delta} + 1_{\hat{\mu}_{nt} \leq \mu_n} \tilde{\mu}_{nt-\delta} + \delta_{n\mu t}, \]
\[ x_{nt} = x_{nt-\delta} + \delta \left( \mu_{nt-\delta} + 1_{\hat{\mu}_{nt-\delta} > \mu_n} \mu_{nt-\delta} + 1_{\hat{\mu}_{nt-\delta} \leq \mu_n} \tilde{\mu}_{nt-\delta} \right) + \delta_{0xt} + \delta_{ntx}, \]
\[ e_{nt} = \delta \left( 1 - \frac{1 - e^{-\theta \delta}}{\theta} \right) \left( \mu_{nt-\delta} + 1_{\hat{\mu}_{nt-\delta} > \mu_n} \mu_{nt-\delta} + 1_{\hat{\mu}_{nt-\delta} \leq \mu_n} \tilde{\mu}_{nt-\delta} \right) \]
\[ + (1 - e^{-\theta \delta}) x_{nt-\delta} + e^{-\theta \delta} e_{nt-\delta} + \delta_{0et} + \delta_{net}, \]
\[ \hat{\mu}_{nt} = (1 - e^{-\nu \delta}) \left( 1_{\hat{\mu}_{nt-\delta} > \mu_n} \mu_{nt-\delta} + 1_{\hat{\mu}_{nt-\delta} \leq \mu_n} \tilde{\mu}_{nt-\delta} \right) \]
\[ + e^{-\nu \delta} 1_{\hat{\mu}_{nt-\delta} > \mu_n} \tilde{\mu}_{nt-\delta} + \delta_{n\hat{\mu} t}, \]
\[ \tilde{\mu}_{nt} = \delta_{\tilde{\mu} \mu t}, \]

where:
\[
\delta_{n\mu t} = \sigma_{\mu} \int_{t-\delta}^{t} d\zeta_{n\mu s}, \\
\delta_{nx t} = \int_{t-\delta}^{t} ((t-s) \sigma_{\mu} d\zeta_{n\mu s} + \sigma_{x} d\zeta_{nxs}), \\
\delta_{net} = \delta_{nx t} - \frac{\delta}{\sigma_{\mu}} + \sigma_{\eta} \int_{t-\delta}^{t} e^{-\theta(t-s)} d\zeta_{n\mu s}, \\
\delta_{n\mu t} = \int_{t-\delta}^{t} (1 - e^{-\nu(t-s)}) \sigma_{\mu} d\zeta_{n\mu s} \\
+ \int_{t-\delta}^{t} e^{-\nu(t-s)} \left( \nu \sigma_{x} d\zeta_{nxs} + \left( \rho_{\mux} \sigma_{\mu} - \nu \rho_{xx} \sigma_{x} \right) d\zeta_{nes} \right).
\] (39)

We can write the system of processes as follows where \( y'_{nt} = \{\mu_{nt}, x_{nt}, \epsilon_{nt}, \bar{\mu}_{nt}, \bar{\mu}_{nt}\} \), \( y'_t = \{y'_{1t}, y'_{2t}, \ldots, y'_{Nt}\} \), \( \delta'_{nt} = \{\delta_{n\mu t}, \delta_{nx t}, \delta_{net}, \delta_{n\mu t}, \delta_{n\mu t}\} \), \( \delta'_{t} = \{\delta'_{1t}, \delta'_{2t}, \ldots, \delta'_{Nt}\} \), and \( \delta'_{n\mu t} = \{\delta_{1n\mu t}, \delta_{2n\mu t}, \ldots, \delta_{Nn\mu t}\} \):

\[
\begin{pmatrix}
\delta_{0t} \\
\delta_{t}
\end{pmatrix} = 
\begin{pmatrix}
A_{00} & 1_N \otimes 0_{1 \times 5} \\
1_N \otimes A_{N0} & I_N \otimes A_{NN}
\end{pmatrix} 
\begin{pmatrix}
y_{0t} \\
y_{t}
\end{pmatrix} \\
+ 
\begin{pmatrix}
I_4 \\
1_N \otimes D_{0N} & 1_N \otimes I_5
\end{pmatrix} 
\begin{pmatrix}
\delta_{0t} \\
\delta_{t}
\end{pmatrix},
\] (40)

where:

\[
A_{00} = \begin{pmatrix}
\frac{1}{\delta} & 0 & 0 & 0 \\
0 & 1 - \frac{1 - e^{-\delta \delta}}{\delta} & 0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix},
\]

\[
A_{N0} = \begin{pmatrix}
\frac{1}{\delta} & 0 & 0 & 0 \\
0 & 1 - \frac{1 - e^{-\delta \delta}}{\delta} & 0 & 0 \\
0 & 0 & 0 & 1 \\
1_{\rho_{n.t.d} < \mu_n} & 0 & 0 & 0
\end{pmatrix},
\]

\[
A_{NN} = \begin{pmatrix}
1_{\rho_{n.t.d} < \mu_n} & 0 & 0 & 0 \\
0 & 1 - \frac{1 - e^{-\delta \delta}}{\delta} & 0 & 0 \\
0 & 0 & 1_{\rho_{n.t.d} < \mu_n} & 0
\end{pmatrix} - \begin{pmatrix}
0 & 0 & 0 & 1_{\rho_{n.t.d} \leq \mu_n} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\]

\[
D_{0N} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\] (41)

To generate the shock vector \( \{\delta_{0t}, \delta'_{t}\} \) from independent normal random variables, it can be written as:

\[
\begin{pmatrix}
\delta_{0t} \\
\delta_{t}
\end{pmatrix} = C_{\delta} b_t,
\] (42)

where \( b_t \) is a \( 4(N + 1) \) vector of standard normal random variables and \( C_{\delta} \) is from
the Cholesky decomposition of the covariance matrix of the shock vector:

\[ C_\delta C'_\delta = E \left[ \delta_{n'} \delta_n' \right] = \left( \begin{array}{cc} \beta^2 \Sigma_{nn} & 1_{N_\mathcal{N}} \otimes 0_{4 	imes 5} \\ 1_{N_\mathcal{N}} \otimes 0_{5 \times 4} & \Sigma_{nn} \otimes 0_4 \otimes 0_4 \end{array} \right), \]

\[ \Sigma_{nn} = E \left[ \delta_{nt} \delta'_{nt} \right] = \left( \begin{array}{cccc} \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} \\ \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} \\ \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} \\ \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} & \sigma_{\mu \mu} \end{array} \right). \] (43)

The elements of \( \Sigma_{nn} \) are:

\[
\begin{align*}
\sigma_{\mu \mu} &= \delta \sigma_{\mu \mu}, \\
\sigma_{xx} &= \delta \left( \sigma_x^2 + \delta \rho_{xx} \sigma_x + \frac{\delta}{3} \sigma_{\mu \mu} \right), \\
\sigma_{ee} &= \sigma_{xx} + \frac{1}{\delta} \sigma_{\mu \mu} - 2 \frac{1}{\delta} \sigma_{\mu x} \\
&\quad + \delta \left( \frac{1-e^{-\delta \delta}}{2 \delta} \sigma_{\eta}^2 + 2 \left( \frac{1-e^{-\delta \delta}}{\delta} \rho_{x \eta} \sigma_x - \frac{e^{-\delta \delta}}{\delta} \rho_{\mu \eta} \sigma_\mu \right) \sigma_\eta \right), \\
\sigma_{\mu \mu} &= \sigma_{\mu \mu} \left( 1 - \frac{(1-e^{-\delta \delta})^2}{\delta} \right) \left( 1 - \rho_{\mu \mu} \right), \\
\sigma_{\mu x} &= \delta \left( \frac{1}{2} \sigma_{\mu \mu} + \rho_{\mu x} \sigma_x \right), \\
\sigma_{\mu e} &= \sigma_{xx} - \frac{1}{\delta} \sigma_{\mu x} + \delta \left( \frac{1}{\delta} \left( \frac{1-e^{-\delta \delta}}{\delta} - e^{-\delta \delta} \right) \rho_{x \eta} \sigma_x + \frac{1-e^{-\delta \delta}}{\delta} \rho_{x \eta} \sigma_x \sigma_\eta \right), \\
\sigma_{\mu \mu} &= \sigma_{\mu \mu} \left( 1 - \frac{(1-e^{-\delta \delta})^2}{\delta} \right) \left( 1 - \rho_{\mu \mu} \right), \\
\sigma_{xx} &= \sigma_{xx} - \frac{1}{\delta} \sigma_{\mu x} + \delta \left( \frac{1}{\delta} \left( \frac{1-e^{-\delta \delta}}{\delta} - e^{-\delta \delta} \right) \rho_{x \eta} \sigma_x + \frac{1-e^{-\delta \delta}}{\delta} \rho_{x \eta} \sigma_x \sigma_\eta \right) + \delta \left( \frac{1}{2} \sigma_{\mu \mu} + \rho_{\mu x} \sigma_x \right), \\
\sigma_{\mu e} &= \sigma_{\mu \mu} - \frac{1}{\delta} \sigma_{\mu \mu} + \delta \left( \frac{1-e^{-\delta \delta}}{\delta} \rho_{\mu \mu} \sigma_x + \frac{1-e^{-\delta \delta}}{(\delta \delta)^2} \left( \rho_{\mu \mu} - \rho_{\mu \mu} \right) \right) \rho_{\mu \mu} \sigma_\eta,
\end{align*}
\]

where:

\[
\begin{align*}
\rho_{\mu \eta} &= \frac{1}{\delta} \left( \rho_{x \eta} \sigma_\mu - \rho_{\mu x} \sigma_x + \frac{1}{\delta} \rho_{\mu \mu} \sigma_\mu \right), \\
\rho_{xx \eta} &= \frac{1}{\delta} \left( \rho_{xx \eta} \sigma_x - \sigma_x + \frac{1}{\delta} \rho_{x \mu} \sigma_\mu \right), \\
\rho_{ee \eta} &= \frac{1}{\delta} \left( \sigma_e - \sigma_x \sigma_\mu \sigma_x + \frac{1}{\delta} \rho_{\mu e} \sigma_\mu \right), \\
\rho_{\mu \mu} &= \rho_{\mu \mu}^2 + \left( \rho_{\mu \mu} - \rho_{\mu \mu} \rho_{\mu \mu} \right) \sqrt{\frac{1-\rho_{\mu \mu}^2}{1-\rho_{xx \mu}^2}}, \\
\rho_{\mu x} &= \sqrt{(1-\rho_{\mu \mu}^2)(1-\rho_{xx \mu}^2)} + \rho_{\mu \mu} \rho_{xx \mu}, \\
\rho_{\mu \eta} &= \frac{1}{\delta} \left( \rho_{x \eta} \sigma_e - \rho_{\mu x} \sigma_x + \frac{1}{\delta} \rho_{\mu \mu} \sigma_e \right).
\end{align*}
\]
References


Appendix

The filtering problem (Section 1.2)

The filtering problem uses the Kalman-Bucy filter (Liptser and Shiryaev 2001, Theorem 10.3). To set up the dynamics for the filtering problem, first note that the ability to directly observe industry-level cash flows and earnings eliminates the need for a given firm \( n \) to utilize the observables from other firms. The filtering problem then reduces to using the industry-level and firm-specific cash flows and earnings to infer industry-level and firm-specific profitability. Denote the vector of profitabilities by \( \mu'_t = \{\mu_{0t}, \mu_{nt}\} \) and denote the vector of observables by \( y'_t = \{x_{0t}, e_{0t}, x_{nt}, e_{nt}\} \).

Denote the vector of agent actions by \( a'_t = \{0, 0, a_{1t}, 0, a_{2t}, 0, \ldots, a_{Nt}, 0\} \). Denote the vector of shocks to profitability by \( z'_{\mu t} = \{z_{0\mu t}, z_{n\mu t}\} \) and the shocks to observables by \( z'_{yt} = \{z_{0xt}, z_{0et}, z_{nxt}, z_{net}\} \). We can express the shocks in terms of the vector \( z_t \) of six independent Brownian motions as:

\[
\begin{pmatrix}
    dz_{\mu t} \\
    dz_{yt}
\end{pmatrix} = C_z dz_t. \tag{A1}
\]

The matrix \( C_z \) is from the Cholesky decomposition of the covariance matrix of \( dz_{\mu t} \) and \( dz_{yt} \):

\[
C_z = \frac{1}{dt} E \left[ \begin{pmatrix}
    dz_{\mu t} \\
    dz_{yt}
\end{pmatrix} \begin{pmatrix}
    dz'_{\mu t} \\
    dz'_{yt}
\end{pmatrix}^T \right] = \begin{pmatrix}
    I_2 & I_2 \otimes r_{\mu y} \\
    I_2 \otimes r_{\mu y} & I_2 \otimes R_{xe}
\end{pmatrix}, \tag{A2}
\]

where \( r_{\mu y} = \{\rho_{\mu x}, \rho_{\mu e}\} \) and:

\[
R_{xe} = \begin{pmatrix}
    1 & \rho_{xe} \\
    \rho_{xe} & 1
\end{pmatrix}. \tag{A3}
\]

The restriction that the covariance matrix is positive semi-definite requires that:

\[
\rho_{\mu x}, \rho_{\mu e}, \rho_{xe} \in [-1, 1], \quad \rho_{xe} \in \rho_{\mu e} \rho_{\mu x} \pm \sqrt{(1 - \rho_{\mu e}^2)(1 - \rho_{\mu x}^2)} \subseteq [-1, 1].
\]
When the agent’s action $a_{nt} = 0$, we can then express the dynamics of the unobservable profitability and observable cash flows and earnings as:

\[
\begin{align*}
d\mu_t &= (\begin{pmatrix} \Sigma_\mu & 0_{2\times4} \end{pmatrix} C_z) d\tilde{z}_t \\
dy_t &= ((A_{\mu} \otimes (\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})) \mu_t dt - a_t dt + (I_2 \otimes (\begin{pmatrix} 0 & 0 \\ 0 & -\theta \end{pmatrix})) y_t dt + (0_{4\times2} \otimes \Sigma_{xx}) C_z d\tilde{z}_t,
\end{align*}
\]

where:

\[
C_y = (\begin{pmatrix} 0 & 1 \\ \beta & 1 \end{pmatrix}), \quad \Sigma_{xx} = (\begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_e \end{pmatrix}).
\]

The filtering problem gives profitability estimates $\hat{\mu}_t = \{\hat{\mu}_0, \hat{\mu}_{nt}\}$ with covariance matrix $\hat{\Gamma}_t = \text{var}_t(\mu_t)$ and the following dynamics:

\[
\begin{align*}
d\hat{\mu}_t &= K_t d\tilde{n}_t, \\
d\tilde{n}_t &= dy_t - (A_{\mu} \hat{\mu}_t - a_t dt + A_y y_t) dt, \\
K_t &= \left( B_\mu B'_y + \hat{\Gamma}_t A'_\mu \right) (B_y B'_y)^{-1}, \\
d\hat{\Gamma}_t &= \left( B_\mu B'_\mu - K_t \left( B_y B'_\mu + A_\mu \hat{\Gamma}_t \right) \right) dt.
\end{align*}
\]

In a steady state, $d\hat{\Gamma}_t = 0_{4\times4}$, which implies steady state posteriors of:

\[
\begin{align*}
\hat{\gamma}_n &= \text{var}_\infty(\mu_{nt}) = \sigma_\mu \sigma_x \left( \sqrt{1-\rho_{\mu e}^2} (1-\rho_{xe}^2) - (\rho_{\mu x} - \rho_{\mu e} \rho_{xe}) \right), \\
\hat{\gamma}_0 &= \text{var}_\infty(\mu_0) = \beta^2 \hat{\gamma}_n.
\end{align*}
\]

The corresponding gain matrix is the following where $\nu = \frac{\sigma_\mu}{\sigma_x} \sqrt{\frac{1-\rho_{\mu e}^2}{1-\rho_{xe}^2}}$:

\[
K_\infty = (\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}) \otimes \left( \nu \frac{\rho_{\mu e} \sigma_{\mu} - \nu \rho_{xe} \sigma_x}{\sigma_e} \right),
\]

which gives expressions (10) and (11). The vector of observable Brownian motions

---

5The other solutions to the steady state either imply negative variances or that the posterior covariance $\text{cov}_\infty(\mu_0, \mu_{nt})$ has a nonzero imaginary component.
\{d\tilde{z}_{0t}, d\tilde{z}_{0xt}, d\tilde{z}_{0ct}, d\tilde{z}_{nt}, d\tilde{z}_{nxt}, dz_{net}\} has the correlation matrix:

\[
I_2 \otimes \begin{pmatrix}
1 & \rho_{\mu x} \\
\rho_{\mu x} & 1
\end{pmatrix},
\]  

(A9)

where:

\[
\rho_{\mu x} = \frac{1}{d^t} E [d\tilde{z}_{nxt}d\tilde{z}_{nxt}] = \frac{1}{d^t} E [d\tilde{z}_{nxt}dz_{net}] = \rho_{\mu e} \rho_{xe} + \sqrt{(1 - \rho_{\mu e}^2)(1 - \rho_{xe}^2)}. 
\]  

(A10)

Proof of Proposition 1

Assume a termination threshold \( \mu \) so that \( \tau = \inf \{ t : \hat{\mu}_{nt} \leq \mu \} \). The value function satisfies the following Hamilton-Jacobi-Bellman (HJB) equation away from the threshold \( \mu \):

\[
v_n (\hat{\mu}_{nt}) = \frac{1}{\tau} \hat{\mu}_{nt} + \frac{\sigma_{\mu}}{2\tau} v''_n (\hat{\mu}_{nt}) \Rightarrow v_n (\hat{\mu}_{nt}) = \frac{1}{\tau} \hat{\mu}_{nt} + c_1 e^{\frac{\sqrt{\tau}}{\sigma_{\mu}} \hat{\mu}_{nt}} + c_2 e^{-\frac{\sqrt{\tau}}{\sigma_{\mu}} \hat{\mu}_{nt}}, \]  

(A11)

for some constants \( c_1 \) and \( c_2 \). The boundary condition that the firm does not terminate as manager ability becomes unbounded \( (v_n (\hat{\mu}_{nt}) \to \frac{1}{\tau} \hat{\mu}_{nt} \text{ as } \hat{\mu}_{nt} \to \infty) \) implies that \( c_1 = 0 \). Continuity at the threshold \( \mu \) gives \( c_2 \):

\[
\left\{ \begin{array}{l}
\frac{1}{\tau} \mu + c_2 e^{-\frac{\sqrt{\tau}}{\sigma_{\mu}} \mu} = c_2 + R (\mu) - k \Rightarrow v_n (\hat{\mu}_{nt}) = \frac{1}{\tau} \hat{\mu}_{nt} + e^{-\frac{\sqrt{\tau}}{\sigma_{\mu}} (\hat{\mu}_{nt} - \mu)/(\sqrt{\tau} P_{ft})} (\frac{1}{\tau} - R) \left( \mu_0 - \mu \right). \\
\end{array} \right.
\]  

(A12)

Smooth pasting \( (v'_n (\mu) = R_\mu) \) gives the condition following condition that yields the surplus-maximizing threshold \( \mu_{fb} \):\footnote{This condition is also equivalent to maximizing \( v_n \) with respect to \( \mu \), which is equivalent to maximizing \( c_2 \).}

\[
\frac{\sqrt{\tau}}{\sigma_{\mu}} \left( \mu_0 - \mu_{fb} \right) = 1 - e^{\frac{\sqrt{\tau}}{\sigma_{\mu}} \mu_{fb}/\sigma_{\mu}}. 
\]  

(A13)
The above condition has two solutions – one positive and the other negative, and the cutoff \( \mu_{fb} \) defined in (21) is the unique \( \mu < 0 \) that satisfies the above condition. Because \( c_2 \) is increasing for all \( \mu < \mu_f \), and decreasing for all \( \mu \in (\mu_{fb}, 0) \), this is the unique optimum. Solving the optimality condition for \( \mu_o - \mu_{fb} \) and substituting into \( v_n \) gives expression (22).

To show \( E_t \left[ e^{-r(\tau-t)} \right] = e^{-\sqrt{2}r(\hat{\mu}_{nt}-\mu_{fb})/\sigma_x} \), the process \( E_t \left[ e^{-r(\tau-t)} \right] \) is a martingale, which we conjecture to be of the form \( e^{-rt} f(\hat{\mu}_{nt}) \). Ito’s lemma then gives:

\[
dE_t \left[ e^{-r(\tau-t)} \right] = -re^{-rt} f(\hat{\mu}_{nt}) dt + e^{-rt} f'(\hat{\mu}_{nt}) \sigma_x d\hat{\epsilon}_{nt} + \frac{1}{2} e^{-rt} f''(\hat{\mu}_{nt}) \sigma_x^2 dt.
\]

Because \( E_t \left[ e^{-r(\tau-t)} \right] \) is a martingale, it has zero drift so that the portion of it that depends on \( \hat{\mu}_{nt} \) solves an equation similar to (A11). The boundary conditions \( \lim_{\hat{\mu}_{nt} \to -\infty} E_t \left[ e^{-r(\tau-t)} \right] = 0 \) and \( \lim_{\hat{\mu}_{nt} \to \mu_{fb}} E_t \left[ e^{-r(\tau-t)} \right] = 1 \) give \( E_t \left[ e^{-r(\tau-t)} \right] = e^{-\sqrt{2}r(\hat{\mu}_{nt}-\mu_{fb})/\sigma_x} \).

If the firm faces only the participation constraint, then the agent payoff can be structured so that the CEO’s expected value exactly meets the reservation value \( w_0 \).

This gives the following payoff to the principal at each contracting date \( \tau_i \):

\[
v_n(0) = E_t \left[ \int_0^{\tau_i} e^{-r(s-t)} \mu_{nt} ds + e^{-r(\tau-t)} (v_n(0) + R(\hat{\mu}_{nt}) - k) \right] = E_t \left[ e^{-r(\tau-t)} \left( v_n(0) + (\frac{1}{r} - R) \left( \mu_o - \hat{\mu}_{nt} \right) \right) \right] \quad (A15)
\]

\[
b_{\tau_i}(0, w_0) = E_{\tau_i} \left[ \int_{\tau_i}^{\tau_{i+1}} e^{-r(s-\tau_i)} (\mu_{ns} - c_s) ds + e^{-r(\tau_{i+1}-\tau)} (b_{\tau_{i+1}}(0, w_0) - k) \right] = E_{\tau_i} \left[ e^{-r(\tau_{i+1}-\tau)} \left( b_{\tau_{i+1}}(0, w_0) + (\frac{1}{r} - R) \left( \mu_o - \hat{\mu}_{nt+1} \right) \right) \right] - w_0
\]

\[
= v_n(0) + E_{\tau_i} \left[ e^{-r(\tau_{i+1}-\tau_i)} \left( b_{\tau_{i+1}}(0, w_0) - v_n(0) \right) \right] - w_0
\]

\[
= e^{\sqrt{2}r \mu_o \sigma_x} \left( b_{\tau_{i+1}}(0, w_0) + (\frac{1}{r} - R) \left( \mu_o - \mu \right) \right) - w_0.
\]

(A16)

The first-order condition implies the optimal threshold \( \mu_{ss}^* = \mu_o - \frac{\sigma_x}{\sqrt{2r}} \frac{b_{\tau_{i+1}}(0, w_0)}{\frac{1}{r} - R} \), and
the second-order condition is satisfied at any \( \mu \) that solves the first-order condition. Substituting back into the objective function gives:

\[
\begin{align*}
  b_{r_i}(0, w_0) &= \frac{\sigma_{\mu}}{\sqrt{2r}} e^{\sqrt{2r} \mu / \sigma_{\mu}} \left( \frac{1}{r} - R_{\mu} \right) - w_0, \\
  &\quad (A17)
\end{align*}
\]

Because \( b_{r_{i+1}}(0, w_0) = b_{r_i}(0, w_0) \), the threshold \( \mu_{**} \) satisfies:

\[
\frac{\sqrt{2r}}{\sigma_{\mu}} \left( \mu_o - \mu_{**} - \frac{1}{r - R_{\mu}} w_0 \right) = 1 - e^{\sqrt{2r} \mu_{**} / \sigma_{\mu}}, \\
  &\quad (A18)
\]

which has a similar solution to the surplus-maximizing threshold \( \mu_{fb} \) with \( \mu_o - \frac{1}{r - R_{\mu}} w_0 \) appearing in the equality rather than \( \mu_o \). Both \( \mu_{fb} \) and \( \mu_{**} \) have the form \( \frac{\sigma_{\mu}}{\sqrt{2r}} m(x) \) where \( m(x) = x - \omega (-e^x) \) and \( x = \frac{\sqrt{2r}}{\sigma_{\mu}} \left( \mu_o - \frac{\sigma_{\mu}}{\sqrt{2r}} \right) \) for \( \mu_{fb} \) and \( x = \frac{\sqrt{2r}}{\sigma_{\mu}} \left( \mu_o - \frac{1}{r - R_{\mu}} w_0 - \frac{\sigma_{\mu}}{\sqrt{2r}} \right) \) for \( \mu_{**} \). The function \( \omega(x) > -1 \), which implies that \( m \) is increasing so that \( \mu_{**} < \mu_{fb} \).

To obtain the function for intermediate times, set compensation equal to \( rR(\mu_{**}) \) with initial lump sum of \( w_0 - R(\mu_{**}) \). The CEO expects:

\[
  w_0 - R(\mu_{**}) + E \left[ \int_t^\tau e^{-r(s-t)} c_s ds + e^{-r(\tau-t)} R(\hat{\mu}_{n\tau}) \right] = w_0 \\
  &\quad (A19)
\]

More generally, if the principal sets the agent’s compensation equal to \( rw_0 \) with a lump sum of \( w_0 - R(\mu) \) at termination, then the principal’s payoff satisfies the HJB equation, and the payoff is maximized at \( \mu = \mu_{**} \). This contract gives (24), which can be derived from direct computations using substitutions from (A18).

**Supporting results for Section 2**

Lemmas 3 and 4 are similar to results in DeMarzo and Sannikov (2017), and we show that they hold in our setting.

**Lemma 3.** In an optimal, incentive-compatible contract, \( a_t = 0 \) and it is unneces-
sary to include the possibility of early termination. Specifically, for any incentive-compatible contract \((c_t, \tau)\) with actions \(a_t\), there exists another incentive-compatible contract \((\tilde{c}_t, \tilde{\tau})\) with actions \(\tilde{a}_t = 0\) that gives the same payoff to the agent and a weakly higher payoff to the principal.

**Proof.** Denote by \(\mathcal{F}_t\) the filtration generated by the observed cash flow and earnings processes \((x_t, c_t)\). Now take an adjusted process \(\tilde{x}_t = x_t - \int_0^t a_s ds\) and denote the filtration \(\tilde{\mathcal{F}}_t\) as the one generated by \((\tilde{x}_t, c_t)\). The original contract has payoff \(c_t(\mathcal{F}_t)\) and stopping time \(\tau(\mathcal{F}_t)\). Now consider a new contract:

\[
\tilde{c}_t = c_t(\tilde{\mathcal{F}}_t) + \lambda a_t, \quad \tilde{\tau} = \tau(\tilde{\mathcal{F}}_t). \tag{A20}
\]

Denote by \(\tilde{a}_t\) the diversion under the new contract. Then \(d\tilde{x}_t = dx_t - a_t dt = (\mu_t - a_t - \tilde{a}_t) dt + \sigma x d\tilde{z}_t\) so that \(c_t(\tilde{\mathcal{F}}_t)\) equals the payoff flow that the agent would obtain by taking action \(a_t + \tilde{a}_t\) under the original contract, and the extra payoff \(\lambda a_t\) under the new contract is the same as the agent’s diversion payoff from taking action \(a_t\) under the original contract. Also, the stopping time \(\tilde{\tau}\) under action \(a_t\) is the same as taking action \(a_t + \tilde{a}_t\) under the original contract. Condition (19c) then implies that \(\tilde{a}_t = 0\).

The the firm-specific portion of the principal’s payoff under the new contract is:

\[
\begin{align*}
E_t \left[ \int_t^{\tilde{\tau}} e^{- r (s-t)} (\hat{\mu}_{ns} - \tilde{a}_s - \tilde{c}_s) \, ds + e^{- r (\tilde{\tau}-t)} (b_n (0, w_0) - k) \right] \\
&= E_t \left[ \int_t^{\tau} e^{- r (s-t)} (\hat{\mu}_{ns} - \lambda a_s - c_s) \, ds + e^{- r (\tau-t)} (b_n (0, w_0) - k) \right] \\
&\geq E_t \left[ \int_t^{\tau} e^{- r (s-t)} (\hat{\mu}_{ns} - a_s - c_s) \, ds + e^{- r (\tau-t)} (b_n (0, w_0) - k) \right] \tag{A21}
\end{align*}
\]

so that the principal is better off. This shows that \(a_t = 0\) in the optimal contract.

To show that the agent will not liquidate before the contract time \(\tau\), the contract
action \( a_t = 0 \) and the restriction that any action \( \hat{a}_t \geq 0 \) imply that any belief divergence \( \hat{\mu}^a_t - \hat{\mu}_t \) must be non-negative. An agent can take a strategy \( \hat{\alpha}_s = a_s + \hat{\mu}^a_t - \hat{\mu}_t \) for all \( s > t \) and (14) implies that the principal’s belief divergence will remain at \( \hat{\mu}^a_t - \hat{\mu}_t \). This gives the agent the payoff:

\[
E_t \left[ \int_t^\tau e^{-r(s-t)}(\lambda \hat{\alpha}_s + c_s)ds + e^{-r(\tau-t)} R(\hat{\mu}_t, \hat{\mu}^a_t) \right] = w_t + E_t \left[ \frac{\lambda}{r} \hat{\mu}^a_t \right] (\hat{\mu}^a_t - \hat{\mu}_t) \\
\geq R(\hat{\mu}_t) + \frac{\lambda}{r} (\hat{\mu}^a_t - \hat{\mu}_t) = \tilde{R}(\hat{\mu}^a_t, \hat{\mu}_t) \quad (A22)
\]

so that the agent is better off remaining with the firm.

**Lemma 4.** The following hold for any incentive-compatible contract \((c_t, \tau)\):

1. The agent’s continuation value \( w_t \) has the representation:

\[
dw_t = (rw_t - c_t) dt + \beta_{mt} d\hat{\mu}_{nt} + \beta_{et} d\tilde{z}_{net}, \quad (A23)
\]

where \( d\tilde{z}_{net} = (dz_{net} - \rho_{\mu e} d\hat{z}_{net})/\sqrt{1 - \rho^2_{\mu e}} \) is the portion of the earnings shock that is orthogonal to shocks to beliefs.

2. The information rent has the representation:

\[
\xi_t = E_{P_t} \left[ \int_t^\tau e^{-(r+\nu)(s-t)} \nu \left( \beta_{\mu s} - \beta_{\hat{\mu}t} \frac{1}{\sigma_{\mu e}^2} \rho_{\mu e} \sqrt{1 - \rho^2_{\mu e}} \right) ds + e^{-(r+\nu)(\tau-t)} \frac{\lambda}{r} \right], \quad (A24)
\]

with the following dynamics for some process \((\chi_{mt}, \chi_{et})\):

\[
d\xi_t = \left( (r + \nu)\xi_t - \nu \left( \beta_{\mu t} - \beta_{\hat{\mu}t} \frac{1}{\sigma_{\mu e}^2} \rho_{\mu e} \sqrt{1 - \rho^2_{\mu e}} \right) \right) dt + \chi_{mt} d\hat{\mu}_{nt} + \chi_{et} d\tilde{z}_{net}. \quad (A25)
\]

3. The IC constraint \( \beta_{\mu t} \geq \frac{1}{\nu} \lambda + \xi_t + \frac{1}{\sigma_{\mu}^2} \beta_{\hat{\mu}t} \frac{\rho_{\mu e}}{\sqrt{1 - \rho^2_{\mu e}}} \) is necessary for the optimality of \( a_t = 0 \).

4. The following lower bound for information rent holds with equality if IC con-
strait is binding at all future dates:

$$\xi_t \geq \frac{\lambda}{\tau}. \quad \text{(A26)}$$

**Proof.** We can write the dynamics of the observed firm-level cash flows and earnings under the equilibrium strategy $a_t = 0$ as follows:

$$
\frac{dx_{nt}}{d\epsilon_{nt}} = \left( \frac{\dot{\mu}_{nt} + \dot{\mu}_{nt}}{\theta(x_{nt} - e_{nt})} \right) dt + \left( \frac{\beta \sigma_x d\tilde{z}_{nt}}{\beta \sigma_e d\tilde{z}_{net}} \right) + C \frac{d\tilde{z}_{nt}}{d\tilde{z}_{nt}} \quad \text{(A27)}
$$

where:

$$
C = \begin{pmatrix}
\rho_{\mu_x \sigma_x} & \rho_{\mu_x \sigma_e} \\
\rho_{\mu_e \sigma_e} & \sqrt{1 - \rho_{\mu_e \sigma_e}^2}
\end{pmatrix}
\begin{pmatrix}
\rho_{\mu_x \sigma_x} \\
\rho_{\mu_e \sigma_e}
\end{pmatrix}
\sqrt{1 - \rho_{\mu_e \sigma_e}^2}.
$$

The vector $\{d\tilde{z}_{nt}, d\tilde{z}_{net}\}$ is a standard, two-dimensional Brownian motion with respect to the beliefs $P$ generated by the equilibrium actions $a_t = 0$. Given a deviation $\hat{a}_t \neq 0$, the dynamics follow:

$$
\frac{dx_{nt}}{d\epsilon_{nt}} = \left( \frac{\mu_{nt} + \mu_{nt} - \hat{a}_t}{\theta(x_{nt} - e_{nt})} \right) dt + \left( \frac{\beta \sigma_x d\tilde{z}_{nt}}{\beta \sigma_e d\tilde{z}_{net}} \right) + C d\tilde{z}_{nt} \frac{d\tilde{z}_{nt}}{d\tilde{z}_{nt}} \quad \text{(A29)}
$$

where:

$$
d\tilde{z}_{nt}^a = \begin{pmatrix}
\frac{d\tilde{z}_{nt}^a}{d\tilde{z}_{nt}^a} \\
\frac{d\tilde{z}_{net}^a}{d\tilde{z}_{net}^a}
\end{pmatrix} = \begin{pmatrix}
\frac{d\tilde{z}_{nt} + \frac{1}{\sigma_p} \nu (\hat{a}_t - (\hat{\mu}_{nt} - \mu_{nt})) dt}{\rho_{\mu_e} \sqrt{1 - \rho_{\mu_e}^2 (d\tilde{z}_{net} - \mu_{nt} d\tilde{z}_{net})}} \\
\frac{d\tilde{z}_{net} + \frac{1}{\sigma_p} \nu (\hat{a}_t - \mu_{nt}) dt}{\rho_{\mu_e} \sqrt{1 - \rho_{\mu_e}^2 (d\tilde{z}_{net} - \mu_{nt} d\tilde{z}_{net})}}
\end{pmatrix}.
$$

The vector $\{d\tilde{z}_{nt}^a, d\tilde{z}_{net}^a\}$ is a standard Brownian motion with respect to the beliefs $\hat{P}$ generated by the actions $\hat{a}_t$. We use the term $\alpha_t = \hat{\mu}_{nt} - \tilde{\mu}_{nt}$ to denote the discrepancy between the principal’s and the agent’s beliefs, with $d\alpha_t = \nu (\hat{a}_t - \alpha_t) dt$ from (14).

Any deviation $\hat{a}_t \neq 0$ must result in beliefs that are absolutely continuous with respect to the beliefs generated by $a_t = 0$ because, otherwise, the principal could enact severe punishments for states that have zero probability under $a_t = 0$ (Williams 2011; DeMarzo and Sannikov 2017). We therefore apply Girsanov’s theorem to obtain a relative density process for the change from the principal’s beliefs $P$ to the agent’s
\( \hat{P} \) (e.g., Øksendal 2003, Theorem 8.6.6). We first require a process \( u_t \) that solves:

\[
C u_t =\left( \begin{array}{c} \hat{\mu}_t \\ \theta(x_{t-} - c_t) \end{array} \right) - \left( \begin{array}{c} \hat{\mu}_t \\ \theta(x_{t-} - c_t) \end{array} \right) = - \left( \begin{array}{c} \alpha_t - \hat{a}_t \\ 0 \end{array} \right) \Rightarrow \quad u_t = - C^{-1} \left( \begin{array}{c} \alpha_t - \hat{a}_t \\ 0 \end{array} \right) = \frac{\nu(\alpha_t - \hat{a}_t)}{\sigma} \left( \begin{array}{c} -1 \\ \sqrt{1 - \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha}} \end{array} \right).
\]

(A31)

There is then a process \( \phi_t \) such that \( d\hat{P}_t = \phi_t dP_t \) where:

\[
\begin{align*}
\phi_t &= \exp \left\{ - \int_0^t u_s' dz_{ns} - \frac{1}{2} \int_0^t u_s' u_s' ds \right\}, \quad d\phi_t = - \phi_t u_t' \, dz_{nt}, \\
\end{align*}
\]

and \( E_P[\phi_t] = \phi_0 = 1 \). The agent’s payoff can be written as:

\[
E_P \left[ \int_0^T e^{-rt} (\lambda \hat{a}_t + c_t) \, dt + e^{-rt} \hat{R}(\hat{\mu}_{nt}, \hat{\mu}_{nt}) \right] = E_P \left[ \int_0^T \phi_t e^{-rt} (\lambda \hat{a}_t + c_t) \, dt + \phi_0 e^{-rt} \hat{R}(\hat{\mu}_{nt}, \hat{\mu}_{nt} + \alpha_t) \right].
\]

(A32)

The problem is solved using a stochastic version of Pontryagin’s maximum principle (e.g., Yong and Zhou 1999, Theorem 3.3). Denote the states by \( y_t = (\phi_t, \alpha_t) \), with the following dynamics:

\[
\left( \begin{array}{c} \frac{d\phi_t}{dy_t} \\ \frac{d\phi_t}{d\alpha_t} \end{array} \right) = \left( \begin{array}{c} \frac{0}{b_{yt}} \\ \frac{\nu(\alpha_t - \hat{a}_t)}{\sigma} \left( \begin{array}{c} 1 - \frac{\rho_{\mu}^{\alpha}}{\sqrt{1 - \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha}}} \\ 0 \end{array} \right) \end{array} \right) \, d\Sigma_{yt}.
\]

(A34)

Denote the costate variables on the drifts by \( p_t = (p_{\phi_t}, p_{\alpha_t}) \), and the costate variables on the volatilities by the matrix \( Q_t = \left( \begin{array}{cc} q_{\phi_t} & q_{\alpha_t} \\ q_{\phi_t} & q_{\alpha_t} \end{array} \right) \). The Hamiltonian is then:

\[
H(t, y_t, \hat{a}_t, p_t, q_t) = \phi_t (\lambda \hat{a}_t + c_t) + p_{\alpha_t} \nu(\alpha_t - \hat{a}_t) + \left( q_{\phi_t} - q_{\alpha_t} \frac{\rho_{\mu}^{\alpha}}{\sqrt{1 - \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha}}} \right) \phi_t \frac{\nu(\alpha_t - \hat{a}_t)}{\sigma},
\]

(A35)

where \( p_{\phi_t} \) does not appear because \( \phi_t \) has zero drift and \( q_{\alpha_t} \) does not appear because \( \alpha_t \) has zero volatility. Differentiating with respect to \( \hat{a}_t \) gives:

\[
\frac{\partial H}{\partial \hat{a}_t} = \phi_t \lambda + p_{\alpha_t} \nu - \frac{\nu}{\sigma} \left( q_{\phi_t} - q_{\alpha_t} \frac{\rho_{\mu}^{\alpha}}{\sqrt{1 - \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha} \rho_{\mu}^{\alpha}}} \right) \phi_t.
\]

(A36)
which must be weakly negative in order for \( \dot{a}_t = 0 \) to be optimal, given the restriction \( \dot{a}_t \geq 0 \). The costate variables on the drifts evolve as follows, where the \( r_{\dot{p}jt} \) terms account for discounting, and \( \dot{a}_t = 0 \) for all \( t \) implies the second equality with \( \alpha_t = 0 \), \( \phi_t = 1 \), and \( d\mathbf{z}_{nt} = d\mathbf{z}_{nt}^0 \):

\[
\begin{align*}
\left( \frac{dp_{\dot{p}}}{d\alpha_t} \right) &= \left( \frac{rp_{\dot{p}} - \partial H}{r_{\dot{p}at} - \partial \alpha} \right) dt + \left( \frac{\alpha_{\dot{p}} q_{\dot{p}\alpha t}}{q_{\dot{p}at} q_{\dot{p}\alpha t}} \right) d\mathbf{z}_{nt} \\
&= \left( \frac{r_{\dot{p}at} - \partial H}{r_{\dot{p}at} - \partial \alpha} \right) dt + \left( \frac{r_{\dot{p}at} - \partial H}{r_{\dot{p}at} - \partial \alpha} \right) \alpha_t \nu \frac{\nu}{\sigma_\mu} dt + \left( \frac{\alpha_{\dot{p}} q_{\dot{p}\alpha t}}{q_{\dot{p}at} q_{\dot{p}\alpha t}} \right) d\mathbf{z}_{nt} \\
&= \left( \frac{r_{\dot{p}at} - \partial H}{r_{\dot{p}at} - \partial \alpha} \right) dt + \left( \frac{\alpha_{\dot{p}} q_{\dot{p}\alpha t}}{q_{\dot{p}at} q_{\dot{p}\alpha t}} \right) d\mathbf{z}_{nt} \quad (A37)
\end{align*}
\]

The boundary conditions are:

\[
\begin{align*}
p_{\dot{p}t} &= \frac{\partial}{\partial \phi_t} \dot{R}(\hat{\mu}_{nt}, \hat{\mu}_{nt} + \alpha_t) \bigg|_{\alpha_t = 0} = \dot{R}(\hat{\mu}_{nt}), \\
p_{\alpha t} &= \frac{\partial}{\partial \alpha} \phi_t \dot{R}(\hat{\mu}_{nt}, \hat{\mu}_{nt} + \alpha_t) \bigg|_{\alpha_t = 0} = \frac{\lambda}{\bar{r}}. \quad (A38)
\end{align*}
\]

The process \( p_{\dot{p}t} \) is given by the agent’s expected payoff from continuing and, because \( p_{\alpha t} \) is the change in payoff with respect to the belief discrepancy \( \alpha_t \), it is the agent’s information rent \( \xi_t \). The two processes are:

\[
\begin{align*}
p_{\dot{p}t} &= \text{E}_{Pt} \left[ \int_{t_0}^{t} e^{-r(s-t)}c_s ds + e^{-r(t-t_0)}R(\hat{\mu}_{nt}) \right], \\
\xi_t &= \text{E}_{Pt} \left[ \int_{t_0}^{t} e^{-(r+\nu)(s-t)} \nu \left( \beta_{\mu s} - \beta_{\mu s} \frac{1}{\bar{r} \sigma_\mu} \right) \frac{\nu}{\sqrt{\bar{r} \sigma_\mu}} dt + e^{-(r+\nu)(t-t_0)} \dot{\lambda} \right]. \quad (A39)
\end{align*}
\]

This can be verified by noting that they match the boundary condition for \( p_{\dot{p}t} \) and \( p_{\alpha t} \) in (A38) and that the dynamics match (A37). Specifically, the following processes, beginning at the hiring date \( t_0 \), are martingales and therefore have zero drift:

\[
\begin{align*}
\text{E}_{Pt} \left[ \int_{t_0}^{t} e^{-r(s-t)}c_s ds + e^{-r(t-t_0)}R(\hat{\mu}_{nt}) \right] &= \int_{t_0}^{t} e^{-r(s-t)}c_s ds + e^{-r(t-t_0)}p_{\dot{p}t}, \\
\text{E}_{Pt} \left[ \int_{t_0}^{t} e^{-(r+\nu)(s-t)} \nu \left( \beta_{\mu s} - \beta_{\mu s} \frac{1}{\bar{r} \sigma_\mu} \right) \frac{\nu}{\sqrt{\bar{r} \sigma_\mu}} dt + e^{-(r+\nu)(t-t_0)} \dot{\lambda} \right] &= \int_{t_0}^{t} e^{-(r+\nu)(s-t)} \nu \left( \beta_{\mu t} - \beta_{\mu t} \frac{1}{\bar{r} \sigma_\mu} \right) \frac{\nu}{\sqrt{\bar{r} \sigma_\mu}} ds + e^{-(r+\nu)(t-t_0)} \xi_t. \quad (A40)
\end{align*}
\]
The zero drift implies:

$$E_{P_t}[dp_{\phi t}] = (r p_{\phi t} - c_t) \, dt,$$

$$E_{P_t}[d\xi_t] = \left((r + \nu) \xi_t - \nu \left(\beta_{\mu t} - \beta_{\hat{\epsilon} t} \frac{1}{\sigma_{\mu}} \frac{\rho_{ue}}{\sqrt{1 - \rho_{ue}^2}}\right)\right) \, dt.$$  \hspace{1cm} (A41)

Put $p_{\phi t} = w_t$ and $p_{\alpha t} = \xi_t$:

$$dw_t = (rw_t - c_t) \, dt + \frac{1}{\sigma_{\mu}} q_{\phi t} d\tilde{\mu}_{nt} + \frac{1}{\sigma_{\nu}} q_{\phi t} d\tilde{z}_{net}$$

$$d\xi_t = \left((r + \nu) \xi_t - \nu \left(\beta_{\mu t} - \beta_{\hat{\epsilon} t} \frac{1}{\sigma_{\mu}} \frac{\rho_{ue}}{\sqrt{1 - \rho_{ue}^2}}\right)\right) \, dt + \frac{1}{\sigma_{\mu}} q_{\alpha t} d\tilde{\mu}_{nt} + \frac{1}{\sigma_{\hat{\epsilon} t}} q_{\alpha t} d\tilde{z}_{net}$$  \hspace{1cm} (A42)

This proves 1 and 2.

With $\phi_t = 1$ and $p_{\alpha t} = \xi_t$, the incentive constraint (A36) can be written as

$$\lambda + \xi_t \nu - \nu \left(\beta_{\mu t} - \frac{1}{\sigma_{\mu}} \beta_{\hat{\epsilon} t} \frac{\rho_{ue}}{\sqrt{1 - \rho_{ue}^2}}\right) \leq 0,$$

giving 3. If this constraint binds, then (A42) gives:

$$d\xi_t = (r \xi_t - \lambda) \, dt + \chi_{\mu t} d\tilde{\mu}_{nt} + \chi_{\hat{\epsilon} t} d\tilde{z}_{net},$$  \hspace{1cm} (A43)

with the following solution that can be verified as before for $\xi_t$:

$$\xi_t = E_{P_t} \left[\int_t^\tau e^{-r(s-t)} \lambda ds + e^{-r(\tau-t)} \frac{\lambda}{r}\right] = \frac{\lambda}{r}$$

$$\Rightarrow \beta_{\mu t} = \lambda \left(\frac{1}{v} + \frac{1}{r}\right) + \frac{1}{\sigma_{\mu}} \beta_{\hat{\epsilon} t} \frac{\rho_{ue}}{\sqrt{1 - \rho_{ue}^2}}.$$  \hspace{1cm} (A44)

Because $\xi_t$ is constant, $\chi_{\mu t} = \chi_{\hat{\epsilon} t} = 0$, which implies that $q_{\alpha t} = q_{\phi t} = 0$

If the constraint does not bind everywhere, then there is a nonnegative process $\varepsilon_t$ such that

$$\beta_{\mu t} - \frac{1}{\sigma_{\mu}} \beta_{\hat{\epsilon} t} \frac{\rho_{ue}}{\sqrt{1 - \rho_{ue}^2}} = \frac{1}{v} \lambda + \xi_t + \varepsilon_t$$

and $d\xi_t = (r \xi_t - \lambda - \nu \varepsilon_t) \, dt + \chi_{\mu t} d\tilde{\mu}_{nt} + \chi_{\hat{\epsilon} t} d\tilde{z}_{net}$

and:

$$\xi_t = E_{P_t} \left[\int_t^\tau e^{-r(s-t)} (\lambda + \varepsilon_s) ds + e^{-r(\tau-t)} \frac{\lambda}{r}\right] = \frac{\lambda}{r} + E_{P_t} \left[\int_t^\tau e^{-r(s-t)} \varepsilon_s ds\right] \geq \frac{\lambda}{r},$$  \hspace{1cm} (A45)
Proof of Proposition 2

Part 1: To implement a contract that terminates when \( \mu_{nt} \) hits \( \mu \), it is necessary to set \( \beta_{\ell t} = 0 \). Otherwise, \( w_t \) may cross \( R(\mu) \), leading to termination, even though \( \mu_{nt} \) has not crossed \( \mu \). With a fixed liquidation contract, \( E_t[e^{-r(t-t')}| \mu_{nt} = \hat{\mu}]/\sigma_\mu \). Incentive compatibility then gives \( \beta_{\mu t} \geq \frac{\lambda}{\nu} + \xi_t \geq \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) \). Taking the agent’s continuation value as a function of \( \mu_{nt} \), we have:

\[
 w_t = R(\mu) + \int_\mu^{\hat{\mu}_{nt}} \beta_{\mu s} \text{d}\hat{\mu}_{ns} \geq R(\mu) + \int_\mu^{\hat{\mu}_{nt}} \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) \text{d}\hat{\mu}_{ns} = w(\mu_{nt}, \mu). \tag{A46}
\]

If the agent’s continuation value \( w_t < w(\mu_{nt}, \mu) \), then incentive compatible contracts have paths that can lead to termination prior to \( \mu_{nt} \) crossing \( \mu \). Therefore, the contract sets the initial reservation value to \( w(0, \mu) \) for all agents with \( w_0 \leq w(0, \mu) \) and the participation constraint does not bind.

Part 2: For any fixed termination contract with threshold \( \mu \), we have:

\[
 w_t = E_t \left[ \int_t^T e^{-r(s-t)} c_s \text{d}s \right] + e^{-\sqrt{2}r(\mu_{nt} - \mu)/\sigma_\mu} R(\mu), \tag{A47}
\]

so that the payments (26) yield \( w_t = w_f(\mu_{nt}, \mu) \). This gives:

\[
 \text{d}w_t = (rw_t - c_t) \text{d}t + \left( \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) + \frac{\sqrt{2}r}{\sigma_\mu} e^{-\sqrt{2}r(\mu_{nt} - \mu)/\sigma_\mu} \max\{0, \frac{1}{r} c - R(\mu)\} \right) \text{d}\mu_{nt}. \tag{A48}
\]
The contract yields the following information rents:

\[ \xi_t = E_t \left[ \int_t^\tau e^{-(r+\nu)(s-t)} \nu \beta_{\mu s} ds + e^{-(r+\nu)(\tau-t)} \frac{\lambda}{\tau} \right] \]

\[ = \beta_{\mu t} - \frac{\lambda}{\nu} - e^{-\sqrt{2(r+\nu)(\tilde{\mu}_{nt}-\mu)}/\sigma_{\mu}} \sqrt{2r} \max\{0, \frac{1}{\tau} - R(\mu)\}, \]

\[ = \frac{\lambda}{\nu} + \frac{\sqrt{2r}}{\sigma_{\mu}} \left( e^{-\sqrt{2r}(\tilde{\mu}_{nt}-\mu)/\sigma_{\mu}} - e^{-\sqrt{2(r+\nu)(\tilde{\mu}_{nt}-\mu)/\sigma_{\mu}}} \right) \max\{0, \frac{1}{\tau} - R(\mu)\}. \]  

(A49)

which implies that \( \beta_{\mu t} \geq \frac{\lambda}{\nu} + \xi_t \) so that the incentive compatibility constraint is satisfied.

To establish global incentive compatibility, we show that if the CEO has followed the policy \( a_t = 0 \) up to some time \( t_0 \), then the agent will not profit from any global deviation \( a_t \geq 0 \) for any \( t > t_0 \). Denoting the deviation in beliefs by

\[ \alpha_t = \nu \int_{t_0}^t e^{-\nu(t-s)} a_s ds \geq 0, \]

the flow compensation is:

\[ c_t = \max\{c, rR(\mu)\} + \lambda \frac{T}{\nu} \left( \tilde{\mu}_{nt} - \alpha_t - \mu \right) \]  

(A50)

Because the principal’s beliefs are \( \tilde{\mu}_{nt} - \alpha_t \leq \hat{\mu}_{nt} \), the termination time \( \tau_a \) with a deviation is weakly earlier than the termination time \( \tau \) without. The agent’s payoff with a deviation is:

\[ w_{t_0}^a = E_t \left[ \int_{t_0}^{\tau_a} e^{-r(s-t_0)} (c_a + \lambda a_s) ds + e^{-r(\tau_a-t_0)} \hat{R}(\mu, \hat{\mu}_{nta}) \right] + E_{t_0} \left[ e^{-r(\tau_a-t_0)} \hat{R}(\mu, \hat{\mu}_{nta}) \right] \]  

(A51)

Continuation value \( w_{t_0} \) with \( a_t = 0 \) \( \forall t > t_0 \)

\[ + E_{t_0} \left[ \int_{t_0}^{\tau_a} e^{-r(s-t_0)} \left( \lambda a_s - \lambda \frac{T}{\nu} \alpha_s \right) ds \right] + E_{t_0} \left[ e^{-r(\tau_a-t_0)} \hat{R}(\mu, \hat{\mu}_{nta}) \right] \]  

(A52)

Incremental value from \( t_0 \) to \( \tau_a \)

\[ - E_{t_0} \left[ e^{-r(\tau_a-t_0)} \left( \int_{\tau_a}^{\tau} e^{-r(s-\tau_a)} \left( \max\{c, rR(\mu)\} + \lambda \frac{T}{\nu} \left( \tilde{\mu}_{ns} - \mu \right) \right) ds + e^{-r(\tau-\tau_a)} R(\mu) \right) \right] \]  

(A53)

Incremental value from \( \tau_a \) to \( \tau \)
The first term is:
\[
E_{t_0} \left[ \int_{t_0}^{\tau_a} e^{-r(s-t)} \alpha_s ds \right] = -\frac{1}{r} E_{t_0} \left[ e^{-r(\tau_a-t_0)} \left( \hat{\mu}_{\tau_a} - \mu \right) - \int_{t_0}^{\tau_a} e^{-r(s-t_0)} \nu (a_s - \alpha_s) ds \right]
\]
\[\Rightarrow E_{t_0} \left[ \int_{t_0}^{\tau_a} e^{-r(s-t_0)} \alpha_s ds \right] = \frac{\nu}{\nu + r} E_{t_0} \left[ \int_{t_0}^{\tau_a} e^{-r(s-t_0)} a_s ds \right] - \frac{1}{\nu + r} E_{t_0} \left[ e^{-r(\tau_a-t_0)} \alpha_{\tau_a} \right],
\]

(A55)

where we used \(d\alpha_t = \nu(a_t - \alpha_t)dt\), \(\alpha_{\tau_a} = \hat{\mu}_{\tau_a} - \underline{\mu}\), and the implicit function theorem.

This gives the following, using \(\hat{R}(\mu, \hat{\mu}_{\tau_a}) = R(\mu) + \frac{1}{r} (\hat{\mu}_{\tau_a} - \mu)\):
\[
E_{t_0} \left[ \int_{t_0}^{\tau_a} e^{-r(s-t_0)} \left( \lambda a_s - \lambda \frac{\nu + r}{\nu} \alpha_s \right) ds \right] + E_{t_0} \left[ e^{-r(\tau_a-t_0)} \hat{R}(\mu, \hat{\mu}_{\tau_a}) \right]
\]
\[= \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) E_{t_0} \left[ e^{-r(\tau_a-t_0)} (\hat{\mu}_{\tau_a} - \underline{\mu}) \right] + E_{t_0} \left[ e^{-r(\tau_a-t_0)} \right] R(\mu).
\]

(A56)

We also have:
\[
E_{t_a} \left[ \int_{t_a}^{\tau} e^{-r(s-t_a)} \left( \max \{ \xi, rR(\mu) \} + \lambda \frac{\nu + r}{\nu} \left( \hat{\mu}_{s_a} - \underline{\mu} \right) \right) ds + e^{-r(\tau-t_a)} R(\mu) \right]
\]
\[= \max \left\{ \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) \left( \hat{\mu}_{\tau_a} - \underline{\mu} \right), E_{t_a} \left[ e^{-r(\tau-t_a)} \right] \max \{ 0, \frac{1}{\nu} \xi - R(\mu) \} \right\}.
\]

(A57)

Combining terms then gives the following for incremental value:
\[
w_{t_0} - w_{t_0} = -E_{t_0} \left[ e^{-r(\tau_a-t_0)} - e^{-r(t-t_0)} \right] \max \{ 0, \frac{1}{\nu} \xi - R(\mu) \} \leq 0,
\]

(A58)

where \(\tau \geq \tau_a\) implies the inequality and the inequality is strict if \(R(\mu) < \frac{1}{\nu} \xi\) and the agent deviates by putting \(a_t > 0\) for some time \(t > t_0\).

The value of firm-specific cash flows to the principal is:
\[
b_n(\hat{\mu}_{nt}, w_t) = E_t \left[ \int_t^{\tau} e^{-r(s-t)} \left( \hat{\mu}_{s_a} - c_s \right) ds + e^{-r(s-t)} \left( b_n \left( 0, \max \{ w_f(0, \mu), w_0 \} \right) - k \right) \right]
\]
\[= \frac{1}{r} \hat{\mu}_{nt} + e^{-\sqrt{2r}(\hat{\mu}_{nt}-\underline{\mu})/\sigma_n} \left( b_n \left( 0, \max \{ w_f(0, \mu), w_0 \} \right) + \left( \frac{1}{r} - R(\mu) \right) (\hat{\mu}_{nt} - \underline{\mu}) \right)
\]

(A59)

(A60)

Setting the parameters to their values at contract initiation and solving for \(b_n(0, \max \{ w(0, \mu), w_0 \})\)
gives:

\[
    b_n \left(0, \max\{w(0, \mu), w_0\}\right) = \frac{e^{\frac{\sigma}{2}R_o/\sigma} - R_o}{1 - e^{\frac{\sigma}{2}R_o/\sigma}} \left(\frac{1}{r} - R_o\right) (\mu_o - \mu) - \frac{1}{1 - e^{\frac{\sigma}{2}R_o/\sigma}} \max\{w(0, \mu), w_0\}. 
\]

(A61)

Substituting back into \(b_n(\hat{\mu}_{nt}, w_t)\) yields (28).

**Part 3:** If \(R(\mu) \geq \frac{1}{r}C\), then a contract with \(c_t\) defined by (26) satisfies the minimum payment condition and the continuation value \(w(\hat{\mu}_{nt}, \mu)\). This gives \(\beta_{\mu t} = \lambda \left(\frac{1}{\nu} + \frac{1}{r}\right)\) and information rents of \(\xi_t = \frac{1}{r}\) so that \(\beta_{\mu t} = \frac{\lambda}{\nu} + \xi_t\) and the incentive constraint is binding at all dates. If the incentive compatibility constraint binds at all dates, then Lemma 4 that information rents \(\xi_t = \frac{1}{r}\) and \(\beta_{\mu t} = \lambda \left(\frac{1}{\nu} + \frac{1}{r}\right)\). Then:

\[
    w_t = R(\mu) + \int_{\mu}^{\hat{\mu}_{nt}} \beta_{\mu s} d\hat{\mu}_{ns} = w(\hat{\mu}_{nt}, \mu). \tag{A62}
\]

This gives \(dw_t = \lambda \left(\frac{1}{\nu} + \frac{1}{r}\right) d\hat{\mu}_{nt}\), so that \(w_t\) has zero drift. Because the drift of \(w_t\) is \(rw_t - c_t\), this implies \(c_t = R(\mu) + \lambda \frac{\nu r}{\nu} \left(\hat{\mu}_{nt} - \mu\right)\), which satisfies the minimum payment constraint if and only if \(R(\mu) \geq \frac{1}{r}C\). This proves Claim 3.

**Derivation of expression (33)**

To show the convexity of \(b_n\), direct computations give:

\[
    \frac{\partial^2 b_n}{\partial \hat{\mu}_{nt}^2} = \frac{2}{\sigma^2} e^{-\frac{\sigma}{2}(\hat{\mu}_{nt} - \mu)/\sigma} \left(b_n(0) + \left(\frac{1}{r} - R_o\right) (\mu_o - \mu) + \max\{0, \frac{1}{r}C - R(\mu)\}\right) > 0, \tag{A63}
\]

where the inequality follows because \(b_n(0) > 0\) for any case where the firm is willing to hire a CEO, and the threshold \(\mu\) is less than the myopic threshold \(\mu_o\). Convexity
implies that:

\[
\frac{\partial b_n}{\partial \mu_{nt}} \geq \frac{\partial b_n}{\partial \mu_{nt}} \bigg|_{\mu_{nt}=\hat{\mu}} = \frac{1}{r} - \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) \frac{\sqrt{2\pi \sigma^2}}{\sigma_\mu} \left( \frac{1}{\nu} R(\mu) - \frac{\mu - \nu}{\nu} - \frac{\max\{0, \frac{1}{r} \nu - R(\mu)\} \nu}{1 - e^{\sqrt{2\pi \sigma^2} \sigma_\mu}} \right) 
\]

\[
= \frac{1}{r} - \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) \frac{\sqrt{2\pi \sigma^2}}{\sigma_\mu} \left( \frac{1}{\nu} R(\mu) - \frac{\mu - \nu}{\nu} + \frac{\max\{0, \frac{1}{r} \nu - R(\mu)\} \nu}{1 - e^{\sqrt{2\pi \sigma^2} \sigma_\mu}} \right) 
\]

\[
+ \frac{\sqrt{2\pi \sigma^2}}{\sigma_\mu} \frac{\max\{0, \omega_0 - w_f(0, \mu)\}}{1 - e^{\sqrt{2\pi \sigma^2} \sigma_\mu}} 
\]

\[
= \left( \frac{1}{r} - \lambda \left( \frac{1}{\nu} + \frac{1}{r} \right) \right) \left( 1 + \frac{\sqrt{2\pi \sigma^2}}{\sigma_\mu} \frac{\sqrt{2\pi \sigma^2}}{\sigma_\mu} \mu \right) + \frac{\sqrt{2\pi \sigma^2}}{\sigma_\mu} \frac{\max\{0, \omega_0 - w_f(0, \mu)\}}{1 - e^{\sqrt{2\pi \sigma^2} \sigma_\mu}} > 0, 
\]

(A64)

where the inequality follows because the function \(1 - \frac{e^{-x}}{1 - e^{-x}} > 0\) for all \(x > 0\) and \(\mu < 0\), and \(\lambda < \frac{\nu}{\nu + r}\). This implies that we can invert \(b_n\) to express beliefs as a function of firm value. We derive expression (33) by rearranging the definition of market value \(b\) to arrive at:

\[
-e^{-\sqrt{2\pi \sigma^2} \sigma_\mu \mu_{nt}} m_0 = e^{\sqrt{2\pi \sigma^2} \sigma_\mu \mu_{nt}} - \mu - m_b, 
\]

(A65)

and applying \(\omega\) to both sides of the equality and solving for \(\hat{\mu}_{nt}\).