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

Judson Caskey
UCLA,

George-Levi Gayle
Washington University, St. Louis,

and

Chen Li
CUNY, Baruch College

May 22, 2018

VERY PRELIMINARY AND INCOMPLETE
DO NOT CITE WITHOUT PERMISSION

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

Abstract

In its current state, this paper develops a dynamic model where a firm chooses compensation schemes to motivate CEO effort and chooses when to replace CEOs. The contract depends on the observed streams of cash flows and accrual-based earnings. In this context, accounting contributes to contracting efficiency to the extent to which shocks to accrual-based earnings negate transitory shocks to cash flows. The optimal turnover choice depends on economizing the net costs of termination and the costs of incentivizing the CEO. In future work, we will structurally estimate the model's parameters to evaluate its descriptive power and to quantify the effects of the information environment on contracting efficiency.

1 Introduction

This paper develops a model that is a step to a broader project in which we will quantify the effects of accounting information on CEO pay and turnover.The model extends (DeMarzo and Sannikov 2017) in two ways. First, (DeMarzo and Sannikov 2017) examine a single employment relationship that ultimately yields terminal payoffs when the relationship and the firm are dissolved. In our model, the firm hires a new CEO at each termination period. Second, (DeMarzo and Sannikov 2017) model only a cash flow process, whereas we introduce an additional earnings process that incorporates key features of accrual-based earnings such as matching to cash flows over time. This allows us to disentangle the effects of information quality from the effects of cash flow volatility.

In the model, the firm hires a CEO and is uncertain of his or her ability. The firm and the CEO share the same prior beliefs about that ability, and the ability varies over time. The firm designs a compensation contract to discourage shirking, and to ensure that the CEO does not leave the firm prematurely. The firm faces costs to replace the CEO, and will terminate the CEO when its beliefs about CEO ability become sufficiently low.

The addition of recurring employment relationships significantly alters the firm's preferred termination policy. With a single employment spell, when there are no agency costs the firm prefers a surplus-maximizing termination policy that has the same form as in standard real options problems. With recurring employment relationships, the surplus-maximizing policy reflects the incentive to economize the net cost of replacing agents, which is the difference between the firm's costs of replacing the CEO versus the value of the CEO's outside opportunities. When there is no net cost of replacing the CEO, the surplus-maximizing policy matches the myopic policy, whereas the surplus-maximizing policy in standard real options problems terminates at strictly lower values than the myopic policy.

When the firm faces no agency costs and must only meet CEOs reservation utility, it prefers a strictly lower termination threshold than the surplus-maximizing threshold. The reason for this is that the firm wishes to economize on its cost of hiring a new CEO, and does not consider the value of the CEO's outside opportunities. By delaying termination, the firm reduces overall surplus, but also reduces the value that it must share with the CEO.

Currently, we derive the optimal termination policy within the class of contracts with a constant termination threshold. Given such a contract, the effect of information quality on firm value and CEO turnover hinges on whether the CEO has a relatively high reservation utility. When the reservation utility is low, firm value is everywhere increasing in information quality, and higher information quality leads to shorter CEO tenure. When the reservation utility is high, greater information quality leads to a lower termination threshold. This occurs because a high reservation utility implies a relatively high cost of hiring a new CEO. The lower termination threshold then implies longer CEO tenure with higher information quality.

In future work, we will further develop the model by deriving the fully optimal contract, without restricting the termination policy. We will also structurally estimate the model's parameters to evaluate the model's fit to data and quantify the effect of accounting information on firm value and CEO turnover.

2 Model

2.1 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 μ_{0t} and the CEO's ability μ_{nt}. Both of these are unobservable so that firm value from the principal's perspective depends on estimates $\hat{\mu}_{0t} = \mathrm{E}_t[\mu_{0t}]$ and $\hat{\mu}_{nt} = \mathrm{E}_t[\mu_{nt}]$. 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 $b_t = b\left(\hat{\mu}_{0t}, \hat{\mu}_{nt}, w_t\right)$ as follows where τ_i denotes the i^{th} stopping time to replace the CEO and the discount rate is r:

$$\begin{aligned} b\left(\hat{\mu}_{0t}, \hat{\mu}_{nt}, w_t\right) &= \mathrm{E}_t\left[\int_t^{\tau_1} \mathrm{e}^{-r(s-t)}\left(\mathrm{d}x_{ns} - c_s \mathrm{d}s\right) + \sum_{i=1}^{\infty}\left(\int_{\tau_i}^{\tau_{i+1}} \mathrm{e}^{-r(s-t)}\left(\mathrm{d}x_{ns} - c_s \mathrm{d}s\right) - \mathrm{e}^{-r(\tau_i - t)} k\right)\right] \\ &= \mathrm{E}_t\left[\int_t^{\tau_1} \mathrm{e}^{-r(s-t)}\left(\mathrm{d}x_{ns} - c_s \mathrm{d}s\right) + \mathrm{e}^{-r(\tau_1 - t)}\left(b\left(\hat{\mu}_{0\tau_1}, 0, w_{\tau_1}\right) - k\right)\right]. \end{aligned} \tag{1}$$

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:

$$\mathrm{d}x_{nt} = \left(\mu_{0t} + \mu_{nt} - a_{nt}\right)\mathrm{d}t + \sigma_x\left(\beta \mathrm{d}z_{0xt} + \mathrm{d}z_{nxt}\right), \quad x_{n0} = 0, \tag{2}$$

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

$$\mathrm{d}\mu_{0t} = \beta\sigma_\mu \mathrm{d}z_{0\mu t}, \qquad \mathrm{d}\mu_{nt} = \sigma_\mu \mathrm{d}z_{n\mu t}, \tag{3}$$

where σ_μ is a parameter and $z_{0\mu t}$ and $z_{n\mu t}$ are independent standard Brownian motions that represent shocks to the respective industry and CEO-specific portions of profitability. The term μ_{0t} includes any correlated aspects of CEO ability so that μ_{nt} is purely idiosyncratic.

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

$$\mathrm{d}e_{nt} = \theta\left(x_{nt} - e_{nt}\right)\mathrm{d}t + \sigma_e\left(\beta \mathrm{d}z_{0et} + \mathrm{d}z_{net}\right), \quad e_{n0} = 0, \tag{4}$$

where the parameter θ governs how quickly accruals reverse in the sense of cumulative earnings converging to cumulative cash flows, σ_e is a volatility parameter, and z_{0et} 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:

$$\mathrm{d}x_{0t} = \mu_{0t}\mathrm{d}t + \beta\sigma_x \mathrm{d}z_{0xt}, \qquad \mathrm{d}e_{0t} = \theta\left(x_{0t} - e_{0t}\right)\mathrm{d}t + \beta\sigma_e \mathrm{d}z_{0et}. \tag{5}$$

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 λa_t from shirking, where $\lambda \in (0, \frac{\nu}{\nu+r})$so that the action is socially wasteful.[1] 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 μ_{nt}. In this case, the agent's beliefs $\hat{\mu}^a_{nt}$ 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 $\hat{\mu}^a_{nt}$ and $\hat{\mu}_{nt}$, respectively, then the present value of his future payoffs is $\hat{R}(\hat{\mu}_t, \hat{\mu}^a_t)$. Denoting by τ the stopping time at which the agent leaves the firm, the agent's continuation value is then the following where $\mathrm{E}^a_t[\cdot]$ denotes expectation with respect to the agent's beliefs:

$$w^a_t = \mathrm{E}^a_t\left[\int_t^\tau \mathrm{e}^{-r(s-t)}(\lambda a_s + c_s)\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\hat{R}(\hat{\mu}_t, \hat{\mu}^a_t)\right]. \tag{6}$$

2.2 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 $\boldsymbol{\mu}_{Nt} = \{\mu_{1t}, \mu_{2t}, \ldots, \mu_{Nt}\}$, 1 denotes a vector of ones, $\mathbf{0}$ denotes a vector of zeros, and $\boldsymbol{I}$ denotes the identity matrix:

$$\begin{pmatrix} \mu_{00} \\ \boldsymbol{\mu}_{N0} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \hat{\mu}_{00} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \hat{\gamma}_0 & \mathbf{0}' \\ \mathbf{0} & \hat{\gamma}_n \boldsymbol{I} \end{pmatrix}\right). \tag{7}$$

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

[1] DeMarzo and Sannikov (2017) place the weaker restriction $\lambda \in (0, 1)$. We require the more stringent upper bound on λ because $\lambda > \frac{\nu}{\nu+r}$ 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:

$$\begin{aligned} \rho_{\mu x} &= \tfrac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}z_{0\mu t}\mathrm{d}z_{0xt}\right] = \tfrac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}z_{n\mu t}\mathrm{d}z_{nxt}\right], \\ \rho_{\mu e} &= \tfrac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}z_{0\mu t}\mathrm{d}z_{0et}\right] = \tfrac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}z_{n\mu t}\mathrm{d}z_{net}\right], \\ \rho_{xe} &= \tfrac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}z_{0xt}\mathrm{d}z_{0et}\right] = \tfrac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}z_{nxt}\mathrm{d}z_{net}\right]. \end{aligned} \tag{8}$$

The steady-state variances are:

$$\begin{aligned} \hat{\gamma}_n &= \lim_{t\to\infty} \mathrm{var}_t\left(\mu_{nt}\right) = \sigma_\mu\sigma_x\left(\sqrt{\left(1-\rho_{\mu e}^2\right)\left(1-\rho_{xe}^2\right)} - \left(\rho_{\mu x} - \rho_{\mu e}\rho_{xe}\right)\right) \\ &= \underbrace{\sigma_\mu\sigma_x\sqrt{\left(1-\rho_{\mu e}^2\right)\left(1-\rho_{xe}^2\right)}}_{\mathrm{std}(z_\mu|z_e)\mathrm{std}(z_x|z_e)}\underbrace{\left(1-\frac{\rho_{\mu x}-\rho_{\mu e}\rho_{xe}}{\sqrt{\left(1-\rho_{\mu e}^2\right)\left(1-\rho_{xe}^2\right)}}\right)}_{1-\mathrm{corr}(z_\mu,z_x|z_e)} > 0, \\ \hat{\gamma}_0 &= \lim_{t\to\infty} \mathrm{var}_t\left(\mu_{0t}\right) = \beta^2\hat{\gamma}_n. \end{aligned} \tag{9}$$

The belief processes evolve as:

$$\begin{aligned} \mathrm{d}\hat{\mu}_{0t} &= \beta\sigma_\mu\mathrm{d}\hat{z}_{0\mu t}, \\ \mathrm{d}\hat{\mu}_{nt} &= \sigma_\mu\mathrm{d}\hat{z}_{n\mu t}, \end{aligned} \tag{10}$$

where:

$$\begin{aligned} \mathrm{d}\hat{z}_{0\mu t} &= \nu\tfrac{\sigma_x}{\sigma_\mu}\underbrace{\tfrac{1}{\beta\sigma_x}\left(\mathrm{d}x_{0t}-\hat{\mu}_{0t}\mathrm{d}t\right)}_{\mathrm{d}\hat{z}_{0xt}} + \tfrac{\rho_{\mu e}\sigma_\mu-\nu\rho_{xe}\sigma_x}{\sigma_\mu}\underbrace{\tfrac{1}{\beta\sigma_e}\left(\mathrm{d}e_{0t}-\theta\left(x_{0t}-e_{0t}\right)\right)}_{\mathrm{d}z_{0et}} \\ \mathrm{d}\hat{z}_{n\mu t} &= \nu\tfrac{\sigma_x}{\sigma_\mu}\underbrace{\tfrac{1}{\sigma_x}\left(\mathrm{d}x_{nt}-\left(\hat{\mu}_{0t}+\hat{\mu}_{nt}-a_{nt}\right)\mathrm{d}t-\beta\sigma_x\mathrm{d}\hat{z}_{0xt}\right)}_{\mathrm{d}\hat{z}_{nxt}} \\ &\quad + \tfrac{\rho_{\mu e}\sigma_\mu-\nu\rho_{xe}\sigma_x}{\sigma_\mu}\underbrace{\tfrac{1}{\sigma_e}\left(\mathrm{d}e_{nt}-\theta\left(x_{nt}-e_{nt}\right)\mathrm{d}t-\beta\sigma_e\mathrm{d}z_{0et}\right)}_{\mathrm{d}z_{net}}, \end{aligned} \tag{11}$$

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

$$\mathrm{d}x_{0t} = \hat{\mu}_{0t}\mathrm{d}t + \beta\sigma_x\mathrm{d}\hat{z}_{0xt}, \qquad \mathrm{d}x_{nt} = \left(\hat{\mu}_{0t}+\hat{\mu}_{nt}\right)\mathrm{d}t + \sigma_x\left(\beta\mathrm{d}\hat{z}_{0xt}+\mathrm{d}\hat{z}_{nxt}\right). \tag{12}$$

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 $\hat{\mu}^a_{nt}$ follow:

$$\begin{aligned}
\mathrm{d}\hat{\mu}^a_{nt} &= \sigma_\mu \left(\sqrt{\tfrac{1-\rho^2_{\mu e}}{1-\rho^2_{xe}}} \tfrac{1}{\sigma_x} \left(\mathrm{d}x_{nt} - \left(\hat{\mu}_{0t} + \hat{\mu}^a_{nt} - \hat{a}_{nt} \right) \mathrm{d}t - \left(\mathrm{d}x_{0t} - \hat{\mu}_{0t} \mathrm{d}t \right) \right) \right. \\
&\qquad \left. + \left(\rho_{\mu e} - \rho_{xe} \sqrt{\tfrac{1-\rho^2_{\mu e}}{1-\rho^2_{xe}}} \right) \mathrm{d}z_{net} \right) \\
&= \mathrm{d}\hat{\mu}_{nt} + \nu \left(\hat{a}_{nt} - a_{nt} - \left(\hat{\mu}^a_{nt} - \hat{\mu}_{nt} \right) \right) \mathrm{d}t.
\end{aligned} \tag{13}$$

The above dynamics imply the belief divergence $\alpha_t = \hat{\mu}^a_{nt} - \hat{\mu}_t$ is:

$$\alpha_t = \nu \int_0^t \mathrm{e}^{-\nu(t-s)} \left(\hat{a}_{ns} - a_{ns} \right) \mathrm{d}s, \quad \mathrm{d}\alpha_t = \nu \left(\hat{a}_t - \alpha_t \right) \mathrm{d}t. \tag{14}$$

The term ν 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 ν 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 ρ_{xe}), then earnings will be highly informative about profitability but relatively less useful for controlling agency costs (low $\hat{\gamma}_n$ and $\nu < \frac{\sigma_\mu}{\sigma_x}$).

For convenience, we express the dynamics in terms of the information generated by the shock $\mathrm{d}\hat{z}_{n\mu t}$ to beliefs about μ_{nt} and the portion of the earnings shock $\mathrm{d}z_{net}$ that is orthogonal to $\mathrm{d}\hat{z}_{n\mu t}$, which we denote by $\mathrm{d}\tilde{z}_{net} = \frac{1}{\sqrt{1-\rho^2_{\mu e}}} \left(\mathrm{d}z_{net} - \rho_{\mu e} \mathrm{d}\hat{z}_{n\mu t} \right)$. The filtration generated by $\{\mathrm{d}\hat{z}_{n\mu t}, \mathrm{d}\tilde{z}_{net}\}$ is informationally equivalent to the filtration generated by $\{\mathrm{d}\hat{z}_{nxt}, \mathrm{d}z_{net}\}$.

2.3 Contracting and the first-best solution

The firm pays flow compensation c_t to the CEO, which must satisfy $c_t \geq \underline{c} \geq 0$. We follow DeMarzo and Sannikov (2017) and use a linear form for the CEO's termination payoff:

$$\hat{R}(\hat{\mu}_t, \hat{\mu}^a_t) = \underbrace{R_0 + R_\mu \hat{\mu}_{nt}}_{R(\hat{\mu}_t)} + \tfrac{\lambda - \psi}{r} \left(\hat{\mu}^a_{nt} - \hat{\mu}_{nt} \right), \tag{15}$$

where the term $\frac{\lambda-\psi}{r}(\hat{\mu}^a_{nt}-\hat{\mu}_{nt})$ reflects that the CEO can earn $\lambda-\psi$ from the deviation $\hat{\mu}^a_{nt}-\hat{\mu}_{nt}$ given in (14) in perpetuity, and the parameter $\psi \in [0,\lambda]$ represents the extent to which the agent's information rents are firm-specific. Because R_0 and R_μ are fixed, this specification omits the possibility that the CEO's outside opportunities change during the contract period. In other words, the firm can anticipate any 'poaching' of the CEO. Furthermore, principals do not choose R_0 and R_μ, 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 $R_\mu \in [0, \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)-\frac{\psi}{r})$ or, equivalently, $\lambda > \frac{\nu}{\nu+r}(rR_\mu+\psi)$ so that $\lambda \in \left(\frac{\nu}{\nu+r}(rR_\mu+\psi), \frac{\nu}{\nu+r}\right)$.

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

$$\begin{aligned} v_n(\hat{\mu}_{nt}) &= \mathrm{E}_t\left[\int_t^\tau \mathrm{e}^{-r(s-t)}\hat{\mu}_{ns}\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\left(v_n(0)+R(\hat{\mu}_{n\tau})-k\right)\right] \\ &= \tfrac{1}{r}\hat{\mu}_{nt} + \mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\left(v_n(0)-\tfrac{1}{r}\hat{\mu}_{n\tau}+R(\hat{\mu}_{n\tau})-k\right)\right] \end{aligned} \tag{16}$$

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

$$\tfrac{1}{r}\underline{\mu}_o = R(\underline{\mu}_o)-k \quad \Rightarrow \quad \underline{\mu}_o = -\frac{k-R_0}{\frac{1}{r}-R_\mu}. \tag{17}$$

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,τ) where c_t denotes cash payments, and the stopping time τ 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 initial expected value of working for the firm:

$$b\left(\hat{\mu}_{0t}, \hat{\mu}_{nt}, w_t\right) = \tfrac{1}{r}\hat{\mu}_{0t} + \underbrace{\mathrm{E}_t\left[\int_t^{\tau} \mathrm{e}^{-r(s-t)}\left(\hat{\mu}_{ns} - a_s - c_s\right)\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\left(b_n\left(0, w_0\right) - k\right)\right]}_{b_n(\hat{\mu}_{nt}, w_t)}. \tag{18}$$

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

$$\underline{w}_0 \leq \underbrace{\mathrm{E}\left[\int_{\tau_i}^{\tau_{i+1}} \mathrm{e}^{-r(s-\tau_i)}(\lambda a_s + c_s)\mathrm{d}s + \mathrm{e}^{-r(\tau_{i+1}-\tau_i)}R(\hat{\mu}_{n\tau_{i+1}})\right]}_{w_0}, \tag{19a}$$

$$w_t = \mathrm{E}\left[\int_t^{\tau_{i+1}} \mathrm{e}^{-r(s-t)}(\lambda a_s + c_s)\mathrm{d}s + \mathrm{e}^{-r(\tau_{i+1}-t)}R(\hat{\mu}_{n\tau_{i+1}})\right] \geq R\left(\hat{\mu}_{nt}\right), \tag{19b}$$

$$w_{\tau_i} \geq \mathrm{E}\left[\int_{\tau_i}^{\tau_{i+1}} \mathrm{e}^{-r(s-\tau_i)}(\lambda \hat{a}_s + c_s)\mathrm{d}s + \mathrm{e}^{-r(\tau_{i+1}-\tau_i)}\hat{R}(\hat{\mu}_{n\tau_{i+1}}, \hat{\mu}^a_{n\tau_{i+1}})\right], \ \forall \hat{a}_t \tag{19c}$$

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 $\underline{w}_0$ 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:

$$\begin{aligned} b_n(\hat{\mu}_{nt}, w_t) &= \mathrm{E}_t\left[\int_t^{\tau} \mathrm{e}^{-r(s-t)}\left(\hat{\mu}_{ns} - c_s\right)\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\left(b_n(0, w_0) - k\right)\right] \\ &= \mathrm{E}_t\left[\int_t^{\tau} \mathrm{e}^{-r(s-t)}\hat{\mu}_{ns}\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\left(b_n(0, w_0) + R(\hat{\mu}_{\tau}) - k\right)\right] - w_t \\ &= \tfrac{1}{r}\hat{\mu}_{nt} + \mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\left(b_n(0, w_0) - \tfrac{1}{r}\hat{\mu}_{n\tau} + R(\hat{\mu}_{\tau}) - k\right)\right] - w_t \end{aligned} \tag{20}$$

The following proposition summarizes the termination rule $\underline{\mu}_{fb}$ that maximizes the surplus v_n, and the firm's preferred termination policy $\underline{\mu}_p$ that maximizes the firm's payoff b_n when only the participation constraint must be considered.

Proposition 1. *1. The surplus-maximizing termination policy is:*

$$\underline{\mu}_{fb} = \underline{\mu}_o - \frac{\sigma_\mu}{\sqrt{2r}} \left(1 + \omega \left(-\mathrm{e}^{\frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o - \frac{\sigma_\mu}{\sqrt{2r}}\right)} \right) \right) < \underline{\mu}_o, \tag{21}$$

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

$$v_n(\hat{\mu}_{nt}; \underline{\mu}_{fb}) = \tfrac{1}{r}\hat{\mu}_{nt} + \tfrac{\sigma_\mu}{\sqrt{2r}} \mathrm{e}^{-\sqrt{2r}(\hat{\mu}_{nt} - \underline{\mu}_{fb})/\sigma_\mu} \left(\tfrac{1}{r} - R_\mu \right), \tag{22}$$

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

2. *The firm's preferred termination policy $\underline{\mu}_p$ (including only the participation constraint) is:*

$$\underline{\mu}_p = \underline{\mu}_o - \frac{1}{\frac{1}{r} - R_\mu}\underline{w}_0 - \frac{\sigma_\mu}{\sqrt{2r}} \left(1 + \omega \left(-\mathrm{e}^{\frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o - \frac{1}{\frac{1}{r} - R_\mu}\underline{w}_0 - \frac{\sigma_\mu}{\sqrt{2r}}\right)} \right) \right) < \underline{\mu}_{fb}. \tag{23}$$

The value of firm-specific profits is:

$$b_n(\hat{\mu}_{nt}, \underline{w}_0; \underline{\mu}_p) = \tfrac{1}{r}\hat{\mu}_{nt} + \tfrac{\sigma_\mu}{\sqrt{2r}} \mathrm{e}^{-\sqrt{2r}(\hat{\mu}_{nt} - \underline{\mu}_p)/\sigma_\mu} \left(\tfrac{1}{r} - R_\mu \right) - \underline{w}_0. \tag{24}$$

This value can be implemented with a contract that pays the agent $c_t = r\underline{w}_0$ at all dates $t < \tau$ and a lump sum of $\underline{w}_0 - R(\underline{\mu}_p)$ at contract termination.

3. *The value of firm-specific profits is positive if and only if $\underline{w}_0 < \bar{w} = \left(\frac{1}{r} - R_\mu\right) \frac{\sigma_\mu}{\sqrt{2r}} \mathrm{e}^{-\left(1 - \sqrt{2r}\underline{\mu}_o/\sigma_\mu\right)}$. If $\underline{w}_0 < \bar{w}$, then $\underline{\mu}_o - \frac{\sigma_\mu}{\sqrt{2r}} < \underline{\mu}_p < \underline{\mu}_{fb} < \underline{\mu}_o$.*

In standard termination problems, the optimal threshold is $\underline{\mu}_o - \frac{\sigma_\mu}{\sqrt{2r}} \leq \underline{\mu}_{fb}$. Here, the threshold $\underline{\mu}_{fb}$ reflects the ability to economize on the net cost $k - R_0$ of firing. 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 $\underline{\mu}_o - \frac{\sigma_\mu}{\sqrt{2r}}$ 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 the same horizon and the firm prefers the surplus-maximizing threshold when there are no agency costs. If each agent is held to his reservation value of $\underline{w}_0$, then the first-best threshold $\underline{\mu}_{fb}$ yields the following expected surplus to the firm at each hiring date:

$$\underbrace{\frac{\sigma_\mu}{\sqrt{2r}} \mathrm{e}^{\sqrt{2r}\underline{\mu}_{fb}/\sigma_\mu} \left(\frac{1}{r} - R_\mu\right)}_{\text{Total surplus}} - \underbrace{\frac{1}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}_{fb}/\sigma_\mu}} \underline{w}_0}_{\substack{\text{Present value} \\ \text{of payments}}} . \tag{25}$$

The firm's preferred policy $\underline{\mu}_p < \underline{\mu}_{fb}$ incurs the cost of a lower surplus in order to reduce the present value of payments to agents, which can be seen by the absence of a $\frac{1}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}$ term multiplying $\underline{w}_0$ in (24). 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 $\underline{\mu}_p$ equals the surplus-maximizing threshold $\underline{\mu}_{fb}$ only if the agent has a zero reservation value ($\underline{w}_0 = 0$). 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.

3 Constant termination contracts

We consider contracts that have a constant termination threshold $\underline{\mu}$ throughout the contract's life. When the minimum payments $\underline{c}$ are sufficiently low relative to the agent's termination value $R(\underline{\mu})$, it is possible to construct contract where the incentive compatibility constraints are everywhere binding. The resulting contract sets the agent's continuation value to the minimum level required to maintain incentive compatibility.

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

1. *If* $\underline{w}_0 < \underline{w}(0, \underline{\mu})$*, then the agent's participation constraint does not bind.*

2. *If* $\frac{1}{r}\underline{c} > R(\underline{\mu}) - \frac{\sigma_\mu}{\sqrt{2r}}\frac{\lambda}{r}$*, then the incentive compatibility constraints are not binding for some* $\hat{\mu}_{nt}$*.*

3. *If* $\frac{1}{r}\underline{c} \leq R(\underline{\mu}) - \frac{\sigma_\mu}{\sqrt{2r}}\frac{\psi}{r}$*, then the unique incentive compatible contract with threshold* $\underline{\mu}$ *has payments:*

$$c_t = rR(\underline{\mu}) - \psi\frac{\sigma_\mu}{\sqrt{2r}} + \lambda\frac{\nu + r}{\nu}\left(\hat{\mu}_{nt} - \underline{\mu}\right) \geq \underline{c}, \tag{26}$$

that yields the following continuation value:

$$\underline{w}(\hat{\mu}_{nt}, \underline{\mu}) = R(\underline{\mu}) + \lambda\left(\frac{1}{\nu} + \frac{1}{r}\right)\left(\hat{\mu}_{nt} - \underline{\mu}\right) - \frac{\psi}{r}\frac{\sigma_\mu}{\sqrt{2r}}\left(1 - \mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt} - \underline{\mu}\right)/\sigma_\mu}\right). \tag{27}$$

The fastest possible payments include a lump-sum of $\max\{0, \underline{w}_0 - \underline{w}(\hat{\mu}_{nt}, \underline{\mu})\}$ *at contract initiation to satisfy the CEO's participation constraint. The value of* *firm-specific cash flows to the principal is:*

$$b_n(\hat{\mu}_{nt}) = \tfrac{1}{r}\hat{\mu}_{nt} + \tfrac{e^{-\sqrt{2r}(\hat{\mu}_{nt}-\underline{\mu})/\sigma_\mu}}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\left(\left(\tfrac{1}{r}-R_\mu\right)\left(\underline{\mu}_o-\underline{\mu}\right)-\max\{\underline{w}(0,\underline{\mu}),\underline{w}_0\}\right) - \underline{w}(\hat{\mu}_{nt},\underline{\mu}). \tag{28}$$

and at the initiation of each CEO:

$$b_n(0) - \max\{0, \underline{w}_0 - \underline{w}(\hat{\mu}_{nt},\underline{\mu})\} = \underbrace{\tfrac{e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\left(\tfrac{1}{r}-R_\mu\right)\left(\underline{\mu}_o-\underline{\mu}\right)}_{\textit{Total surplus}} - \underbrace{\tfrac{1}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\max\{\underline{w}(0,\underline{\mu}),\underline{w}_0\}}_{\textit{Value of payments}}. \tag{29}$$

4. *The smooth-pasting condition* $b_n'(\underline{\mu}) = 0$ *yields the optimal termination threshold* $\underline{\mu}_c$*, and yields the firm-specific cash flows:*

$$b_n(\hat{\mu}_{nt};\underline{\mu}_c) = \tfrac{1}{r}\hat{\mu}_{nt} + e^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}_c\right)/\sigma_\mu}\tfrac{\sigma_\mu}{\sqrt{2r}}\left(\tfrac{1}{r}-\lambda\left(\tfrac{1}{\nu}+\tfrac{1}{r}\right)+\tfrac{\psi}{r}\right) - \underline{w}(\hat{\mu}_{nt},\underline{\mu}_c). \tag{30}$$

The periodic compensation is:

$$\tfrac{e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\left(\tfrac{1}{r}-R_\mu\right)\left(\underline{\mu}_o-\underline{\mu}\right) - \tfrac{e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\max\{\underline{w}(0,\underline{\mu}),\underline{w}_0\} - \tfrac{e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\underline{w}(0,\underline{\mu}) \tag{31}$$

$$\begin{aligned} \int_t^{t+1} c_s \mathrm{d}s &= rR(\underline{\mu}) - \psi\tfrac{\sigma_\mu}{\sqrt{2r}} + \lambda\tfrac{\nu+r}{\nu}\left(\int_t^{t+1}\hat{\mu}_{ns}\mathrm{d}s - \underline{\mu}\right), \\ \mathrm{E}_t\left[\int_t^{t+1} c_s \mathrm{d}s\right] &= rR(\underline{\mu}) - \psi\tfrac{\sigma_\mu}{\sqrt{2r}} + \lambda\tfrac{\nu+r}{\nu}\left(\hat{\mu}_{nt} - \underline{\mu}\right), \end{aligned} \tag{32}$$

Earnings impacts the compensation via $\nu = \frac{\sigma_\mu}{\sigma_x}\sqrt{\frac{1-\rho_{\mu e}^2}{1-\rho_{xe}^2}}$. The more correlated earnings is with fundamentals (higher $\rho_{\mu e}^2$), the lower is ν and the greater is the incentive pay. A higher correlation with cash flows (higher ρ_{xe}^2) increases ν and lowers incentive pay. Both of these effects obtain because the purpose of incentive pay in the model is to discourage the CEO from shirking. From (11) and (14), we see that ν reflects the impact of cash flows, and the effect of shirking on cash flows, on beliefs $\hat{\mu}_{nt}$. When $\rho_{\mu e}^2$ is high, earnings receive relatively high weight when forming beliefs, and cash flows have relatively low weight. The firm must accordingly make incentives very sensitive to beliefs to make pay sufficiently sensitive to shirking. When ρ_{xe}^2 is high, cash flows receive relatively high weight when forming beliefs because earnings are somewhat redundant. Because beliefs are very sensitive to cash flows, lower incentives are needed to make pay sensitive to shirking.

Part 4 of Proposition 2 implies the following corollary:

Corollary 2.1. *If the firm chooses an optimal constant termination threshold, then the value $b_n(\hat{\mu}_{nt})$ is an increasing, convex function of beliefs $\hat{\mu}_{nt}$ and beliefs can be expressed as an increasing, concave function of b_n.*

Corollary 2.1 is useful because it provides a means to use observable market values to estimate unobservable beliefs about CEO ability.

Corollary 2.1 also implies that compensation will be a concave function of market values, which differs from the typical intuition that compensation is a convex function of market value. The reason is that compensation is linear in beliefs about CEO ability, and the termination option renders market values a convex function of beliefs about CEO ability. The model includes no opposing forces such as risk aversion or adverse selection that might lead to convex contracts (See, e.g., Hemmer et al. 1999; Beyer et al. 2014).

Proposition 2 also implies the following effects of ν on the termination threshold and information quality:

Corollary 2.2. *If the firm chooses an optimal constant termination threshold, then:*

1. *If the reservation value is sufficiently low ($\underline{w}_0 < \underline{w}(0, \underline{\mu})$), then firm value $b_n(\hat{\mu}_{nt}; \underline{\mu}_c)$ is increasing in ν for all $\hat{\mu}_{nt}$ and the termination threshold $\underline{\mu}_c$ is increasing in ν.*

2. *If the reservation value is sufficiently high* ($\underline{w}_0 > \underline{w}(0, \underline{\mu})$)*, then firm value* $b_n(\hat{\mu}_{nt}; \underline{\mu}_c)$ *is increasing (decreasing) in* ν *for high (low)* $\hat{\mu}_{nt}$ *and the termination threshold* $\underline{\mu}_c$ *is decreasing in* ν.

Corollary 2.2 implies that higher information quality, in the sense of higher ν, is associated with shorter CEO tenure when CEOs have low reservation utilities, and vice versa when CEOs have high reservation utilities. Specifically, the density of the time-to termination τ is:[2]

$$f\left(s, \frac{\hat{\mu}_{nt} - \underline{\mu}}{\sigma_\mu}\right) = \frac{\hat{\mu}_{nt} - \underline{\mu}}{\sigma_\mu} \frac{\hat{\mu}_{nt} - \underline{\mu}}{\sqrt{2\pi(s-t)^3}} \mathrm{e}^{-\left(\frac{\hat{\mu}_{nt} - \underline{\mu}}{\sigma_\mu}\right)^2 / 2(s-t)}, \tag{33}$$

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:

$$\mathrm{P}_t\left(\tau \leq T\right) = 2\Phi\left(-\frac{\hat{\mu}_{nt} - \underline{\mu}}{\sigma_\mu \sqrt{T-t}}\right). \tag{34}$$

This probability is increasing in $\underline{\mu}$ for all T. When the CEO reservation utility is high, firms must make upfront payments when hiring a new CEO, which creates incentives to delay termination.

The effects of information quality ν on the termination threshold $\underline{\mu}_c$ arise because of the role played in overall compensation to the CEO. Over the life of the contract, the firm and the new CEO expect pay valued at the greater of the reservation wage $\underline{w}_0$ or the value of payments from the firm $\underline{w}(0, \underline{\mu})$. The value of $\underline{w}(0, \underline{\mu})$ is decreasing in ν, making it relatively less expensive to hire a new CEO. This effect swamps the lower cash compensation that a higher ν allows the firm to pay the current CEO. When the reservation wage $\underline{w}_0$ is high, a higher ν has no marginal impact on the cost of obtaining a new CEO, so that the only effect of ν is to make it less expensive

[2] This is a somewhat standard result that one can derive applying a reverse Laplace transform to $\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right] = \mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt} - \underline{\mu}\right)/\sigma_\mu}$.

to compensate the current CEO. This leads to a lower termination threshold when CEOs have a high reservation utility.

While it may seem counterintuitive that higher information quality ν can reduce firm value, as in Part 2 of Corollary 2.2, the reason is straightforward given the effect of ν on the termination threshold. The value b_n of firm-specific cash flows reach their lowest point as $\hat{\mu}_{nt}$ approaches the termination threshold. A higher ν reduces the termination threshold, which means that the firm will allow the value to drift lower before terminating the CEO. This naturally implies that a higher ν will reduce b_n for small values of $\hat{\mu}_{nt}$. As $\hat{\mu}_{nt}$ increases, the effect of ν on reducing agency costs dominates and higher ν increases firm value.

4 Generating discrete data

To estimate the model, we will use data observed at discrete intervals. In this section, we derive the behavior of discrete observations of data from our continuous time model. The dynamics of the industry-level processes imply that we can express discrete changes as follows for a time increment of size δ:

$$\begin{aligned}
\mu_{0t} &= \mu_{0,t-\delta} + \delta_{0\mu t}, \\
x_{0t} &= x_{0,t-\delta} + \delta\mu_{0,t-\delta} + \delta_{0xt}, \\
e_{0t} &= \delta\left(1 - \tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\right)\mu_{0,t-\delta} + \left(1-\mathrm{e}^{-\theta\delta}\right)x_{0,t-\delta} + \mathrm{e}^{-\theta\delta}e_{0,t-\delta} + \delta_{0et}, \\
\hat{\mu}_{0t} &= \left(1-\mathrm{e}^{-\nu\delta}\right)\mu_{0,t-\delta} + \mathrm{e}^{-\nu\delta}\hat{\mu}_{0,t-\delta} + \delta_{0\hat{\mu}t},
\end{aligned} \tag{35}$$

where:

$$
\begin{aligned}
\delta_{0\mu t} &= \beta\sigma_\mu \int_{t-\delta}^{t} \mathrm{d}z_{0\mu s}, \\
\delta_{0xt} &= \beta \int_{t-\delta}^{t} \left((t-s)\,\sigma_\mu \mathrm{d}z_{0\mu s} + \sigma_x \mathrm{d}z_{0xs} \right), \\
\delta_{0et} &= \delta_{0xt} - \tfrac{1}{\theta}\delta_{0\mu t} + \beta\sigma_\eta \int_{t-\delta}^{t} \mathrm{e}^{-\theta(t-s)} \mathrm{d}z_{0\eta s}, \\
\delta_{0\hat{\mu}t} &= \beta\Big(\int_{t-\delta}^{t} \left(1-\mathrm{e}^{-\nu(t-s)}\right) \sigma_\mu \mathrm{d}z_{0\mu s} \\
&\quad + \int_{t-\delta}^{t} \mathrm{e}^{-\nu(t-s)} \left(\nu\sigma_x \mathrm{d}z_{0xs} + \left(\rho_{\mu e}\sigma_\mu - \nu\rho_{xe}\sigma_x\right) \mathrm{d}z_{0es} \right) \Big),
\end{aligned}
\tag{36}
$$

and the accrual shocks are:

$$
\begin{aligned}
\mathrm{d}z_{0\eta t} &= \tfrac{1}{\sigma_\eta} \left(\sigma_e \mathrm{d}z_{0et} - \sigma_x \mathrm{d}z_{0xt} + \tfrac{1}{\theta}\sigma_\mu \mathrm{d}z_{0\mu t} \right), \\
\sigma_\eta &= \sqrt{\sigma_e^2 + \sigma_x^2 + \tfrac{1}{\theta^2}\sigma_\mu^2 + 2\left(\tfrac{1}{\theta}\left(\rho_{\mu e}\sigma_e - \rho_{\mu x}\sigma_x\right)\sigma_\mu - \rho_{xe}\sigma_x\sigma_e \right)}.
\end{aligned}
\tag{37}
$$

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

$$
\begin{aligned}
\mu_{nt} &= 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n} \mu_{n,t-\delta} + 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n} \tilde{\mu}_{n,t-\delta} + \delta_{n\mu t}, \\
x_{nt} &= x_{n,t-\delta} + \delta \left(\mu_{0,t-\delta} + 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n} \mu_{n,t-\delta} + 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n} \tilde{\mu}_{n,t-\delta} \right) + \delta_{0xt} + \delta_{nxt}, \\
e_{nt} &= \delta \left(1 - \tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta} \right) \left(\mu_{0,t-\delta} + 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n} \mu_{n,t-\delta} + 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n} \tilde{\mu}_{n,t-\delta} \right) \\
&\quad + \left(1 - \mathrm{e}^{-\theta\delta}\right) x_{n,t-\delta} + \mathrm{e}^{-\theta\delta} e_{n,t-\delta} + \delta_{0et} + \delta_{net}, \\
\hat{\mu}_{nt} &= \left(1 - \mathrm{e}^{-\nu\delta}\right) \left(1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n} \mu_{n,t-\delta} + 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n} \tilde{\mu}_{n,t-\delta} \right) \\
&\quad + \mathrm{e}^{-\nu\delta} 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n} \hat{\mu}_{n,t-\delta} + \delta_{n\hat{\mu}t}, \\
\tilde{\mu}_{nt} &= \delta_{n\tilde{\mu}t},
\end{aligned}
\tag{38}
$$

where $\delta_{n\tilde{\mu}t}$ is a draw from a normal distribution with mean 0 and variance $\hat{\gamma}_n$, and:

$$
\begin{aligned}
\delta_{n\mu t} &= \sigma_\mu \int_{t-\delta}^{t} \mathrm{d}z_{n\mu s}, \\
\delta_{nxt} &= \int_{t-\delta}^{t} \left((t-s)\,\sigma_\mu \mathrm{d}z_{n\mu s} + \sigma_x \mathrm{d}z_{nxs}\right), \\
\delta_{net} &= \delta_{nxt} - \tfrac{1}{\theta}\delta_{n\mu t} + \sigma_\eta \int_{t-\delta}^{t} \mathrm{e}^{-\theta(t-s)} \mathrm{d}z_{n\eta s}, \\
\delta_{n\hat{\mu}t} &= \int_{t-\delta}^{t} \left(1-\mathrm{e}^{-\nu(t-s)}\right) \sigma_\mu \mathrm{d}z_{n\mu s} \\
&\quad + \int_{t-\delta}^{t} \mathrm{e}^{-\nu(t-s)} \left(\nu\sigma_x \mathrm{d}z_{nxs} + \left(\rho_{\mu e}\sigma_\mu - \nu\rho_{xe}\sigma_x\right) \mathrm{d}z_{nes}\right).
\end{aligned} \tag{39}
$$

We can write the system of processes as follows where $\boldsymbol{y}'_{nt} = \{\mu_{nt}, x_{nt}, e_{nt}, \hat{\mu}_{nt}, \tilde{\mu}_{nt}\}$, $\boldsymbol{y}'_t = \{\boldsymbol{y}'_{1t}, \boldsymbol{y}'_{2t}, \ldots, \boldsymbol{y}'_{Nt}\}$, $\boldsymbol{\delta}'_{nt} = \{\delta_{n\mu t}, \delta_{nxt}, \delta_{net}, \delta_{n\hat{\mu}t}, \delta_{n\tilde{\mu}t}\}$, and $\boldsymbol{\delta}'_t = \{\boldsymbol{\delta}'_{1t}, \boldsymbol{\delta}'_{2t}, \ldots, \boldsymbol{\delta}'_{Nt}\}$:

$$
\begin{aligned}
\begin{pmatrix} \boldsymbol{y}_{0t} \\ \boldsymbol{y}_t \end{pmatrix} &= \begin{pmatrix} \boldsymbol{A}_{00} & \boldsymbol{1}'_N \otimes \boldsymbol{0}_{4\times 5} \\ \boldsymbol{1}_N \otimes \boldsymbol{A}_{N0} & \boldsymbol{I}_N \otimes \boldsymbol{A}_{NN} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_{0,t-\delta} \\ \boldsymbol{y}_{t-\delta} \end{pmatrix} \\
&\quad + \begin{pmatrix} \boldsymbol{I}_4 & \boldsymbol{1}'_N \otimes \boldsymbol{0}_{4\times 5} \\ \boldsymbol{1}_N \otimes \boldsymbol{D}_{0N} & \boldsymbol{I}_N \otimes \boldsymbol{I}_5 \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{0t} \\ \boldsymbol{\delta}_t \end{pmatrix},
\end{aligned} \tag{40}
$$

where:

$$
\begin{aligned}
\boldsymbol{A}_{00} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ \delta & 1 & 0 & 0 \\ \delta\left(1-\frac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\right) & 1-\mathrm{e}^{-\theta\delta} & \mathrm{e}^{-\theta\delta} & 0 \\ 1-\mathrm{e}^{-\nu\delta} & 0 & 0 & \mathrm{e}^{-\nu\delta} \end{pmatrix}, \\
\boldsymbol{A}_{N0} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ \delta & 0 & 0 & 0 \\ \delta\left(1-\frac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
\boldsymbol{A}_{NN} &= \begin{pmatrix} 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n} & 0 & 0 & 0 & 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n} \\ 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n}\delta & 1 & 0 & 0 & 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n}\delta \\ 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n}\delta\left(1-\frac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\right) & 1-\mathrm{e}^{-\theta\delta} & \mathrm{e}^{-\theta\delta} & 0 & 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n}\delta\left(1-\frac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\right) \\ 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n}\left(1-\mathrm{e}^{-\nu\delta}\right) & 0 & 0 & 1_{\hat{\mu}_{n,t-\delta} > \underline{\mu}_n}\mathrm{e}^{-\nu\delta} & 1_{\hat{\mu}_{n,t-\delta} \le \underline{\mu}_n}\left(1-\mathrm{e}^{-\nu\delta}\right) \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \\
\boldsymbol{D}_{0N} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned} \tag{41}
$$

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

$$
\begin{pmatrix} \boldsymbol{\delta}_{0t} \\ \boldsymbol{\delta}_t \end{pmatrix} = \boldsymbol{C}_\delta \boldsymbol{b}_t, \tag{42}
$$

where $\boldsymbol{b}_t$ is a $4(N+1)$ vector of standard normal random variables and $\boldsymbol{C}_\delta$ is from the Cholesky decomposition of the covariance matrix of the shock vector:

$$\begin{aligned} \boldsymbol{C}_\delta \boldsymbol{C}_\delta' &= \mathrm{E}\left[\begin{pmatrix} \boldsymbol{\delta}_{0t} \\ \boldsymbol{\delta}_t \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{0t} & \boldsymbol{\delta}_t' \end{pmatrix}\right] = \begin{pmatrix} \beta^2 \boldsymbol{\Sigma}_{nn} & \mathbf{1}_N' \otimes \mathbf{0}_{4\times 5} \\ \mathbf{1}_N \otimes \mathbf{0}_{5\times 4} & \boldsymbol{I}_N \otimes \begin{pmatrix} \boldsymbol{\Sigma}_{nn} & \mathbf{0}_4 \\ \mathbf{0}_4' & \hat{\gamma}_n \end{pmatrix} \end{pmatrix}, \\ \boldsymbol{\Sigma}_{nn} &= \mathrm{E}\left[\boldsymbol{\delta}_{nt} \boldsymbol{\delta}_{nt}'\right] = \begin{pmatrix} \sigma_{\mu\mu} & \sigma_{\mu x} & \sigma_{\mu e} & \sigma_{\mu\hat{\mu}} \\ \sigma_{\mu x} & \sigma_{xx} & \sigma_{xe} & \sigma_{\hat{\mu}x} \\ \sigma_{\mu e} & \sigma_{xe} & \sigma_{ee} & \sigma_{\hat{\mu}e} \\ \sigma_{\mu\hat{\mu}} & \sigma_{\hat{\mu}x} & \sigma_{\hat{\mu}e} & \sigma_{\hat{\mu}\hat{\mu}} \end{pmatrix}. \end{aligned} \tag{43}$$

The elements of $\boldsymbol{\Sigma}_{nn}$ are:

$$\begin{aligned} \sigma_{\mu\mu} &= \delta\sigma_\mu^2, \\ \sigma_{xx} &= \delta\left(\sigma_x^2 + \delta\rho_{\mu x}\sigma_\mu\sigma_x + \tfrac{\delta}{3}\sigma_{\mu\mu}\right), \\ \sigma_{ee} &= \sigma_{xx} + \tfrac{1}{\theta^2}\sigma_{\mu\mu} - 2\tfrac{1}{\theta}\sigma_{\mu x} \\ &\quad + \delta\left(\tfrac{1-\mathrm{e}^{-2\theta\delta}}{2\theta\delta}\sigma_\eta^2 + 2\left(\tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\rho_{x\eta}\sigma_x - \tfrac{\mathrm{e}^{-\theta\delta}}{\theta}\rho_{\mu\eta}\sigma_\mu\right)\sigma_\eta\right), \\ \sigma_{\hat{\mu}\hat{\mu}} &= \sigma_{\mu\mu}\left(1 - \tfrac{\left(1-\mathrm{e}^{-\nu\delta}\right)^2}{\nu\delta}\left(1-\rho_{\mu\hat{\mu}}\right)\right), \\ \sigma_{\mu x} &= \delta\left(\tfrac{1}{2}\sigma_{\mu\mu} + \rho_{\mu x}\sigma_\mu\sigma_x\right), \\ \sigma_{\mu e} &= \sigma_{\mu x} - \tfrac{1}{\theta}\sigma_{\mu\mu} + \delta\tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\rho_{\mu\eta}\sigma_\mu\sigma_\eta, \\ \sigma_{\mu\hat{\mu}} &= \sigma_{\mu\mu}\left(1 - \tfrac{1-\mathrm{e}^{-\nu\delta}}{\nu\delta}\left(1-\rho_{\mu\hat{\mu}}\right)\right), \\ \sigma_{xe} &= \sigma_{xx} - \tfrac{1}{\theta}\sigma_{\mu x} + \delta\left(\tfrac{1}{\theta}\left(\tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta} - \mathrm{e}^{-\theta\delta}\right)\rho_{\mu\eta}\sigma_\mu\sigma_\eta + \tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\rho_{x\eta}\sigma_x\sigma_\eta\right), \\ \sigma_{\hat{\mu}x} &= \delta\left(\tfrac{1}{2}\sigma_{\mu\mu} + \mathrm{e}^{-\nu\delta}\rho_{\hat{\mu}x}\sigma_\mu\sigma_x + \left(1-\mathrm{e}^{-\nu\delta}\right)\rho_{\mu x}\sigma_\mu\sigma_x\right), \\ \sigma_{\hat{\mu}e} &= \sigma_{\hat{\mu}x} - \tfrac{1}{\theta}\sigma_{\mu\hat{\mu}} + \delta\left(\tfrac{1-\mathrm{e}^{-\theta\delta}}{\theta\delta}\rho_{\mu\eta} + \tfrac{1-\mathrm{e}^{-(\theta+\nu)\delta}}{(\theta+\nu)\delta}\left(\rho_{\hat{\mu}\eta} - \rho_{\mu\eta}\right)\right)\sigma_\mu\sigma_\eta, \end{aligned} \tag{44}$$

where:

$$\begin{aligned} \rho_{\mu\eta} &= \tfrac{1}{\sigma_\eta}\left(\rho_{\mu e}\sigma_e - \rho_{\mu x}\sigma_x + \tfrac{1}{\theta}\sigma_\mu\right) \\ \rho_{x\eta} &= \tfrac{1}{\sigma_\eta}\left(\rho_{xe}\sigma_e - \sigma_x + \tfrac{1}{\theta}\rho_{\mu x}\sigma_\mu\right) \\ \rho_{e\eta} &= \tfrac{1}{\sigma_\eta}\left(\sigma_e - \rho_{xe}\sigma_x + \tfrac{1}{\theta}\rho_{\mu e}\sigma_\mu\right) \\ \rho_{\mu\hat{\mu}} &= \rho_{\mu e}^2 + \left(\rho_{\mu x} - \rho_{\mu e}\rho_{xe}\right)\sqrt{\tfrac{1-\rho_{\mu e}^2}{1-\rho_{xe}^2}}, \\ \rho_{\hat{\mu}x} &= \sqrt{\left(1-\rho_{\mu e}^2\right)\left(1-\rho_{xe}^2\right)} + \rho_{\mu e}\rho_{xe}, \\ \rho_{\hat{\mu}\eta} &= \tfrac{1}{\sigma_\eta}\left(\rho_{\mu e}\sigma_e - \rho_{\hat{\mu}x}\sigma_x + \tfrac{1}{\theta}\rho_{\mu\hat{\mu}}\sigma_\mu\right). \end{aligned} \tag{45}$$

In future work, we will use the above processes to simulate data and estimate the model using simulated method of moments.

5 Conclusion and future work

This paper develops a continuous-time, dynamic contracting model that we will use to quantify the effects of accounting on CEO turnover and firm value. Within the context of our model, accounting quality depends solely on the ability to use earnings to filter out transitory shocks to cash flows. Filtering out these shocks reduces the cost of inducing CEO effort.

The repeated termination and hiring in our model introduces issues and complexities that are absent in models where there is only one employment period. Specifically, the firm will never prefer the surplus-maximizing contract even if there are no agency conflicts. The reason for this is that the firm always has an incentive to reduce the overall surplus in order to reduce the value it must share with current and future agents – a problem that it cannot resolve with a contract with the current agent.

In future work, we will derive the fully optimal contract without restricting to constant termination policies, and will structurally estimate the model's parameters.

References

Beyer, A., I. Guttman, and I. Marinovic. 2014. Optimal contracts with performance manipulation. *Journal of Accounting Research* 52(4): 817–847.

DeMarzo, P., and Y. Sannikov. 2017. Learning, termination, and payout policy in dynamic incentive contracts. *Review of Economic Studies* 84(1): 182–236.

Hemmer, T., O. Kim, and R. Verrecchia. 1999. Introducing convexity into optimal compensation contracts. *Journal of Accounting and Economics* 28(3): 307–327.

Liptser, R., and A. Shiryaev. 2001. *Statistics of Random Processes I. General Theory.* Springer-Verlag Berlin Heidelberg 2nd edition.

Øksendal, B. 2003. *Stochastic Differential Equations.* Springer Berlin, Germany 6 edition.

Williams, N. 2011. Persistent private information. *Econometrica* 79(4): 1233–1275.

Yong, J., and X. Zhou. 1999. *Stochastic Controls: Hamiltonian Systems and HJB Equations.* Springer-Verlag New York, USA.

Appendix

The filtering problem (Section 2.2)

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

$$\begin{pmatrix} \mathrm{d}\boldsymbol{z}_{\mu t} \\ \mathrm{d}\boldsymbol{z}_{yt} \end{pmatrix} = \boldsymbol{C}_z \mathrm{d}\boldsymbol{z}_t. \tag{A1}$$

The matrix $\boldsymbol{C}_z$ is from the Cholesky decomposition of the covariance matrix of $\mathrm{d}\boldsymbol{z}_{\mu t}$ and $\mathrm{d}\boldsymbol{z}_{yt}$:

$$\boldsymbol{C}_z \boldsymbol{C}'_z = \frac{1}{\mathrm{d}t} \mathrm{E}\left[\begin{pmatrix} \mathrm{d}\boldsymbol{z}_{\mu t} \\ \mathrm{d}\boldsymbol{z}_{yt} \end{pmatrix} \begin{pmatrix} \mathrm{d}\boldsymbol{z}'_{\mu t} & \mathrm{d}\boldsymbol{z}'_{yt} \end{pmatrix} \right] = \begin{pmatrix} \boldsymbol{I}_2 & \boldsymbol{I}_2 \otimes \boldsymbol{r}'_{\mu y} \\ \boldsymbol{I}_2 \otimes \boldsymbol{r}_{\mu y} & \boldsymbol{I}_2 \otimes \boldsymbol{R}_{xe} \end{pmatrix}, \tag{A2}$$

where $\boldsymbol{r}'_{\mu y} = \{\rho_{\mu x}, \rho_{\mu e}\}$ and:

$$\boldsymbol{R}_{xe} = \begin{pmatrix} 1 & \rho_{xe} \\ \rho_{xe} & 1 \end{pmatrix}. \tag{A3}$$

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

$$\rho_{\mu x}, \rho_{\mu e}, \rho_{xe} \in [-1, 1], \qquad \rho_{xe} \in \rho_{\mu e}\rho_{\mu x} \pm \sqrt{\left(1 - \rho_{\mu e}^2\right)\left(1 - \rho_{\mu x}^2\right)} \subseteq [-1, 1].$$

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:

$$\begin{aligned} \mathrm{d}\boldsymbol{\mu}_t &= \underbrace{\left(\boldsymbol{\Sigma}_\mu \ \mathbf{0}_{2\times 4} \right) \boldsymbol{C}_z}_{\boldsymbol{B}_\mu} \mathrm{d}\boldsymbol{z}_t \\ \mathrm{d}\boldsymbol{y}_t &= \underbrace{\left(\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right) \otimes \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \right)}_{\boldsymbol{A}_\mu} \boldsymbol{\mu}_t \mathrm{d}t - \boldsymbol{a}_t \mathrm{d}t + \underbrace{\left(\boldsymbol{I}_2 \otimes \left(\begin{smallmatrix} 0 & 0 \\ \theta & -\theta \end{smallmatrix}\right) \right)}_{\boldsymbol{A}_y} \boldsymbol{y}_t \mathrm{d}t + \underbrace{\left(\mathbf{0}_{4\times 2} \ \boldsymbol{C}_y \otimes \boldsymbol{\Sigma}_{xe} \right) \boldsymbol{C}_z}_{\boldsymbol{B}_y} \mathrm{d}\boldsymbol{z}_t, \end{aligned} \tag{A4}$$

where:

$$\boldsymbol{C}_y = \left(\begin{smallmatrix} \beta & 0 \\ \beta & 1 \end{smallmatrix}\right), \qquad \boldsymbol{\Sigma}_{xe} = \left(\begin{smallmatrix} \sigma_x & 0 \\ 0 & \sigma_e \end{smallmatrix}\right). \tag{A5}$$

The filtering problem gives profitability estimates $\hat{\boldsymbol{\mu}}_t' = \{\hat{\mu}_{0t}, \hat{\mu}_{nt}\}$ with covariance matrix $\hat{\boldsymbol{\Gamma}}_t = \mathrm{var}_t(\boldsymbol{\mu}_t)$ and the following dynamics:

$$\begin{aligned} \mathrm{d}\hat{\boldsymbol{\mu}}_t &= \boldsymbol{K}_t \mathrm{d}\boldsymbol{n}_t, \\ \mathrm{d}\boldsymbol{n}_t &= \mathrm{d}\boldsymbol{y}_t - \left(\boldsymbol{A}_\mu \hat{\boldsymbol{\mu}}_t - \boldsymbol{a}_t \mathrm{d}t + \boldsymbol{A}_y \boldsymbol{y}_t \right) \mathrm{d}t, \\ \boldsymbol{K}_t &= \left(\boldsymbol{B}_\mu \boldsymbol{B}_y' + \hat{\boldsymbol{\Gamma}}_t \boldsymbol{A}_\mu' \right) \left(\boldsymbol{B}_y \boldsymbol{B}_y' \right)^{-1}, \\ \mathrm{d}\hat{\boldsymbol{\Gamma}}_t &= \left(\boldsymbol{B}_\mu \boldsymbol{B}_\mu' - \boldsymbol{K}_t \left(\boldsymbol{B}_y \boldsymbol{B}_\mu' + \boldsymbol{A}_\mu \hat{\boldsymbol{\Gamma}}_t \right) \right) \mathrm{d}t. \end{aligned} \tag{A6}$$

In a steady state, $\mathrm{d}\hat{\boldsymbol{\Gamma}}_t = \mathbf{0}_{4\times 4}$, which implies steady state posteriors of:[3]

$$\begin{aligned} \hat{\gamma}_n &= \mathrm{var}_\infty(\mu_{nt}) = \sigma_\mu \sigma_x \left(\sqrt{\left(1 - \rho_{\mu e}^2\right)\left(1 - \rho_{xe}^2\right)} - \left(\rho_{\mu x} - \rho_{\mu e}\rho_{xe} \right) \right), \\ \hat{\gamma}_0 &= \mathrm{var}_\infty(\mu_{0t}) = \beta^2 \hat{\gamma}_n. \end{aligned} \tag{A7}$$

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

$$\boldsymbol{K}_\infty = \left(\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right) \otimes \left(\nu \ \ \frac{\rho_{\mu e}\sigma_\mu - \nu \rho_{xe}\sigma_x}{\sigma_e} \right), \tag{A8}$$

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

[3] The other solutions to the steady state either imply negative variances or that the posterior covariance $\mathrm{cov}_\infty(\mu_{0t}, \mu_{nt})$ has a nonzero imaginary component.

$\{\mathrm{d}\hat{z}_{0\mu t}, \mathrm{d}\hat{z}_{0xt}, \mathrm{d}z_{0et}, \mathrm{d}\hat{z}_{n\mu t}, \mathrm{d}\hat{z}_{nxt}, \mathrm{d}z_{net}\}$ has the correlation matrix:

$$\boldsymbol{I}_2 \otimes \begin{pmatrix} 1 & \rho_{\hat{\mu}x} & \rho_{\mu e} \\ \rho_{\hat{\mu}x} & 1 & \rho_{xe} \\ \rho_{\mu e} & \rho_{xe} & 1 \end{pmatrix}, \tag{A9}$$

where:

$$\rho_{\hat{\mu}x} = \frac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}\hat{z}_{n\mu t}\mathrm{d}\hat{z}_{nxt}\right] = \frac{1}{\mathrm{d}t}\mathrm{E}\left[\mathrm{d}\hat{z}_{n\mu t}\mathrm{d}z_{nxt}\right] = \rho_{\mu e}\rho_{xe} + \sqrt{\left(1-\rho_{\mu e}^2\right)\left(1-\rho_{xe}^2\right)}. \tag{A10}$$

■

Supporting results

Lemmas A1 and A2 are similar to results in DeMarzo and Sannikov (2017), and we show that they hold in our setting.

Lemma A1. *In an optimal, incentive-compatible contract,* $a_t = 0$ *and it is unnecessary to include the possibility of early termination. Specifically, for any incentive-compatible contract* (c_t, τ) *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, e_t). Now take an adjusted process $\tilde{x}_t = x_t - \int_0^t a_s \mathrm{d}s$ and denote the filtration $\tilde{\mathcal{F}}_t$ as the one generated by $(\tilde{x}_t, e_t)$. The original contract has payoff $c_t(\mathcal{F}_t)$ and stopping time $\tau(\mathcal{F}_t)$. Now consider a new contract:

$$\tilde{c}_t = c_t(\tilde{\mathcal{F}}_t) + \lambda a_t, \qquad \tilde{\tau} = \tau(\tilde{\mathcal{F}}_t). \tag{A11}$$

Denote by $\tilde{a}_t$ the diversion under the new contract. Then $\mathrm{d}\tilde{x}_t = \mathrm{d}x_t - a_t\mathrm{d}t = (\hat{\mu}_t - a_t - \tilde{a}_t)\mathrm{d}t + \sigma_x\mathrm{d}\hat{z}_{xt}$ 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 λa_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:

$$\begin{aligned} \mathrm{E}_t \left[\int_t^{\tilde{\tau}} \mathrm{e}^{-r(s-t)} \left(\hat{\mu}_{ns} - \tilde{a}_s - \tilde{c}_s \right) \mathrm{d}s + \mathrm{e}^{-r(\tilde{\tau}-t)} \left(b_n \left(0, w_0\right) - k \right) \right] & \\ = \mathrm{E}_t \left[\int_t^{\tau} \mathrm{e}^{-r(s-t)} \left(\hat{\mu}_{ns} - \lambda a_s - c_s \right) \mathrm{d}s + \mathrm{e}^{-r(\tau-t)} \left(b_n \left(0, w_0\right) - k \right) \right] & \\ \geq \mathrm{E}_t \left[\int_t^{\tau} \mathrm{e}^{-r(s-t)} \left(\hat{\mu}_{ns} - a_s - c_s \right) \mathrm{d}s + \mathrm{e}^{-r(\tau-t)} \left(b_n \left(0, w_0\right) - k \right) \right] & \quad \text{(A12)} \end{aligned}$$

so that the principal is better off. This shows that $a_t = 0$ in the optimal contract. ■

Lemma A2. *The following hold for any incentive-compatible contract* (c_t, τ)*:*

1. *The agent's continuation value* w_t *has the representation:*

$$\mathrm{d}w_t = \left(rw_t - c_t\right) \mathrm{d}t + \beta_{\mu t} \mathrm{d}\hat{\mu}_{nt} + \beta_{et} \mathrm{d}\tilde{e}_{nt}, \tag{A13}$$

where $\mathrm{d}\tilde{e}_{nt} = \frac{\sqrt{1-\rho_{\mu e}^2}}{\rho_{\mu e}} \sigma_\mu \mathrm{d}\tilde{z}_{net}$.

2. *The information rent has the representation:*

$$\xi_t = \mathrm{E}_t \left[\int_t^{\tau} \mathrm{e}^{-(r+\nu)(s-t)} \nu \left(\beta_{\mu s} - \beta_{es} \right) \mathrm{d}s + \mathrm{e}^{-(r+\nu)(\tau-t)} \frac{\lambda - \psi}{r} \right], \tag{A14}$$

with the following dynamics for some process $\left(\chi_{\mu t}, \chi_{et}\right)$*:*

$$\mathrm{d}\xi_t = \left(\left(r + \nu\right) \xi_t - \nu \left(\beta_{\mu t} - \beta_{et} \right) \right) \mathrm{d}t + \chi_{\mu t} \mathrm{d}\hat{z}_{n\mu t} + \chi_{et} \mathrm{d}\tilde{z}_{net}. \tag{A15}$$

3. *The IC constraint* $\beta_{\mu t} \geq \frac{\lambda}{\nu} + \xi_t + \beta_{et}$ *is necessary for the optimality of* $a_t = 0$.

4. *The following lower bound for information rent holds with equality if IC con-*

straint is binding at all future dates:

$$\xi_t \geq \zeta_t = \tfrac{1}{r}\left(\lambda - \psi \mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right]\right). \tag{A16}$$

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

$$\begin{pmatrix} \mathrm{d}x_{nt} \\ \mathrm{d}e_{nt} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_{0t}+\hat{\mu}_{nt} \\ \theta(x_{nt}-e_{nt}) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \beta\sigma_x \mathrm{d}\hat{z}_{0xt} \\ \beta\sigma_e \mathrm{d}z_{0et} \end{pmatrix} + \tilde{\boldsymbol{C}}\mathrm{d}\tilde{\boldsymbol{z}}_t, \tag{A17}$$

where $\tilde{\boldsymbol{z}}_t$ is a two-dimensional standard Brownian motion with:

$$\begin{pmatrix} \sigma_x \mathrm{d}\hat{z}_{nxt} \\ \sigma_e \mathrm{d}z_{net} \end{pmatrix} = \underbrace{\begin{pmatrix} \rho_{\hat{\mu}x}\sigma_x & \frac{\rho_{xe}}{\sqrt{1-\rho_{xe}^2}}\left(\rho_{\hat{\mu}x} - \frac{\rho_{\mu e}}{\rho_{xe}}\right)\sigma_x \\ \rho_{\mu e}\sigma_e & \sigma_e\sqrt{1-\rho_{\mu e}^2} \end{pmatrix}}_{\tilde{\boldsymbol{C}}} \underbrace{\begin{pmatrix} \mathrm{d}\hat{z}_{n\mu t} \\ \mathrm{d}\tilde{z}_{net} \end{pmatrix}}_{\mathrm{d}\tilde{\boldsymbol{z}}_{nt}}. \tag{A18}$$

The vector $\tilde{\boldsymbol{z}}_{nt}$ 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 of $\{x_{nt}, e_{nt}\}$ are as follows where $\alpha_t = \hat{\mu}_{nt}^a - \hat{\mu}_{nt}$:

$$\begin{pmatrix} \mathrm{d}x_{nt} \\ \mathrm{d}e_{nt} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_{0t}+\hat{\mu}_{nt}+\alpha_t-\hat{a}_t \\ \theta(x_{nt}-e_{nt}) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \beta\sigma_x \mathrm{d}\hat{z}_{0xt} \\ \beta\sigma_e \mathrm{d}z_{0et} \end{pmatrix} + \tilde{\boldsymbol{C}}\mathrm{d}\tilde{\boldsymbol{z}}_{nt}^a, \tag{A19}$$

and:

$$\mathrm{d}\tilde{\boldsymbol{z}}_{nt}^a = \begin{pmatrix} \mathrm{d}\hat{z}_{n\mu t}^a \\ \mathrm{d}\tilde{z}_{net}^a \end{pmatrix} = \begin{pmatrix} \mathrm{d}\hat{z}_{n\mu t} + \frac{\nu}{\sigma_\mu}(\hat{a}_{nt}-\alpha_t)\mathrm{d}t \\ \frac{1}{\sqrt{1-\rho_{\mu e}^2}}\left(\mathrm{d}z_{net} - \rho_{\mu e}\mathrm{d}\hat{z}_{n\mu t}^a\right) \end{pmatrix}. \tag{A20}$$

The process $\tilde{\boldsymbol{z}}_{nt}^a$ is a Brownian motion with respect to the beliefs $\hat{P}$ generated by the actions $\hat{a}_t$, with $\mathrm{d}\alpha_t = \nu\left(\hat{a}_t - \alpha_t\right)\mathrm{d}t$ 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). In other words, $\mathrm{d}\hat{P}_t = \phi_t \mathrm{d}P_t$ and the agent's payoff can be written as:

$$\mathrm{E}_{\hat{P}}\left[\int_0^\tau \mathrm{e}^{-rt}\left(\lambda \hat{a}_t + c_t\right)\mathrm{d}t + \mathrm{e}^{-r\tau} R\right] = \mathrm{E}_P\left[\int_0^\tau \phi_t \mathrm{e}^{-rt}\left(\lambda \hat{a}_t + c_t\right)\mathrm{d}t + \phi_\tau \mathrm{e}^{-r\tau} R\right]. \quad \text{(A21)}$$

The process ϕ_t satisfies:

$$\phi_t = \exp\left\{-\int_0^t \boldsymbol{u}_s' \mathrm{d}\tilde{\boldsymbol{z}}_{ns} - \tfrac{1}{2}\int_0^t \boldsymbol{u}_s' \boldsymbol{u}_s \mathrm{d}s\right\}, \quad \mathrm{d}\phi_t = -\phi_t \boldsymbol{u}_t' \mathrm{d}\tilde{\boldsymbol{z}}_{nt}, \quad \text{(A22)}$$

with $\mathrm{E}_P[\phi_t] = \phi_0 = 1$, where the process $\boldsymbol{u}_t$ solves:

$$\tilde{\boldsymbol{C}}\boldsymbol{u}_t = \begin{pmatrix} \hat{\mu}_{nt} \\ \theta(x_{nt}-e_{nt}) \end{pmatrix} - \begin{pmatrix} \hat{\mu}^a_{nt}-\hat{a}_t \\ \theta(x_{nt}-e_{nt}) \end{pmatrix} = -\begin{pmatrix} \alpha_t-\hat{a}_t \\ 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{u}_t = -\tilde{\boldsymbol{C}}^{-1}\begin{pmatrix} \alpha_t-\hat{a}_t \\ 0 \end{pmatrix} = \frac{\nu(\alpha_t-\hat{a}_t)}{\sigma_\mu}\begin{pmatrix} -1 \\ \frac{\rho_{\mu e}}{\sqrt{1-\rho^2_{\mu e}}} \end{pmatrix}. \quad \text{(A23)}$$

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 $\boldsymbol{y}_t = (\phi_t, \alpha_t)$, with the following dynamics:

$$\underbrace{\begin{pmatrix} \mathrm{d}\phi_t \\ \mathrm{d}\alpha_t \end{pmatrix}}_{\mathrm{d}\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} 0 \\ \nu(\hat{a}_t-\alpha_t) \end{pmatrix}}_{\boldsymbol{b}_{yt}} \mathrm{d}t + \underbrace{\phi_t \frac{\nu(\alpha_t-\hat{a}_t)}{\sigma_\mu}\begin{pmatrix} 1 & -\frac{\rho_{\mu e}}{\sqrt{1-\rho^2_{\mu e}}} \\ 0 & 0 \end{pmatrix}}_{\boldsymbol{\Sigma}_{yt}} \mathrm{d}\boldsymbol{z}_{nt}. \quad \text{(A24)}$$

Denote the costate variables on the drifts by $\boldsymbol{p}_t = (p_{\phi t}, p_{\alpha t})$, and their diffusion coefficients by the matrix $\boldsymbol{Q}_t = \begin{pmatrix} q_{\phi\mu t} & q_{\phi e t} \\ q_{\alpha\mu t} & q_{\alpha e t} \end{pmatrix}$ (i.e., the diffusion term of $\mathrm{d}\boldsymbol{p}_t$ is $\boldsymbol{Q}_t \mathrm{d}\tilde{\boldsymbol{z}}_{nt}$).

The (current value) Hamiltonian is then:

$$H(t,\boldsymbol{y}_t,\hat{a}_t,\boldsymbol{p}_t,\boldsymbol{q}_t)=\phi_t\left(\lambda\hat{a}_t+c_t\right)+\underbrace{p_{\alpha t}\nu\left(\hat{a}_t-\alpha_t\right)}_{\boldsymbol{b}_{yt}'\boldsymbol{p}_t}+\underbrace{\left(q_{\phi\mu t}-q_{\phi et}\frac{\rho_{\mu e}}{\sqrt{1-\rho_{\mu e}^2}}\right)\phi_t\frac{\nu(\alpha_t-\hat{a}_t)}{\sigma_\mu}}_{\mathrm{tr}(\boldsymbol{Q}_t'\boldsymbol{\Sigma}_{yt})},\tag{A25}$$

where $p_{\phi t}$ does not appear because ϕ_t has zero drift and the $\boldsymbol{Q}_t$ terms that correspond to α_t do not appear because α_t has zero volatility. Differentiating with respect to $\hat{a}_t$ gives:

$$\frac{\partial H}{\partial\hat{a}_t}=\phi_t\lambda+p_{\alpha t}\nu-\frac{\nu}{\sigma_\mu}\left(q_{\phi\mu t}-q_{\phi et}\frac{\rho_{\mu e}}{\sqrt{1-\rho_{\mu e}^2}}\right)\phi_t,\tag{A26}$$

which must be weakly negative in order for $\hat{a}_t=0$ to be optimal, given the restriction $\hat{a}_t\geq 0$. The costate variables on the drifts evolve as follows, where the $r\boldsymbol{p}_t$ term accounts for discounting, and $\hat{a}_t=0$ for all t implies the second equality with $\alpha_t=0$, $\phi_t=1$, and $\mathrm{d}\tilde{\boldsymbol{z}}_{nt}=\mathrm{d}\tilde{\boldsymbol{z}}_{nt}^a$:

$$\begin{aligned}\underbrace{\begin{pmatrix}\mathrm{d}p_{\phi t}\\ \mathrm{d}p_{\alpha t}\end{pmatrix}}_{\mathrm{d}\boldsymbol{p}_t}&=\left(r\boldsymbol{p}_t-\mathrm{D}_{\boldsymbol{y}}H\big|_{\alpha_t=\hat{a}_t=0}\right)\mathrm{d}t+\boldsymbol{Q}_t\mathrm{d}\tilde{\boldsymbol{z}}_t\\ &=\begin{pmatrix}rp_{\phi t}-c_t\\ (r+\nu)p_{\alpha t}-\frac{\nu}{\sigma_\mu}\left(q_{\phi\mu t}-q_{\phi et}\frac{\rho_{\mu e}}{\sqrt{1-\rho_{\mu e}^2}}\right)\end{pmatrix}\mathrm{d}t+\boldsymbol{Q}_t\mathrm{d}\tilde{\boldsymbol{z}}_{nt}.\end{aligned}\tag{A27}$$

The boundary conditions are:

$$\begin{aligned}p_{\phi\tau}&=\frac{\partial}{\partial\phi}\phi_\tau R\Big|_{\hat{a}_\tau=0}&=R(\hat{\mu}_{n\tau}),\\ p_{\alpha\tau}&=\frac{\partial}{\partial\alpha}\phi_\tau R\Big|_{\alpha_\tau=0}&=\frac{\lambda-\psi}{r}.\end{aligned}\tag{A28}$$

The process $p_{\phi t}$ is given by the agent's expected payoff from continuing w_t. To see this, conjecture that:

$$p_{\phi t}=w_t=\mathrm{E}_t\left[\int_t^\tau \mathrm{e}^{-r(s-t)}c_s\mathrm{d}s+\mathrm{e}^{-r(\tau-t)}R(\hat{\mu}_{n\tau})\right],\tag{A29}$$

and put:

$$\hat{p}_{\phi t} = \mathrm{E}_t\left[\int_0^{\tau} \mathrm{e}^{-r(s-t)} c_s \mathrm{d}s + \mathrm{e}^{-r(\tau-t)} R(\hat{\mu}_{n\tau})\right] = \int_0^{t} \mathrm{e}^{-rs} c_s \mathrm{d}s + \mathrm{e}^{-rt} p_{\phi t}. \tag{A30}$$

Because $\hat{p}_{\phi t}$ is a martingale, we have:

$$0 = \underbrace{\mathrm{e}^{-rt}\left(\mathrm{E}[\mathrm{d}p_{\phi t}] - (r p_{\phi t} - c_t)\,\mathrm{d}t\right)}_{\mathrm{E}\left[\mathrm{d}\hat{p}_{\phi t}\right]}, \tag{A31}$$

which matches the drift in (A27). We can then write the dynamics of the agent's continuation value as in (A13) where $\beta_{\mu t} = \frac{1}{\sigma_\mu} q_{\phi \mu t}$, $\beta_{et} = \frac{1}{\sigma_\mu} \frac{\rho_{\mu e}}{\sqrt{1-\rho_{\mu e}^2}} q_{\phi e t}$, and $\mathrm{d}\tilde{e}_{nt} = \frac{\sqrt{1-\rho_{\mu e}^2}}{\rho_{\mu e}} \sigma_\mu \mathrm{d}\tilde{z}_{net}$.

The process $p_{\alpha t}$ is the change in payoff with respect to the belief discrepancy α_t, so it is the agent's information rent ξ_t. Can write dynamics of $p_{\alpha t}$ as follows when $\phi_t = 1$:

$$\mathrm{d}p_{\alpha t} = \left((r+\nu)\, p_{\alpha t} - \nu\left(\beta_{\mu t} - \beta_{et}\right)\right)\mathrm{d}t + q_{\alpha \mu t}\mathrm{d}\hat{z}_{n\mu t} + q_{\alpha e t}\mathrm{d}\tilde{z}_{net}. \tag{A32}$$

Conjecture that $p_{\alpha t}$ equals ξ_t in expression (A14). The process:

$$\hat{\xi}_t = \mathrm{E}_{Pt}\left[\int_0^{\tau} \mathrm{e}^{-(r+\nu)s} \nu\left(\beta_{\mu s} - \beta_{es}\right)\mathrm{d}s + \mathrm{e}^{-(r+\nu)\tau} \frac{\lambda-\psi}{r}\right] = \int_0^{t} \mathrm{e}^{-(r+\nu)s} \nu\left(\beta_{\mu s} - \beta_{es}\right)\mathrm{d}s + \mathrm{e}^{-(r+\nu)t} \xi_t, \tag{A33}$$

is a martingale, implying:

$$0 = \underbrace{-\mathrm{e}^{-(r+\nu)t}\left(\left((r+\nu)\xi_t - \nu\left(\beta_{\mu t} - \beta_{et}\right)\right)\mathrm{d}t - \mathrm{E}\left[\mathrm{d}\xi_t\right]\right)}_{\mathrm{E}[\mathrm{d}\hat{\xi}_t]}, \tag{A34}$$

so that the drift matches (A32). We can write the dynamics of the information rent process as in (A15) where $\chi_{\mu t} = q_{\alpha\mu t}$ and $\chi_{et} = q_{\alpha et}$. With $\phi_t = 1$ and $p_{\alpha t} = \xi_t$, the incentive compatibility constraint (A26) can be written as $\beta_{\mu t} \geq \frac{\lambda}{\nu} + \xi_t + \beta_{et}$.

If the IC constraint binds everywhere, then:

$$\mathrm{d}\xi_t = (r\xi_t - \lambda)\,\mathrm{d}t + \chi_{\mu t}\mathrm{d}\hat{z}_{n\mu t} + \chi_{et}\mathrm{d}\tilde{z}_{net}, \tag{A35}$$

and:

$$\xi_t = \mathrm{E}_{Pt}\left[\int_t^\tau \mathrm{e}^{-r(s-t)}\lambda\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\tfrac{\lambda-\psi}{r}\right] = \tfrac{1}{r}\left(\lambda - \psi\mathrm{E}_{Pt}\left[\mathrm{e}^{-r(\tau-t)}\right]\right). \tag{A36}$$

To see this, put:

$$\hat{\xi}_t = \mathrm{E}_{Pt}\left[\int_0^\tau \mathrm{e}^{-rs}\lambda\mathrm{d}s + \mathrm{e}^{-r\tau}\tfrac{\lambda-\psi}{r}\right] = \int_0^t \mathrm{e}^{-rs}\lambda\mathrm{d}s + \mathrm{e}^{-rt}\xi_t, \tag{A37}$$

which is a martingale so that:

$$0 = \underbrace{-\mathrm{e}^{-rt}\left((r\xi_t - \lambda)\,\mathrm{d}t - \mathrm{E}[\mathrm{d}\xi_t]\right)}_{\mathrm{E}[\mathrm{d}\hat{\xi}_t]}, \tag{A38}$$

and the drift matches the dynamics (A35).

If the constraint does not bind everywhere, then there is a nonnegative process ε_t such that $\beta_{\mu t} = \frac{\lambda}{\nu} + \xi_t + \beta_{et} + \varepsilon_t$, $\mathrm{d}\xi_t = (r\xi_t - \lambda - \nu\varepsilon_t)\,\mathrm{d}t + \chi_{\mu t}\mathrm{d}\hat{z}_{n\mu t} + \chi_{et}\mathrm{d}\tilde{z}_{net}$ and:

$$\begin{aligned}\xi_t = \mathrm{E}_{Pt}&\left[\int_t^\tau \mathrm{e}^{-r(s-t)}(\lambda + \nu\varepsilon_s)\mathrm{d}s + \mathrm{e}^{-r(\tau-t)}\tfrac{\lambda-\psi}{r}\right] \\ &= \zeta_t + \nu\mathrm{E}_{Pt}\left[\int_t^\tau \mathrm{e}^{-r(s-t)}\varepsilon_s\mathrm{d}s\right] \geq \zeta_t,\end{aligned} \tag{A39}$$

giving 4. ■

The following result is used in deriving optimal thresholds:

Claim A3. *The solution z^* to the equation:*

$$0 = k_0 - \tfrac{1}{1-\mathrm{e}^z}(k_1 - z), \tag{A40}$$

is $z^* = k_1 - k_0 - \omega\left(-\mathrm{e}^{k_1-k_0}k_0\right)$, subject to the condition $-\mathrm{e}^{k_1-k_0}k_0 \geq -\mathrm{e}^{-1}$. If $k_1 < 0$, then the right-hand-side of (A40) *is increasing in z and $\frac{\mathrm{d}z}{\mathrm{d}k_1} > 0$, $\frac{\mathrm{d}z}{\mathrm{d}k_0} < 0$.*

Proof. Expression (A40) can be rearranged as:

$$-\mathrm{e}^{k_1-k_0}k_0 = \underbrace{(k_1 - k_0 - z)\mathrm{e}^{k_1-k_0-z}}_{\omega^{-1}(k_1-k_0-z)}, \tag{A41}$$

immediately implying the result. The condition $-\mathrm{e}^{k_1-k_0}k_0 > -\mathrm{e}^{-1}$ is required because Lambert's W function is valid for arguments that weakly exceed $-\mathrm{e}^{-1}$. The function $g(z) = k_0 - \frac{1}{1-\mathrm{e}^z}(k_1 - z)$ has $g'(z) = \frac{\mathrm{e}^z}{(1-\mathrm{e}^z)^2}\left(z + \mathrm{e}^{-z} - 1 - k_1\right)$, where $z + \mathrm{e}^{-z} - 1 > 0$ for all $z < 0$ so that $k_1 < 0$ guarantees $g'(z) > 0$ and gives $\frac{\mathrm{d}z}{\mathrm{d}k_1} > 0$, $\frac{\mathrm{d}z}{\mathrm{d}k_0} < 0$. ■

Proof of Proposition 1

Part 1

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

$$v_n(\hat{\mu}_{nt}) = \tfrac{1}{r}\hat{\mu}_{nt} + \tfrac{\sigma_\mu^2}{2r}v_n''(\hat{\mu}_{nt}) \quad \Rightarrow \quad v_n(\hat{\mu}_{nt}) = \tfrac{1}{r}\hat{\mu}_{nt} + c_1\mathrm{e}^{\sqrt{2r}\hat{\mu}_{nt}/\sigma_\mu} + c_2\mathrm{e}^{-\sqrt{2r}\hat{\mu}_{nt}/\sigma_\mu}, \tag{A42}$$

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

$$\underbrace{\tfrac{1}{r}\underline{\mu} + c_2 \mathrm{e}^{-\sqrt{2r}\underline{\mu}/\sigma_\mu}}_{v_n\left(\underline{\mu}\right)} = \underbrace{c_2}_{v_n(0)} + R(\underline{\mu}) - k \quad \Rightarrow \quad v_n\left(\hat{\mu}_{nt}\right) = \tfrac{1}{r}\hat{\mu}_{nt} + \frac{\mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}\right)/\sigma_\mu}}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\left(\tfrac{1}{r} - R_\mu\right)\left(\underline{\mu}_o - \underline{\mu}\right).$$

Smooth pasting $\left(v_n'\left(\underline{\mu}\right) = R_\mu\right)$ gives the condition following condition that yields the surplus-maximizing threshold $\underline{\mu}_{fb}$:[4]

$$0 = 1 - \frac{1}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}} \frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o - \underline{\mu}\right). \tag{A43}$$

The solution (21) follows from Claim A3, putting $z = \frac{\sqrt{2r}}{\sigma_\mu}\underline{\mu}_{fb}$, $k_0 = 1$, and $k_1 = \frac{\sqrt{2r}}{\sigma_\mu}\underline{\mu}_o$. Because c_2 is increasing for all $\underline{\mu} < \underline{\mu}_{fb}$ and decreasing for all $\underline{\mu} \in (\underline{\mu}_{fb}, 0)$, this is the unique optimum. Solving the optimality condition for $\underline{\mu}_o - \underline{\mu}_{fb}$ and substituting into v_n gives expression (22).

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

$$\mathrm{d}\,\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right] = -r\mathrm{e}^{-rt}f\left(\hat{\mu}_{nt}\right)\mathrm{d}t + \mathrm{e}^{-rt}f'\left(\hat{\mu}_{nt}\right)\underbrace{\sigma_\mu \mathrm{d}\hat{z}_{n\mu t}}_{\mathrm{d}\hat{\mu}_{nt}} + \tfrac{1}{2}\mathrm{e}^{-rt}f''\left(\hat{\mu}_{nt}\right)\underbrace{\sigma_\mu^2 \mathrm{d}t}_{\left(\mathrm{d}\hat{\mu}_{nt}\right)^2}. \tag{A44}$$

Because $\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right]$ is a martingale, it has zero drift so that the portion of it that depends on $\hat{\mu}_{nt}$ solves an equation similar to (A42). The boundary conditions $\lim_{\hat{\mu}_{nt}\to\infty}\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right] = 0$ and $\lim_{\hat{\mu}_{nt}\to\underline{\mu}_{fb}}\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right] = 1$ give $\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right] = \mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}_{fb}\right)/\sigma_\mu}$.

Part 2

If the firm faces only the participation constraint, then the agent payoff can be

[4] This condition is also equivalent to maximizing v_n with respect to $\underline{\mu}$, which is equivalent to maximizing c_2.

structured so that the CEO's expected value exactly meets the reservation value $\underline{w}_0$. This gives the following payoff to the principal at each contracting date τ_i:

$$\begin{aligned} b_{\tau_i}(0,\underline{w}_0) &= \mathrm{E}_{\tau_i}\left[\int_{\tau_i}^{\tau_{i+1}} \mathrm{e}^{-r(s-\tau_i)}\left(\hat{\mu}_{ns}-c_s\right)\mathrm{d}s + \mathrm{e}^{-r(\tau_{i+1}-\tau_i)}\left(b_{\tau_{i+1}}(0,\underline{w}_0)-k\right)\right] \\ &= \mathrm{E}_{\tau_i}\left[\mathrm{e}^{-r(\tau_{i+1}-\tau_i)}\left(b_{\tau_{i+1}}(0,\underline{w}_0)+\tfrac{1}{r}\left(\underline{\mu}_o-\hat{\mu}_{n\tau_{i+1}}\right)\right)\right]-\underline{w}_0 \\ &= v_n\left(0\right)+\mathrm{E}_{\tau_i}\left[\mathrm{e}^{-r(\tau_{i+1}-\tau_i)}\left(b_{\tau_{i+1}}(0,\underline{w}_0)-v_n\left(0\right)\right)\right]-\underline{w}_0 \\ &= \mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}\left(b_{\tau_{i+1}}(0,\underline{w}_0)+\left(\tfrac{1}{r}-R_\mu\right)\left(\underline{\mu}_o-\underline{\mu}\right)\right)-\underline{w}_0. \end{aligned} \tag{A45}$$

The first-order condition implies the optimal threshold $\underline{\mu}_{**}=\underline{\mu}_o-\frac{\sigma_\mu}{\sqrt{2r}}+\frac{b_{\tau_{i+1}}(0,\underline{w}_0)}{\frac{1}{r}-R_\mu}$, and the second-order condition is satisfied at any $\underline{\mu}$ that solves the first-order condition. Substituting back into the objective function gives:

$$b_{\tau_i}(0,\underline{w}_0)=\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}\frac{\sigma_\mu}{\sqrt{2r}}\left(\tfrac{1}{r}-R_\mu\right)-\underline{w}_0. \tag{A46}$$

Because $b_{\tau_i+1}(0,\underline{w}_0)=b_{\tau_i}(0,\underline{w}_0)$, the threshold $\underline{\mu}_p$ satisfies:

$$\begin{aligned} \underline{\mu}_p=\underline{\mu}_o-\frac{\sigma_\mu}{\sqrt{2r}}+\frac{b_{\tau_{i+1}}(0,\underline{w}_0)}{\frac{1}{r}-R_\mu} &= \underline{\mu}_o-\frac{\sigma_\mu}{\sqrt{2r}}\left(1-\mathrm{e}^{\sqrt{2r}\underline{\mu}_p/\sigma_\mu}\right)-\frac{1}{\frac{1}{r}-R_\mu}\underline{w}_0 \\ \Rightarrow\quad 0 &= 1-\frac{1}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}_p/\sigma_\mu}}\frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o-\frac{1}{\frac{1}{r}-R_\mu}\underline{w}_0-\underline{\mu}_p\right), \end{aligned} \tag{A47}$$

which has a similar solution to the surplus-maximizing threshold $\underline{\mu}_{fb}$ with $\underline{\mu}_o-\frac{1}{\frac{1}{r}-R_\mu}\underline{w}_0$ appearing in the equality rather than $\underline{\mu}_o$. Both $\underline{\mu}_{fb}$ and $\underline{\mu}_p$ have the form $\frac{\sigma_\mu}{\sqrt{2r}}m(x)$ where $m(x)=x-\omega\left(-\mathrm{e}^x\right)$ and $x=\frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o-\frac{\sigma_\mu}{\sqrt{2r}}\right)$ for $\underline{\mu}_{fb}$ and $x=\frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o-\frac{\sigma_\mu}{\sqrt{2r}}-\frac{1}{\frac{1}{r}-R_\mu}\underline{w}_0\right)$ for $\underline{\mu}_p$. Because $m'(x)=\frac{1}{1+\omega(-\mathrm{e}^x)}$ and the function $\omega(x)>-1$, m is increasing so that $\underline{\mu}_p<\underline{\mu}_{fb}$.

To obtain the function for intermediate times, set compensation equal to $r\underline{w}_0$ with terminal lump sum of $\underline{w}_0-R(\underline{\mu}_p)$. The CEO expects:

$$\mathrm{E}\left[\int_t^\tau \mathrm{e}^{-r(s-t)}r\underline{w}_0\mathrm{d}s+\mathrm{e}^{-r(\tau-t)}\left(R(\hat{\mu}_{n\tau})+\underline{w}_0-R(\underline{\mu}_p)\right)\right]=\underline{w}_0. \tag{A48}$$

More generally, if the principal sets the agent's compensation equal to $r\underline{w}_0$ with a lump sum of $\underline{w}_0 - R(\underline{\mu})$ at termination, then the principal's payoff satisfies the HJB equation, and the payoff is maximized at $\underline{\mu} = \underline{\mu}_p$. This contract gives (24), which can be derived from direct computations using substitutions from (A47).

Part 3

Solving $b_n(0, \underline{w}_0; \underline{\mu}_p) > 0$ yields the following inequality after substituting from (A47):

$$b_n(0, \underline{w}_0; \underline{\mu}_p) > 0 \quad \Leftrightarrow \quad \underline{\mu}_p > \underline{\mu}_o - \tfrac{\sigma_\mu}{\sqrt{2r}}. \tag{A49}$$

The condition (A47) implies that $\underline{\mu}_p$ is strictly decreasing in $\underline{w}_0$, which implies that $\underline{\mu}_p > \underline{\mu}_o - \frac{\sigma_\mu}{\sqrt{2r}}$ if and only if $\underline{w}_0 > \bar{w}$. ■

Proof of Proposition 2

Part 1: To implement a contract that terminates when $\hat{\mu}_{nt}$ hits $\underline{\mu}$, it is necessary to set $\beta_{\tilde{e}t} = 0$. Otherwise, w_t may cross $R(\underline{\mu})$, leading to termination, even though $\hat{\mu}_{nt}$ has not crossed $\underline{\mu}$. With a fixed termination contract, $\mathrm{E}_t[\mathrm{e}^{-r(\tau - t)}] = \mathrm{e}^{-\sqrt{2r}(\hat{\mu}_{nt} - \underline{\mu})/\sigma_\mu}$. Incentive compatibility then gives:

$$\tfrac{\mathrm{d}w_t}{\mathrm{d}\hat{\mu}_{nt}} = \beta_{\mu t} \geq \tfrac{\lambda}{\nu} + \xi_t \geq \tfrac{\lambda}{\nu} + \zeta_t = \lambda\left(\tfrac{1}{\nu} + \tfrac{1}{r}\right) - \tfrac{\psi}{r}\mathrm{e}^{-\sqrt{2r}(\hat{\mu}_{nt} - \underline{\mu})/\sigma_\mu}. \tag{A50}$$

Taking the agent's continuation value as a function of $\hat{\mu}_{nt}$, we have:

$$\begin{aligned} w_t &= \underbrace{R(\underline{\mu})}_{\substack{\text{Value of}\\ \text{leaving}\\ \text{at } \hat{\mu}_{nt} = \underline{\mu}}} + \underbrace{\int_{\underline{\mu}}^{\hat{\mu}_{nt}} \frac{\mathrm{d}w_s}{\mathrm{d}\hat{\mu}_{ns}} \mathrm{d}\hat{\mu}_{ns}}_{\substack{\text{Value of}\\ \text{pay until}\\ \hat{\mu}_{nt}=\underline{\mu}}} \\ &\geq R(\underline{\mu}) + \lambda\left(\tfrac{1}{\nu} + \tfrac{1}{r}\right)\left(\hat{\mu}_{nt} - \underline{\mu}\right) - \tfrac{\psi}{r}\int_{\underline{\mu}}^{\hat{\mu}_{nt}} \mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{ns}-\underline{\mu}\right)/\sigma_\mu} \mathrm{d}\hat{\mu}_{ns} \qquad \text{(A51)} \\ &= \underbrace{R(\underline{\mu}) + \lambda\left(\tfrac{1}{\nu} + \tfrac{1}{r}\right)\left(\hat{\mu}_{nt} - \underline{\mu}\right) - \tfrac{\psi}{r}\tfrac{\sigma_\mu}{\sqrt{2r}}\left(1 - \mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}\right)/\sigma_\mu}\right)}_{\underline{w}(\hat{\mu}_{nt},\underline{\mu})}. \end{aligned}$$

If the agent's continuation value $w_t < \underline{w}(\hat{\mu}_{nt}, \underline{\mu})$, then incentive compatible contracts have paths that can lead to termination prior to $\hat{\mu}_{nt}$ crossing $\underline{\mu}$. Therefore, the contract sets the initial reservation value to at least $\underline{w}(0, \underline{\mu})$, and the participation constraint does not bind for agents with $\underline{w}_0 \leq \underline{w}(0, \underline{\mu})$.

Part 3: From Lemma A2, if the participation constraint binds everywhere and $\beta_{et} = 0$, then:

$$\beta_{\mu t} = \lambda\left(\tfrac{1}{\nu} + \tfrac{1}{r}\right) - \tfrac{\psi}{r}\underbrace{\mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}\right)/\sigma_\mu}}_{\mathrm{E}_t\left[\mathrm{e}^{-r(\tau-t)}\right]}. \qquad \text{(A52)}$$

In this case, $w_t = \underline{w}(\hat{\mu}_{nt}, \underline{\mu})$, which implies:

$$\mathrm{d}w_t = (rw_t - c_t)\,\mathrm{d}t + \beta_{\mu t}\mathrm{d}\hat{\mu}_{nt}, \qquad \text{(A53)}$$

where c_t is given by (26). To have $c_t \geq \underline{c}$ for all $\hat{\mu}_{nt} \geq \underline{\mu}$, it must be that $R(\underline{\mu}) - \frac{\sigma_\mu}{\sqrt{2r}}\frac{\psi}{r} \geq \frac{1}{r}\underline{c}$. Otherwise, the payments must be higher, which implies that the incentive compatibility constraints must not be binding for some $\hat{\mu}_{nt}$.

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

$\alpha_t = \nu \int_{t_0}^{t} \mathrm{e}^{-\nu(t-s)} a_s \mathrm{d}s \geq 0$, the flow compensation is:

$$c_t = rR(\underline{\mu}) - \psi \frac{\sigma_\mu}{\sqrt{2r}} + \lambda \frac{\nu+r}{\nu} \left(\hat{\mu}_{nt} - \alpha_t - \underline{\mu} \right). \tag{A54}$$

Because the principal's beliefs are $\hat{\mu}_{nt} - \alpha_t \leq \hat{\mu}_{nt}$, the termination time τ_a with a deviation is weakly earlier than the termination time τ without. The agent's payoff with a deviation is:

$$\begin{aligned} w_{t_0}^a &= \mathrm{E}_{t_0}\left[\int_{t_0}^{\tau_a} \mathrm{e}^{-r(s-t_0)}(c_s + \lambda a_s)\mathrm{d}s + \mathrm{e}^{-r(\tau_a - t_0)} \hat{R}(\underline{\mu}, \hat{\mu}_{n\tau_a})\right] \\ &= \underbrace{\mathrm{E}_{t_0}\left[\int_{t_0}^{\tau} \mathrm{e}^{-r(s-t_0)} \left(rR(\underline{\mu}) - \psi \frac{\sigma_\mu}{\sqrt{2r}} + \lambda \frac{\nu+r}{\nu} \left(\hat{\mu}_{ns} - \underline{\mu} \right) \right) \mathrm{d}s + \mathrm{e}^{-r(\tau - t_0)} R(\underline{\mu})\right]}_{\text{Continuation value } w_{t_0} \text{ with } a_t = 0\ \forall t > t_0} \\ &\quad \underbrace{+ \mathrm{E}_{t_0}\left[\int_{t_0}^{\tau_a} \mathrm{e}^{-r(s-t_0)} \lambda \left(a_s - \frac{\nu+r}{\nu} \alpha_s \right) \mathrm{d}s + \mathrm{e}^{-r(\tau_a - t_0)} \hat{R}(\underline{\mu}, \hat{\mu}_{n\tau_a})\right]}_{\text{Incremental value from } t_0 \text{ to } \tau_a} \\ &\quad \underbrace{- \mathrm{E}_{t_0}\left[\int_{\tau_a}^{\tau} \mathrm{e}^{-r(s-t_0)} \left(rR(\underline{\mu}) - \psi \frac{\sigma_\mu}{\sqrt{2r}} + \lambda \frac{\nu+r}{\nu} \left(\hat{\mu}_{ns} - \underline{\mu} \right) \right) \mathrm{d}s + \mathrm{e}^{-r(\tau - t_0)} R(\underline{\mu})\right]}_{\text{Incremental value from } \tau_a \text{ to } \tau} \end{aligned} \tag{A55}$$

Applying the implicit function theorem gives the following using $\alpha_{\tau_a} = \hat{\mu}_{n\tau_a} - \underline{\mu}$, $\alpha_{t_0} = 0$, and $\mathrm{d}\alpha_t = \nu(a_t - \alpha_t)\mathrm{d}t$:

$$\begin{aligned} &\mathrm{E}_{t_0}\left[\int_{t_0}^{\tau_a} \mathrm{e}^{-r(s-t_0)} \alpha_s \mathrm{d}s\right] = -\frac{1}{r}\mathrm{E}_{t_0}\left[\mathrm{e}^{-r(\tau_a - t_0)} \alpha_{\tau_a} - \int_{t_0}^{\tau_a} \mathrm{e}^{-r(s-t_0)} \nu (a_s - \alpha_s) \mathrm{d}s\right] \\ \Rightarrow \quad &\mathrm{E}_{t_0}\left[\int_{t_0}^{\tau_a} \mathrm{e}^{-r(s-t_0)} \alpha_s \mathrm{d}s\right] = \frac{\nu}{\nu+r}\mathrm{E}_{t_0}\left[\int_{t_0}^{\tau_a} \mathrm{e}^{-r(s-t_0)} a_s \mathrm{d}s\right] - \frac{1}{\nu+r}\mathrm{E}_{t_0}\left[\mathrm{e}^{-r(\tau_a - t_0)} \left(\hat{\mu}_{n\tau_a} - \underline{\mu} \right)\right]. \end{aligned} \tag{A56}$$

The incremental value from t_0 to τ_a is then:

$$\mathrm{E}_{t_0}\left[\mathrm{e}^{-r(\tau_a - t_0)} \left(R(\underline{\mu}) + \left(\lambda \left(\frac{1}{\nu} + \frac{1}{r} \right) - \frac{\psi}{r} \right) \left(\hat{\mu}_{n\tau_a} - \underline{\mu} \right) \right)\right]. \tag{A57}$$

The incremental value from τ_a to τ is the following using $\hat{\mu}_{n\tau} = \underline{\mu}$ and $\mathrm{E}_{\tau_a}[\mathrm{e}^{-r(\tau - \tau_a)}] =$

$\mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{n\tau_a}-\underline{\mu}\right)/\sigma_\mu}$:

$$\mathrm{E}_{t_0}\left[\mathrm{e}^{-r(\tau_a-t_0)}\left(R(\underline{\mu})-\tfrac{\psi}{r}\tfrac{\sigma_\mu}{\sqrt{2r}}\left(1-\mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{n\tau_a}-\underline{\mu}\right)/\sigma_\mu}\right)+\lambda\left(\tfrac{1}{\nu}+\tfrac{1}{r}\right)\left(\hat{\mu}_{n\tau_a}-\underline{\mu}\right)\right)\right]. \tag{A58}$$

Combining terms gives:

$$w_t-w^a_{t_0}=\tfrac{\psi}{r}\tfrac{\sigma_\mu}{\sqrt{2r}}\mathrm{E}_{t_0}\left[\mathrm{e}^{-r(\tau_a-t_0)}\left(\tfrac{\sqrt{2r}}{\sigma_\mu}\left(\hat{\mu}_{n\tau_a}-\underline{\mu}\right)+\mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{n\tau_a}-\underline{\mu}\right)/\sigma_\mu}-1\right)\right], \tag{A59}$$

which is positive since the function $x+\mathrm{e}^{-x}-1$ is increasing for all positive x and $\hat{\mu}_{n\tau_a}\geq\underline{\mu}$. This establishes global incentive compatibility.

The value of firm-specific cash flows to the principal is:

$$\begin{aligned} b_n(\hat{\mu}_{nt},w_t)&=\mathrm{E}_t\left[\int_t^\tau \mathrm{e}^{-r(s-t)}\left(\hat{\mu}_{ns}-c_s\right)\mathrm{d}s+\mathrm{e}^{-r(\tau-t)}\left(b_n\left(0,\max\{\underline{w}(0,\underline{\mu}),\underline{w}_0\}\right)-k\right)\right]\\ &=\tfrac{1}{r}\hat{\mu}_{nt}+\mathrm{e}^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}\right)/\sigma_\mu}\left(b_n\left(0,\max\{\underline{w}(0,\underline{\mu}),\underline{w}_0\}\right)+\left(\tfrac{1}{r}-R_\mu\right)\left(\underline{\mu}_o-\underline{\mu}\right)\right)-w_t. \end{aligned} \tag{A60}$$

Setting the parameters to their values at contract initiation and solving for $b_n\left(0,\max\{\underline{w}(0,\underline{\mu}),\underline{w}_0\}\right)$ gives (29). Substituting back into $b_n(\hat{\mu}_{nt},w_t)$ yields (28).

Proof of part 4

The continuation value (28) satisfies the HJB equation:

$$rb_n\left(\hat{\mu}_{nt}\right)=\hat{\mu}_{nt}-c_t+\tfrac{\sigma_\mu^2}{2}b_n''\left(\hat{\mu}_{nt}\right), \tag{A61}$$

and the value matching condition:

$$b_n\left(\underline{\mu}\right)=b_n\left(0\right)-k-\max\left\{0,\underline{w}_0-\underline{w}(0,\underline{\mu})\right\}, \tag{A62}$$

where c_t is given by (26). Smooth pasting condition $b_n'\left(\underline{\mu}\right)=0$ gives the optimal threshold, where the derivative equals zero since the firm's continuation payoff does not depend on $\hat{\mu}_{nt}$. This gives the condition:

$$0 = \frac{1}{r} - \lambda\left(\frac{1}{\nu} + \frac{1}{r}\right) + \frac{\psi}{r} - \frac{\sqrt{2r}}{\sigma_\mu}\frac{1}{1-e^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\left(\left(\frac{1}{r} - R_\mu\right)\left(\underline{\mu}_o - \underline{\mu}\right) - \max\{\underline{w}(0,\underline{\mu}), \underline{w}_0\}\right). \tag{A63}$$

Substituting back into (28) then gives (30) after a substitution from (A63). Claim A3 can be used to solve expression (A63) for $\underline{\mu}_c$ by putting $z = \frac{\sqrt{2r}}{\sigma_\mu}\underline{\mu}$, $k_0 = \frac{\frac{1}{r}-\lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)+\frac{\psi}{r}}{\frac{1}{r}-R_\mu} \in (0,1)$, and $k_1 = \frac{\sqrt{2r}}{\sigma_\mu}\left(\underline{\mu}_o - \frac{1}{\frac{1}{r}-R_\mu}\max\{\underline{w}(0,\underline{\mu}), \underline{w}_0\}\right) < 0$. ■

Proof of Corollary 2.1

We can write (30) as follows after explicitly writing $\underline{w}(\hat{\mu}_{nt}, \underline{\mu}_c)$ and combining terms:

$$b_n(\hat{\mu}_{nt}) = \frac{\sigma_\mu}{\sqrt{2r}}\left(\frac{1}{r} - \lambda\left(\frac{1}{\nu} + \frac{1}{r}\right)\right)\left(\frac{\sqrt{2r}}{\sigma_\mu}\left(\hat{\mu}_{nt} - \underline{\mu}_c\right) + e^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}_c\right)/\sigma_\mu}\right) + \frac{1}{r}\underline{\mu}_c + \frac{\psi}{r}\frac{\sigma_\mu}{\sqrt{2r}} - R(\underline{\mu}_c), \tag{A64}$$

which, because $\lambda < \frac{\nu}{\nu+r}$, implies that $b_n(\hat{\mu}_{nt})$ is increasing and convex in $\hat{\mu}_{nt}$. Because $b_n(\hat{\mu}_{nt})$ is monotonic, we can invert it to express beliefs $\hat{\mu}_{nt}$ as a function of the firm-specific portion of market value b_t. Specifically, expression (A64) can be rearranged as:

$$\frac{\sqrt{2r}}{\sigma_\mu}\left(\hat{\mu}_{nt} - \underline{\mu}_c\right) + e^{-\sqrt{2r}\left(\hat{\mu}_{nt}-\underline{\mu}_c\right)/\sigma_\mu} = \frac{\frac{\sqrt{2r}}{\sigma_\mu}\left(b_t - \frac{1}{r}\underline{\mu}_c - \frac{\psi}{r}\frac{\sigma_\mu}{\sqrt{2r}} + R(\underline{\mu}_c)\right)}{\frac{1}{r} - \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)}, \tag{A65}$$

The function $z + e^{-z} = k$ has the solution $z = k + \omega\left(-e^{-k}\right)$, giving:

$$\hat{\mu}_{nt} = \frac{b_t - \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)\underline{\mu} - \frac{\psi}{r}\frac{\sigma_\mu}{\sqrt{2r}} + R(\underline{\mu})}{\frac{1}{r} - \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)} + \frac{\sigma_\mu}{\sqrt{2r}}\omega\left(-\exp\left\{-\frac{\frac{\sqrt{2r}}{\sigma_\mu}\left(b_t - \frac{1}{r}\underline{\mu} - \frac{\psi}{r}\frac{\sigma_\mu}{\sqrt{2r}} + R(\underline{\mu})\right)}{\frac{1}{r} - \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)}\right\}\right). \tag{A66}$$

Because $b_n(\hat{\mu}_{nt})$ is increasing and convex, $\hat{\mu}_{nt}(b_{nt})$ is increasing and concave. ■

.0.1 Proof of Corollary 2.2

Define the equation on the right-hand-side of (A63) by $g(\underline{\mu})$. Direct computations give:

$$\begin{aligned} \frac{\partial g}{\partial \nu} &= \frac{\lambda}{\nu^2}\left(1 + 1_{\underline{w}(0,\underline{\mu})>\underline{w}_0}\frac{1}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\frac{\sqrt{2r}}{\sigma_\mu}\underline{\mu}\right), \\ \frac{\partial g}{\partial \underline{\mu}} &= \frac{\sqrt{2r}}{\sigma_\mu}\left(\frac{1}{r} - \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)\right) + 1_{\underline{w}(0,\underline{\mu})<\underline{w}_0}\frac{\sqrt{2r}}{\sigma_\mu}\frac{1}{1-\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu}}\left(\lambda\left(\frac{1}{\nu}+\frac{1}{r}\right) - \frac{\psi}{r}\mathrm{e}^{\sqrt{2r}\underline{\mu}/\sigma_\mu} - R_\mu\right). \end{aligned} \tag{A67}$$

The parameter restrictions $R_\mu < \lambda\left(\frac{1}{\nu}+\frac{1}{r}\right) - \frac{\psi}{r}$ and $\lambda < \frac{\nu}{\nu+r}$ imply that $\frac{\partial g}{\partial \underline{\mu}} > 0$. The function $1 + \frac{z}{1-\mathrm{e}^z}$ is negative for all $z < 0$, which implies that $\frac{\partial g}{\partial \nu} < 0$ if and only if $\underline{w}(0,\underline{\mu}) > \underline{w}_0$. The implicit function theorem then gives $\frac{\mathrm{d}\underline{\mu}}{\mathrm{d}\nu} = -\frac{\partial g/\partial \nu}{\partial g/\partial \underline{\mu}}$, which is positive if and only if $\underline{w}(0,\underline{\mu}) > \underline{w}_0$.

For the effect of ν on b_n, we have:

$$\frac{\mathrm{d}b_n(\hat{\mu}_{nt})}{\mathrm{d}\nu} = \underbrace{\frac{\partial b_n(\hat{\mu}_{nt})}{\partial \nu}}_{>0} + \underbrace{\frac{\partial b_n(\hat{\mu}_{nt})}{\partial \underline{\mu}}}_{>0}\frac{\mathrm{d}\underline{\mu}}{\mathrm{d}\nu}. \tag{A68}$$

When $\underline{w}(0,\underline{\mu}) > \underline{w}_0$, $\frac{\mathrm{d}\underline{\mu}}{\mathrm{d}\nu} > 0$ and $\frac{\mathrm{d}b_n(\hat{\mu}_{nt})}{\mathrm{d}\nu} > 0$. When $\underline{w}(0,\underline{\mu}) < \underline{w}_0$, direct computations give:

$$\frac{\mathrm{d}b_n(\hat{\mu}_{nt})}{\mathrm{d}\nu} = \frac{\sigma_\mu}{\sqrt{2r}}\frac{\lambda}{\nu^2}\left(\frac{\sqrt{2r}}{\sigma_\mu}\left(\hat{\mu}_{nt} - \underline{\mu}\right) - \frac{\lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)-R_\mu}{\frac{1}{r}-\lambda\left(\frac{1}{\nu}+\frac{1}{r}\right)}\right), \tag{A69}$$

which, because $\lambda\left(\frac{1}{\nu}+\frac{1}{r}\right) - R_\mu > 0$, will be negative for $\hat{\mu}_{nt}$ sufficiently small and positive for $\hat{\mu}_{nt}$ sufficiently large. ■