Managerial Turnover and Entrenchment *

Zenan Wu † Xi Weng ‡

Abstract

In this paper, we consider a two-period model in which the success of the firm is realized in the second period and it depends both on period-one (the incumbent) manager's effort and the ability of the manager in office in the second period, and the board makes a retention decision at the end of the first period on the incumbent manager based on noisy signals, which is not directly contractible. We show that the board's information technology of assessing the incumbent manager's ability is an important determinant of the optimal contract and managerial turnover. The board needs to use the contract to both provide incentives for the incumbent manager to exert effort and to ensure that the second-period manager is of high ability (probabilistically). We show that severance pay in the contract serves as a costly device to provide commitment to the incumbent manager in order to induce effort. The optimal replacement policy balances incentive provision, manager selection and commitment provision. Different from models in the existing literature, we characterize the conditions under which entrenchment and anti-entrenchment can emerge under optimal contract.

Keywords: severance pay; entrenchment; managerial turnover; optimal-contracting; moral hazard; rho-concavity.

JEL classification: $D86$, $J33$, $M52$.

*We are deeply indebted to Hanming Fang and Aislinn Bohren for their guidance and continuous support. We are grateful to David Dillenberger and Andrew Postlewaite for their invaluable advice. We would like to thank Navin Kartik, SangMok Lee, George Mailath, Steven Matthews, Lucian Taylor and seminar participants at University of Pennsylvania for helpful comments and suggestions. All errors remain ours.

†Department of Economics, University of Pennsylvania. zenan@sas.upenn.edu

‡Guanghua School of Management, Peking University. wengxi125@gsm.pku.edu.cn

1 Introduction

The design of managerial contract and the decision to replace the CEO are arguably two important decisions made by the board of directors. These two decisions are closely related through a key component of the contracts between firms and CEOs: the severance agreement, which specifies payments to the CEO upon his future forced departure. According to Rusticus (2006), approximately 50% of the CEO compensation contracts involve some form of severance agreements. The percentage of S&P firms that include severance agreement in the CEO compensation contracts has increased from 20% in 1993 to more than 55% in 2007 (Huang, 2011). In general, a contract with positive severance payment creates explicit cost to the board's retention decision and makes replacement more difficult relative to a compensation contract without such agreement.

This paper investigates how the optimal design of the severance agreement influences CEOs entrenchment. In this paper, CEOs are said to be entrenched if board retains some CEOs that are believed to perform worse than a potential average replacement CEO. It is widely believed that CEOs are entrenched.[1] CEOs can be entrenched for many reasons and two popular views regarding entrenchment are present. One views entrenchment as an instance of governance failure in the form of a captive board of directors (see, e.g., Inderst and Mueller, 2010; Shleifer and Vishny, 1989; Hermalin and Weisbach, 1998). The other views entrenchment as the solution to overcoming moral hazard problem (see, e.g., Almazan and Suarez, 2003; Casamatta and Guembel, 2010; Manso, 2011).

Taylor (2010) makes the first attempt to measure the cost of entrenchment using a structural model of CEO turnover and finds suggestive evidence on the opposite, which is referred to as anti-entrenchment in this paper. In particular, he finds that boards in large firms fire CEOs with higher frequency than is optimal. This finding can not be rationalized by the existing models on CEO turnover and thus calls for a new model to better understand the determinants of managerial turnover.

[1] Although numerous evidence shows forced CEO turnover is increasing over time and indicates boards are using more aggressive replacement policies, it is yet widely believed that the CEOs are rarely fired and thus entrenched.For instance, Kaplan and Minton (2012) find that board driven turnover increased steadily from 10.93% (1992 – 1999) to 12.47% (2000 – 2007) using data of publicly traded Fortune 500 companies.

This paper proposes a two-period principal-agent model of CEO turnover and identifies conditions that predict the emergence of entrenchment and anti-entrenchment. Formally, we consider a setup in which the incumbent manager can be incentivized by a contract that constitutes performance-related pay and severance pay. Firm's success depends both on the incumbent manager's effort and the ability of the manager in office after board's retention decision in the second period. Thus, the board faces an ability selection problem and a moral hazard problem. We consider the case in which, after the incumbent manager exerts effort, the board observes a non-contractible signal regarding the incumbent's ability. The board can fire the incumbent manager by paying the severance pay specified in the initial contract and hire a replacement manager. Since the board's information of the incumbent manager's ability is non-contractible, it cannot write a contract that specifies retention decision contingent on the signal, lacking commitment power. Severance pay is used as a costly device to provide commitment of not firing the incumbent manager.

By committing a high severance pay, the board enforces itself a low expected profit after replacement and a less aggressive replacement policy is expected to be played in equilibrium. An aggressive replacement policy is associated with low severance pay. Thus, implementing a more (less) aggressive replacement policy generates a net gain (loss) relative to the first best policy. Board's optimal replacement policy balances incentive provision, manager selection and commitment provision.

We show that when the board's monitoring technology is noisy, entrenchment is optimal to the board. Under such scenario, the board gives priority to motivating the incumbent manager rather than optimizing manager's ability. This is because setting an aggressive replacement policy in equilibrium will fire the incumbent of high ability too often and dis-incentivize the incumbent too much. Meanwhile, a contract that induces an aggressive replacement policy saves little severance pay since the posterior belief of the incumbent's ability is not too much different from the prior. As a result, a contract that induces entrenchment is optimal to the board.

Anti-entrenchment is optimal when the board's monitoring technology is sufficiently informative. Different from the case of uninformative monitoring technology, the board is reluctant to provide commitment. On the one hand, a contract that favors the incumbent manager does not increase effort too much because the probability of replacement conditional on the incumbent being high ability is low. On the other hand, an aggressive equilibrium replacement policy can save a substantial amount of severance pay. Thus, anti-entrenchment is optimal to the board.

To the best of our knowledge, we are the first to unveil the interactions between the board's monitoring technology and turnover and argue that contract that involves anti-entrenchment can be optimal.

The stylized model can be applied to a variety of real-world settings. For example, the model can be used to analyze the turnover of founder CEOs in venture-capital-backed companies where the venture capitalist is a large shareholder and engages in active monitoring. Another example is the contract between the head coaches and the professional sports teams.

Related Literature: This paper belongs to the literature on principal-agent model with replacement.[2] One strand of research views entrenchment as a potential source of inefficiency that the board aims to mitigate. Consequently, anti-entrenchment cannot be observed. Inderst and Mueller (2010) solve the optimal contract for the incumbent CEO who holds private information on firm's future performance and can avoid replacement by concealing bad information. Consequently, optimal contract is designed to induce the incumbent to voluntarily step down when evidence suggests low expected profit under his current management. Similarly, entrenchment occurs if the incumbent can make manager-specific investments to create cost of replacement to the board (Shleifer and Vishny, 1989) or there exist close ties between the board and manager (Hermalin and Weisbach, 1998).

Another strand of research views entrenchment as a feature of optimal contract (board structure) that helps overcome moral hazard problem. Manso (2011) shows that tolerance for early failure (entrenchment) can be part of the optimal incentive scheme when motivating a CEO to pursue more innovative business strategies is an important concern to the board. Casamatta and Guembel (2010) study the optimal contract for the incumbent manager with reputational concern. In their model, entrenchment is optimal because the incumbent manager would like to see his strategy succeed and is less costly to motivate than the replacement manager. Almazan and Suarez (2003) study the optimal board structure to incentivize the incumbent CEO. They show that

[2]See Laux (2014) for a comprehensive survey of the theoretical models on this topic.

it can be optimal for shareholders to relinquish some power and choose a weak board, where the incumbent can veto his departure, rather than a strong board, where the board can fire the incumbent at will. In the same spirit, Laux (2008) studies the optimal degree of board independence for shareholders. He shows that some lack of independence can increase shareholder value. In these papers, boards (shareholders) provide better job security to the incumbent by making dismissal more difficult to induce more effort. Our paper contributes to the existing literature by pointing out that despite all the merits of entrenchment on providing incentives, the cost of doing so can be high when the board's monitoring technology is sufficiently informative.

In terms of modeling, the paper is mostly close to Taylor and Yildirim (2011). They study the benefit and cost of different review policies and identify conditions under which principal commits not to utilize agent's information and chooses blind review as optimal policy. We apply their model to analyze managerial turnover by adding a contract stage to endogenize agent's payoff and allow the principal to replace the agent in the interim stage.

The remainder of the paper is organized as follows. Section 2 describes the baseline model. Section 3 introduces the information structure of the board. Section 4 defines entrenchment (anti-entrenchment) and characterizes the optimal replacement policy. Section 5 studies the impact of informativeness on optimal replacement policy. Section 6 and 7 discuss the key assumptions and extensions of the model. Section 8 concludes. All proofs are in the Appendix.

2 Baseline Model

There are two periods $t = 1, 2$ in all.

2.1 Period 1

2.1.1 Contract stage

The board (Principal), acting on behalf of the shareholders of the firm, hires an incumbent manager (Agent) from a pool with unknown ability $\theta_i \in \{0, 1\}$ to work for the firm. $\Pr(\theta_i = 1) = 1 - \Pr(\theta_i = 0) = \mu \in (0, 1)$. The ability is unknown to both sides.

The board offers a contract to manager. What a contract looks like will be described below.

Both the board and the managers are assumed to be risk neutral. Moreover, we assume managers are protected by limited liability. This assumption is necessary since it excludes the possibility that the board sells the whole firm to manager to provide the highest incentive in optimal contract. Finally, we assume that manager's value of outside option $\underline{u}$ is low enough, say 0. This assumption guarantees that individual rationality (IR) constraint never binds and simplifies the analysis.

2.1.2 Effort and retention stage

Next we describe the continuation game that follows after the contract has been signed between the board and the incumbent manger. The incumbent manager exerts effort to create a project by incurring cost $C(q) = \frac{1}{2}q^2$. q can be interpreted as firm's vision set by the incumbent manager.

After the incumbent manger exerts effort q, the board receives a noisy signal $s \in \mathcal{S}$ of manager's ability θ_i, deciding whether to replace the incumbent manager. If incumbent manger is fired, the incumbent is replaced with a new one (replacement manager) with ability θ_r randomly drawn from the same prior distribution.

The project generation process can also be interpreted as project selection process. Alternatively, we can assume true state η is randomly drawn from uniform distribution $\eta \in U[0, 1]$. The incumbent manager is hired to choose a project $a \in [0, 1]$ that matches the state. The quality of the project is 1 if $a = \eta$ and 0 otherwise. The incumbent manager incur cost $C(q)$ to receive a noisy signal ν of the true state. With probability q, ν is equal to true state η; with complementary probability $1 - q$, ν is pure noise and randomly drawnly from $U[0, 1]$. These two specifications lead to the same model.

2.2 Period 2

After retention decision of the board, the manager that stays in office implements the project with no additional effort and payoffs are realized. Implementation is assumed to be costless and depends only on manager's ability.[3] To formalize this idea, we

[3]This assumption is relaxed in Section 7.2.

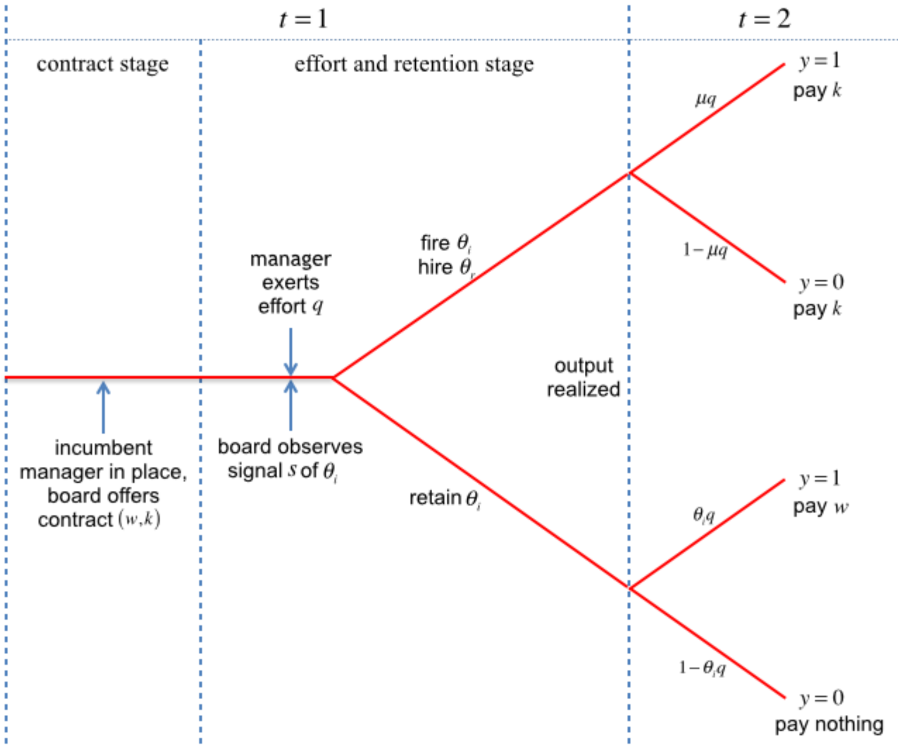

Figure 1: Timeline

assume that the expected quality of the project is equal to $q\tilde{\theta}$, where q is the effort choice of the incumbent manger and $\tilde{\theta}$ is the ability of the manager who stays in office at the beginning of period 2. With probability $q\tilde{\theta}$, the project is of high quality and yields outcome $y = 1$. With complementary probability $1 - q\tilde{\theta}$, the project is of low quality and yields outcome $y = 0$. After payoffs are realized, the incumbent manager, if retained, receives payment according to the contract signed in period 0 and the game comes to an end.

Since only two outcomes will be realized and one of them is normalized to 0, it is obvious that the wage for low output has to be 0. Thus it suffices to calculate the wage rate w when outcome is $y = 1$. Consequently, a contract is defined by the tuple (w, k), where w is the wage rate under the event of $y = 1$ and k is the severance pay to the incumbent manger if the incumbent is forced out by the end of period 1. By the limited liability assumption on the manager, we have $w \geq 0$ and $k \geq 0$.

3 Information Structure

After the incumbent manager exerts effort, the board receives a noisy signal $s \in \mathcal{S}$ about incumbent manager's ability θ_i. s is drawn from distribution with cdf $F_{\theta_i}(\cdot)$ and pdf $f_{\theta_i}(\cdot)$ for $\theta_i \in \{0,1\}$. Without loss of generality, we assume $\mathcal{S} = [0,1]$ and the unconditional distribution of s under $\mu = \frac{1}{2}$ is uniform. That is, $\frac{1}{2}F_1(s) + \frac{1}{2}F_0(s) = s$ for $s \in [0,1]$.[4] The two conditional density functions $\{f_1(s), f_0(s)\}$ suffice to define an information structure under such normalization. Three common assumptions are imposed on the information structure for ease of exposition.

Assumption 1 *Monotone likelihood ratio property (MLRP):* $\frac{f_1(s)}{f_0(s)}$ *is strictly increasing in* s *for* $s \in [0,1]$.

For binary states, MLRP is without loss of generality since signals can always be relabeled according to likelihood ratio to satisfy this assumption.

Assumption 2 *Perfectly informative at extreme signals:* $\lim_{s\to 0} \frac{f_1(s)}{f_0(s)} = 0$ *and* $\lim_{s\to 1} \frac{f_1(s)}{f_0(s)} = +\infty$.

As will be clear later, Assumption 2 guarantees that support of the posterior belief is always $[0,1]$.

The last assumption imposed on the information structure is mirror symmetry. This assumption allows us to define the first best replacement policy on the signal space.

Assumption 3 *Mirror symmetry:* $f_1(s) = f_0(1-s)$ *for all* $s \in [0,1]$.

By Assumption 3, $f_1(\frac{1}{2}) = f_0(\frac{1}{2})$. Thus the likelihood ratio at $s = \frac{1}{2}$ is always 1 and the Bayesian update of the incumbent manager's ability at $\frac{1}{2}$ is equal to the prior μ.

We close this section by introducing an index $\alpha \in (0,\infty)$ to parameterize the information structure. To impose the weakest structure on the parameterized information structure, we assume that $f_{\theta_i}(s;\alpha)$ is continuous in s and α for $\theta_i \in \{0,1\}$ and define the information structures for two extreme values of α as follows.

[4] This assumption is without loss of generality due to the fact any information structure can be normalized via integral probability transformation. See Appendix B for more details.

Assumption 4 (Completely informative/uninformative information structure)

1. *The information structure becomes completely uninformative when* $\alpha \to 0$*, i.e.,* $\lim_{\alpha\to 0}[f_0(s;\alpha) - f_1(s;\alpha)] = 0$ *for* $s \in (0,1)$*.*
2. *The information structure becomes completely informative when* $\alpha \to \infty$*, i.e.,* $\lim_{\alpha\to\infty} f_1(s;\alpha) = 0$ *for* $s \in [0, \frac{1}{2})$ *and* $\lim_{\alpha\to\infty} f_0(s;\alpha) = 0$ *for* $s \in (\frac{1}{2}, 1]$*.*[5]

When information structure becomes completely uninformative ($\alpha \to 0$), the two conditional density functions are the same. When information structure becomes completely informative ($\alpha \to \infty$), board will not observe signal below $\frac{1}{2}$ when the incumbent manager is of high ability and signal above $\frac{1}{2}$ when the incumbent manager's ability is low.

4 Solving for Optimal Cutoff

4.1 The benchmark case: the first best policy

To define entrenchment (anti-entrenchment), it is necessary to first pin down the socially optimal replacement policy. By Assumption 1, the socially optimal replacement policy is a cutoff rule. For the rest of the paper, denote $\hat{s}$ as the cutoff on signal s.

Lemma 1 (First best cutoff) *Suppose the board can contract on effort* q *of the incumbent manager. Then* $q^{\mathcal{FB}} = \mu\Big\{1 + (1-\mu)\big[F_0(\hat{s}^{\mathcal{FB}}) - F_1(\hat{s}^{\mathcal{FB}})\big]\Big\}$*, where the replacement cutoff* $\hat{s}^{\mathcal{FB}} = \frac{1}{2}$*.*

When effort is contractible, board is able to optimize effort and selection separately. Thus, there is no tradeoff between moral hazard problem and selection problem. It is optimal to replace the incumbent manager when the posterior belief of the incumbent's ability falls below the average and retain the incumbent vice versa. By Assumption 3, the likelihood ratio $\frac{f_1(s)}{f_0(s)}$ at $s = \frac{1}{2}$ is always equal to 1. Consequently, the Bayesian update of the incumbent manager's ability is always equal to the prior independent of the informativeness α of information structure. Consequently, the socially optimal cutoff $\hat{s}^{\mathcal{FB}} = \frac{1}{2}$ for all α.

[5] Both compltetely informative and uninformative information structure are defined using pointwise convergence.

Definition 1 *Denote (w^*, k^*) as the optimal contract to the board. Let $(\hat{s}^*, q^*)$ be the equilibrium replacement cutoff and effort of the continuation game induced by the optimal contract. If $\hat{s}^* < \frac{1}{2}$ $(\hat{p}^* < \mu)$, we say the* ***entrenchment*** *is optimal to the board. Otherwise, if $\hat{s}^* > \frac{1}{2}$ $(\hat{p}^* > \mu)$, we say the* ***anti-entrenchment*** *is optimal to the board.*

For the case that $\hat{s}^* = \frac{1}{2}$ $(\hat{p}^* = \mu)$, we say that neither entrenchment nor anti-entrenchment is observed. The replacement policy coincides with the socially optimal. When $\hat{s}^* < \frac{1}{2}$, the replacement policy favors the incumbent manager: the board could have improved implementation by replacing the incumbent. Similarly, the replacement policy is considered aggressive and disadvantages the incumbent manager when $\hat{s}^* > \frac{1}{2}$.

4.2 Board's Optimization Problem

In this section, we solve the equilibrium outcome when effort is non-contractible. The board can only commit to the wage w and severance pay k in the contract. We are interested in the cutoff $\hat{s}^*$ induced by the optimal contract.

4.2.1 Incentives under a given contract (w, k)

We first solve the sub-game in period 1 after a contract (w, k). A contract (w, k) induces a simultaneous move game. The incumbent manager's effort q and the board's replacement policy $\hat{s}$ will be determined in a Cournot-Nash equilibrium.

For a fixed contract (w, k) and belief on cutoff $\hat{s}$, the incumbent manager chooses effort q to maximize,

$$\mu\big[1 - F_1(\hat{s})\big]qw + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]k - C(q).$$

$$\Longrightarrow q = \mu\big[1 - F_1(\hat{s})\big]w. \tag{1}$$

Note that the board can provide incentive on effort by increasing wage w directly and lowering equilibrium cutoff $\hat{s}$ indirectly. Fix w, the equilibrium cutoff $\hat{s}$ is decreasing in k. By committing to a higher severance pay, the board chooses a lower replacement cutoff in equilibrium and is able to induce more effort.[6]

[6] The clear relation between k and q is due to the assumption that board observes signal of the incumbent

For a fixed contract (w, k) and belief of the incumbent manager's effort level q, the board chooses cutoff $\hat{s}$ to maximize,

$$\mu\big[1 - F_1(\hat{s})\big]q(1-w) + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big](\mu q - k).$$

$$\implies \frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})} q(1-w) = \mu q - k. \tag{2}$$

Since higher cutoff implies higher posterior belief of the incumbent manager's ability, the board chooses a cutoff that the expected profit created by the marginal incumbent manager is equal to the expected profit under replacement in equilibrium.

Equilibrium replacement cutoff and effort level $(\hat{s}, q)$ are pinned down by equation (1) and (2) together. We can reversely calculate the corresponding contract (w, k) that induces any tuple $(\hat{s}, q)$ as follows,

$$w(\hat{s}, q) = \frac{q}{\mu\big[1 - F_1(\hat{s})\big]}$$

and

$$k(\hat{s}, q) = \mu q - \frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})} q\big[1 - w(\hat{s}, q)\big].$$ [7]

4.2.2 Optimal cutoff to the board

The board chooses contract (w, k) to maximize:

$$\max_{\{w,k\}} \mu\big[1 - F_1(\hat{s})\big]q(1-w) + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big](\mu q - k)$$

s.t.

$$q = \mu\big[1 - F_1(\hat{s})\big]w$$

and

$$\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})} q(1-w) = \mu q - k.$$

Equivalently, firm is maximizing expected profit over $(\hat{s}, q)$. After some algebra, it

manager's type rather than outcome. This assumption is relaxed later in Section 7.3.

[7] Notice that the non-negativity assumption on k is not always satisfied for all $\hat{s}$ and q. We ignore this limited liability constraint at this moment and solve the unconstrained problem. This is not a big concern since it can be proved later that the optimal wage is $w^* = \frac{1}{2}$. It can be verified that the non-negativity constraint on k is satisfied if $\mu \geq \frac{1}{2}$.

can be verified that $w^* = \frac{1}{2}$ and $q = \frac{1}{2}\mu\big[1 - F_1(\hat{s})\big]$ under optimal contract. Thus the expected profit can be rewritten in terms of $\hat{s}$ alone,

$$\pi(\hat{s}) = \frac{1}{4}\mu\big[1-F_1(\hat{s})\big]\left\{\mu\big[1 - F_1(\hat{s})\big] + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu)f_0(\hat{s})}\right\}.$$

The optimal cutoff highly depends on the information structure $\big\{f_1(\cdot), f_0(\cdot)\big\}$. Informativeness α is the key. To better understand the driving forces in determining the optimal cutoff $\hat{s}^*$, it is beneficial to rewrite the expected profit as follows:

$$\begin{aligned}\pi(\hat{s}) = \frac{1}{4}\mu\underbrace{\Big[1 - F_1(\hat{s})\Big]}_{\text{incentive effect}}\Bigg\{&\underbrace{\Big[\mu\big[1 - F_1(\hat{s})\big] + \mu\big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\Big]}_{\text{selection effect}} \\ &+ \underbrace{\Big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\Big]\Big(\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu)f_0(\hat{s})} - \mu\Big)}_{\text{commitment cost effect}}\Bigg\}.\end{aligned}$$

4.2.3 Three effects in determining $\hat{s}^*$

Three effects play important roles of determination of the optimal cutoff. $\Big[\mu[1-F_1(\hat{s})] + \mu[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})]\Big]$ is called the selection effect. This is the expected ability of the manager in period 2. Increasing $\hat{s}$ will increase the expected ability of the manager in office when $\hat{s} < \frac{1}{2}$ and decrease the expected ability when $\hat{s} \geq \frac{1}{2}$. To optimize ability selection alone, the board sets $\hat{s} = \frac{1}{2}$.

Since outcome also depends on the effort choice of the incumbent manager, board faces a moral hazard problem in addition to ability selection and needs to incentivize the incumbent. This is captured by $\big[1 - F_1(\hat{s})\big]$, which is referred to as the incentive effect. As the equilibrium replacement cutoff $\hat{s}$ increases, the incumbent manager expects a lower retaining probability in equilibrium and exerts less effort accordingly. The board provides more job security to better incentivize the incumbent manager in response. By this effect alone, the board sets $\hat{s} = 0$.

By selection effect and incentive effect, a cutoff below $\frac{1}{2}$ is optimal to the board and entrenchment is expected to emerge under optimal contract. However, this intuition is not correct as it firstly seems to be. Since signal is non-contractible, board lacks commitment power on replacement policy and severance pay serves as a costly device to provide commitment of not replacing the incumbent manager. As the promised severance pay increases, the board commits more to not replacing the incumbent by lowering the expected payoff of replacement. In equilibrium the expected profit of replacement is equal to the expected profit created by the marginal incumbent manager. When board lowers the cutoff ($\hat{s} < \frac{1}{2}$) to provide more incentive on effort, it has to increase severance pay to make the equilibrium replacement policy credible. This generates a net loss compared to the first best replacement policy. It is captured by $\left[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\right]\left(\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s})+(1-\mu)f_0(\hat{s})} - \mu\right)$, which is referred to as the commitment cost effect. Compared to the first best cutoff $\hat{s} = \frac{1}{2}$, the board obtains a net commitment gain by providing less commitment and designing a contract that induces cutoff above $\frac{1}{2}$. Similarly, the board suffers a commitment loss by committing a cutoff that is below $\frac{1}{2}$. The net commitment cost effect is shown by $\left(\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s})+(1-\mu)f_0(\hat{s})} - \mu\right)$. Multiplied by the probability of replacement yields the total net commitment gain/loss. By this effect alone, the board sets $\hat{s} = 1$.

Among all three effects, incentive effect and commitment cost effect are the key to determine the optimal cutoff $\hat{s}^*$. If incentive effect dominates commitment cost effect, entrenchment is optimal to the board. Otherwise, anti-entrenchment is expected to emerge in optimal contract.

5 Information Structure and Entrenchment (Anti-entrenchment)

In this section, we study how the optimal replacement policy varies in the informativeness of the board's monitoring technology.

5.1 Limit result of entrenchment (anti-entrenchment)

Proposition 1 *Suppose $\left\{f_1(s;\alpha), f_0(s;\alpha)\right\}$ satisfies Assumption 1 - 4. Then there exists $\overline{\alpha}$ and $\underline{\alpha}$ such that,*

1. *$\hat{s}^*(\alpha) > \frac{1}{2}$ for $\alpha > \overline{\alpha}$;*
2. *$\hat{s}^*(\alpha) < \frac{1}{2}$ for $\alpha < \underline{\alpha}$.*

When information structure is noisy, providing incentive is easier than obtaining net commitment gain. On the one hand, choosing $\hat{s} > \frac{1}{2}$ saves little severance pay. The reason is that the Bayesian update around $\hat{s} = \frac{1}{2}$ changes very slowly. Thus the expected ability of the incumbent manger on the margin does not vary too much compared to the one at $\frac{1}{2}$. On the other hand, choosing $\hat{s} > \frac{1}{2}$ dis-incentivizes the incumbent manager's effort. Consequently, it is optimal for the board to design a contract that induces entrenchment.

When board's monitoring technology is sufficiently informative, obtaining commitment gain is of higher priority. On the one hand, choosing $\hat{s} < \frac{1}{2}$ is not in the interest of the board. Since the probability of firing a high ability manger is very small for all signals below $\frac{1}{2}$, lowering the equilibrium replacement cutoff does not increase the incumbent's effort too much. On the other hand, it is easy to obtain commitment gain. The expected ability of the manager on the right neighborhood of $\frac{1}{2}$ is very close to 1 when information structure is sufficiently informative. That is, board can largely reduces the severance pay by choosing a cutoff slightly above $\frac{1}{2}$. Thus, anti-entrenchment is optimal to the board.

5.2 Non-limit result of entrenchment (anti-entrenchment)

Proposition 1 does not characterize the equilibrium replacement policy in optimal contract for moderate α. To do this, it is necessary to introduce a new information order.

5.2.1 Distribution of posterior beliefs

Denote $p = \varphi(s, \mu)$ as the posterior belief of θ after observing signal s given prior belief μ. Then $\varphi(s, \mu) = \frac{\mu f_1(s)}{\mu f_1(s)+(1-\mu) f_0(s)}$. By Assumption 1, $\varphi(s)$ is strictly increasing in s. By Assumption 2, the support of p is $[0, 1]$. Denote $g(p, \mu)$ as the corresponding density function. Since $\mathbb{E}(\mathbb{E}(\theta|s)) = \mu$, the only constraint we impose on $g(\cdot)$ is that $\int_0^1 pg(p, \mu)dp = \mu$.

Given an information structure $\{f_1(\cdot), f_0(\cdot)\}$, the density function of posterior belief p can be calculated as follows,

$$g(p, \mu) = \left[\mu f_1(\varphi^{-1}(p, \mu)) + (1-\mu) f_0(\varphi^{-1}(p, \mu))\right] \frac{\partial \varphi^{-1}(p, \mu)}{\partial p}.$$

Lemma 2 *For any density function $g(\cdot,\mu)$ with support $[0,1]$ that satisfies $\int_0^1 pg(p,\mu)dp = \mu$, there exists a unique information structure $\{f_1(\cdot), f_0(\cdot)\}$ that induces $g(\cdot,\mu)$.*

By Lemma 2, there exists a one-to-one mapping between information structure $\{f_1(\cdot), f_0(\cdot)\}$ and $g(\cdot,\mu)$ given prior μ. Thus working on information structure $\{f_1(\cdot), f_0(\cdot)\}$ is equivalent to working on distribution of posterior belief $g(\cdot,\mu)$. Consequently, we can define information order on $g(\cdot)$. In the rest of the paper, we focus on the case where $\mu = \frac{1}{2}$ and drop μ in $G(\cdot,\mu)$ and $g(\cdot,\mu)$.[8] When $\mu = \frac{1}{2}$, by Assumption 3 on $\{f_1(\cdot), f_0(\cdot)\}$, $g(p) = g(1-p)$ and $G(p) = 1 - G(1-p)$ for $p \in [0,1]$. Thus, it suffices to order different information structures based on $G(p)$ for $p \in [0, \frac{1}{2}]$.

5.2.2 Information order: ρ-concave order

We use ρ-concavity to define the informativeness of the information structure. [9] To the best of our knowledge, this is the first paper that defines information order using ρ-concavity.

Given $G(\cdot)$, define local ρ-concavity at p as,

$$\rho(p) \equiv 1 - \frac{G(p)g'(p)}{g^2(p)}.$$

By definition, $\rho(p)$ is the power of $G(\cdot)$ such that the second order Taylor expansion at p drops out. Thus, $\rho(p)$ is a measure of concavity of $G(\cdot)$ at point p. Log-concavity is equivalent to $\rho(p) \geq 0$ and concavity is equivalent to $\rho(p) \geq 1$. We focus on the distributions such that $\rho(p) \in (0,\infty)$. This assumption is a necessary condition to guarantee the initial condition $G(0) = 0$ is satisfied.[10]

Definition 2 (ρ-concave order) *$G_1(p)$ is said to be more informative than $G_2(p)$ in the ρ-concave order if $\rho(p|G_1) > \rho(p|G_2)$ for all $p \in [0, \frac{1}{2}]$.*

By definition, $G_1(p)$ is more informative than $G_2(p)$ if $G_1(p)$ is everywhere more concave than $G_2(p)$ measured by local ρ-concavity. It can be proved that concave order

[8] To generalize the information order to $\mu \neq \frac{1}{2}$, one can simply treat $\frac{1}{2}$ as an operator.

[9] For more applications of ρ-concavity in economics, see Mares and Swinkels (2014) in auction theory; Anderson and Renault (2003), Weyl and Fabinger (2013) in industry organization context.

[10] Imposing this non-negativity assumption on $\rho(\cdot)$ is without loss of generality: completely uninformative information structure can still be defined under this constraint.

implies Blackwell's sufficiency and rotation order under $\mu = \frac{1}{2}$.[11]

Assume that $\max_{p\in[0,\frac{1}{2}]}\left\{\rho(p;\alpha)\right\}$ and $\max_{p\in[0,\frac{1}{2}]}\left\{\rho(p;\alpha)\right\}$ exist for all $\alpha \in (0,\infty)$. Denote $\overline{\rho}(\alpha) = \max_{p\in[0,\frac{1}{2}]}\left\{\rho(p;\alpha)\right\}$ and $\underline{\rho}(\alpha) = \min_{p\in[0,\frac{1}{2}]}\left\{\rho(p;\alpha)\right\}$ for notational convenience.

Lemma 3 *Suppose* $0 < \underline{\rho} \leq \overline{\rho} < \infty$. *Then* $\frac{1}{2}(2p)^{\frac{1}{\underline{\rho}}} \leq G(p) \leq \frac{1}{2}(2p)^{\frac{1}{\overline{\rho}}}$ *for* $p \in [0, \frac{1}{2}]$.

By Lemma 3, $G(p)$ can be bounded by two constant cumulative density functions with constant ρ-concavity. Perfectly informative information structure corresponds to the case that $\lim_{\alpha\to\infty}\underline{\rho}(\alpha) = \infty$ and perfectly uninformative information structure is equivalent to $\lim_{\alpha\to\infty}\overline{\rho}(\alpha) = 0$.[12] The following assumptions are imposed on the family of distribution $\left\{G(\cdot;\alpha)\right\}$ indexed by $\alpha \in (0,\infty)$.

Assumption 5 (Log concavity) $\rho(p;\alpha) \in (0,\infty)$ *for* $(p,\alpha) \in [0,\frac{1}{2}] \times (0,\infty)$.

Assumption 6 (Concave order) *If* $\alpha_1 > \alpha_2$, $\rho(p;\alpha_1) > \rho(p;\alpha_2)$ *for* $p \in [0,\frac{1}{2}]$.

Assumption 7 (Regularity 1**)** $\forall\alpha$, $\rho(p;\alpha)$ *is weakly decreasing in* p *for* $p \in [0,\frac{1}{2}]$.[13]

Assumption 8 (Regularity 2**)** *There exists* α *such that* $\rho(p;\alpha) = 1$ *for all* $p \in [0,\frac{1}{2}]$.

Assumption 9 (Normalization) $\lim_{\alpha\to\infty}\underline{\rho}(\alpha) = \infty$ *and* $\lim_{\alpha\to 0}\overline{\rho}(\alpha) = 0$.

By Assumption 5, we focus on $G(p;\alpha)$ that is log-concave in $p \in [0,\frac{1}{2}]$. Assumption 7 together with Assumption 8 guarantees that the concavity/convexity of $G(\cdot)$ will not change for given α. Assumption 9 restates Definition 4 in the language of ρ-concavity.

5.2.3 Non-limit result of entrenchment (anti-entrenchment)

Since working on the signal space is equivalent to working on the posterior belief, board's expected profit can be rewritten in terms of cutoff of posterior beliefs. Denote $\tilde{\pi}(\hat{p})$ as the profit function in terms of cutoff posterior belief $\hat{p}$,

$$\tilde{\pi}(\hat{p}) = \frac{1}{4}\underbrace{\int_{\hat{p}}^{1} tg(t)dt}_{\text{incentive effect}}\left\{\underbrace{\frac{1}{2}G(\hat{p}) + \int_{\hat{p}}^{1} tg(t)dt}_{\text{selection effect}} + \underbrace{(\hat{p}-\frac{1}{2})\int_{0}^{\hat{p}} g(t)dt}_{\text{commitment cost effect}}\right\}.$$

[11] See Appendix C for more details.

[12] See Appendix C for detailed proof.

[13] As will be clear later, this assumption generates a well-behaved profit function for $p \in [0,\frac{1}{2}]$.

The profit function can be further simplified combining selection effect and commitment cost effect,

$$\tilde{\pi}(\hat{p}) = \frac{1}{4} \underbrace{\int_{\hat{p}}^{1} tg(t)dt}_{\text{incentive effect}} \left\{ \underbrace{\int_{\hat{p}}^{1} tg(t)dt \quad + \quad \hat{p}G(\hat{p})}_{\text{selection + commitment cost effect}} \right\}.$$

The expression of the total selection and commitment cost effect is intuitive. In equilibrium board's expected profit of replacement is equal to the expected profit created by the marginal incumbent manager with expected ability $\hat{p}$. Hence the board is replacing the incumbent manager of ability $p \leq \hat{p}$ with $\hat{p}$ taking into consideration of commitment cost. It can be verified that the total selection + commitment cost effect is increasing in $\hat{p}$ and thus maximized at $\hat{p} = 1$.

First order derivative with respect to $\hat{p}$ yields,

$$\tilde{\pi}'(\hat{p}) = \frac{1}{4}\left[-\hat{p}g(\hat{p})\left(1 - \int_{\hat{p}}^{1} G(t)dt\right) + G(\hat{p})\int_{\hat{p}}^{1} tg(t)dt\right].$$

$$\Longrightarrow \qquad \tilde{\pi}'(\hat{p}) \lesseqqgtr 0 \Longleftrightarrow \frac{\hat{p}g(\hat{p})}{G(\hat{p})} \gtreqqless \frac{\int_{\hat{p}}^{1} tg(t)dt}{1 - \int_{\hat{p}}^{1} G(t)dt}.$$

From the first order condition, $\hat{p}g(\hat{p})$ is the marginal incentive effect and $G(\hat{p})$ is the marginal selection + commitment cost effect. Whether profit in increasing or decreasing in $\hat{p}$ largely depends on the ratio between these two marginal effects, which is also the elasticity of $G(\cdot)$ at point $\hat{p}$. Since $\tilde{\pi}(1) = 0$, incentive effect dominates selection and commitment cost effect when $\hat{p}$ is close to 1. To relate ρ-concavity to the profit function, notice that $\frac{\hat{p}g(\hat{p})}{G(\hat{p})} = \left(\frac{\int_0^{\hat{p}} \rho(t)dt}{\hat{p}}\right)^{-1}$, which is the inverse of the average ρ-concavity of $G(\cdot)$ from 0 to $\hat{p}$. This ratio is weakly increasing if $\rho(p;\alpha)$ is weakly decreasing in p for $p \in [0, \frac{1}{2}]$ by Assumption 7. This assumption guarantees that marginal incentive effect changes faster than marginal selection + commitment cost effect and yields well-behaved profit function for $\hat{p} \in [0, \frac{1}{2}]$. Concave order (Assumption 6) guarantees that marginal selection + commitment cost effect changes faster than marginal incentive effect for given $\hat{p} \in [0, \frac{1}{2}]$ as α increases. Consequently, selection and commitment cost effect takes over as board's monitoring technology improves and anti-entrenchment is more likely to emerge.

Proposition 2 *Suppose the family of distribution* $\{G(\cdot;\alpha)\}$ *indexed by* $\alpha \in (0,\infty)$ *satisfies Assumption 5-9. Then there exists* α_1 *and* α_2 *such that* $\alpha_2 > \alpha_1$ *and,*

1. $\hat{s}^*(\alpha) = 0$ *for* $\alpha \in (0, \alpha_1]$*;*

2. $\hat{s}^*(\alpha) \in (0, \frac{1}{2})$ *for* $\alpha \in (\alpha_1, \alpha_2)$*;*

3. $\hat{s}^*(\alpha) \in (\frac{1}{2}, 1)$ *for* $\alpha \in (\alpha_2, \infty)$*,*

where α_1 *satisfies* $\rho(p;\alpha_1) = 1\ \forall\, p \in [0, \frac{1}{2}]$.

Proposition 2 characterizes optimal replacement policy for all α. Once anti-entrenchment is optimal for a certain informativeness level α', the optimal replacement policy never involves entrenchment with more informative information structure $\alpha > \alpha'$.

5.2.4 A tractable example

Example 1 *Suppose* $\mu = \frac{1}{2}$ *and* $G(p)$ *has the following functional form,*

$$G(p) = \begin{cases} \frac{1}{2}(2p)^{\frac{1}{\alpha}} & \textit{for} \quad \hat{p} \in [0, \frac{1}{2}] \\ 1 - \frac{1}{2}[2(1-p)]^{\frac{1}{\alpha}} & \textit{for} \quad \hat{p} \in (\frac{1}{2}, 1] \end{cases}.$$

1. *For* $\alpha \leq 1$*, optimal cutoff* $\hat{p}^* = 0$.

2. *For* $1 < \alpha < \frac{\sqrt{5}+1}{2}$*, optimal cutoff* $\hat{p}^* \in (0, \frac{1}{2})$.

3. *For* $\alpha > \frac{\sqrt{5}+1}{2}$*, optimal cutoff* $\hat{p}^* \in (\frac{1}{2}, 1)$.

Given $G(\cdot)$, the two corresponding conditional density functions are,

$$f_1(s) = \begin{cases} (2s)^{\alpha} & \text{for} \quad s \in [0, \frac{1}{2}] \\ 2 - \left[2(1-s)\right]^{\alpha} & \text{for} \quad s \in (\frac{1}{2}, 1] \end{cases}$$

and

$$f_0(s) = \begin{cases} 2 - (2s)^{\alpha} & \text{for} \quad s \in [0, \frac{1}{2}] \\ \left[2(1-s)\right]^{\alpha} & \text{for} \quad s \in (\frac{1}{2}, 1] \end{cases}.$$

Figure 2 shows the optimal cutoff with different informativeness of monitoring technology. In general, it is difficult to analyze the change of managerial turnover in informativeness α. However, under this special case, it is possible to do comparative static.

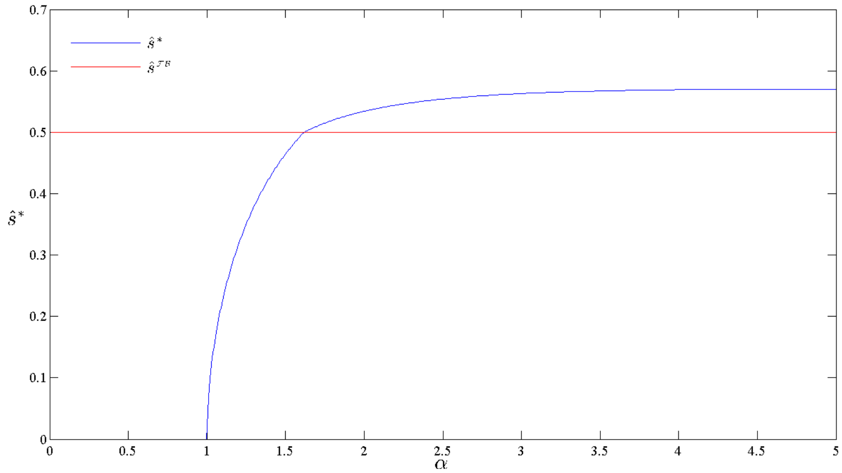

Figure 2: Optimal replacement policy

When $\mu = \frac{1}{2}$, turnover coincides with the cutoff $\hat{s}^*(\alpha)$ since $\frac{1}{2}F_1(\hat{s}^*) + \frac{1}{2}F_0(\hat{s}^*) = \hat{s}^*$. It can be proved that turnover is increasing for $\alpha \in \left[1, \frac{\sqrt{5}+1}{2}\right]$. The relationship between turnover and informativeness of board's monitoring technology is an inverted-U shape for $\alpha > \frac{\sqrt{5}+1}{2}$. When α is approaching infinity, the optimal cutoff is approaching $\frac{1}{2}$.

Figure 3 shows that severance pay in optimal contract is decreasing in informativeness of board's monitoring technology. When information structure is becoming more informative, it is easier for the board to obtain net commitment gain. Thus the board is less willing to provide commitment of not replacing the incumbent manager and the size of severance pay offered in the optimal contract decreases as a result. This generates a testable implication of the model: the size of severance package is decreasing in board's monitoring technology.

6 Discussion: Reasons of Entrenchment

The main result of optimal replacement policy stems from two important assumptions. The first assumption is that signal is non-contractible. The second assumption is that severance pay is constant with respect to outcome, i.e., the board can not provide performance-based severance pay. Lacking any one of these two assumptions, neither entrenchment nor anti-entrenchment emerges in optimal contract. The optimal

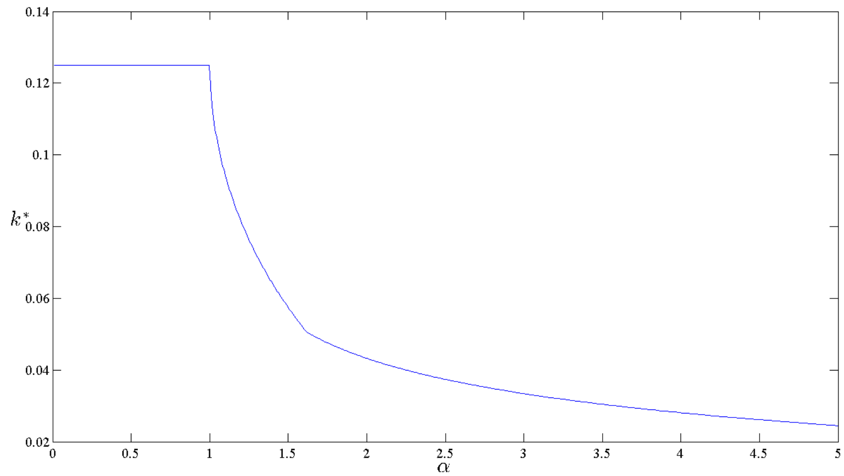

Figure 3: Severance pay in optimal contract

replacement policy is always $\hat{s}^* = \frac{1}{2}$.

6.1 Contractible signal

The non-contractibility assumption of board's signal is relaxed in this section. We maintain the assumption that severance pay is constant. A contract is fully characterized by $\{w(s), r(s), k(s)\}$, where $s \in [0, 1]$. $\{w(s), k(s)\}$ is the promised wage and severance pay after signal s. $r(s) \in [0, 1]$ specifies the retaining probability of the incumbent manager at signal s. In particular, $r(s) = 1$ indicates that the incumbent manager is retained while $r(s) = 0$ indicates that the incumbent is fired.[14]

Proposition 3 *Suppose signal is contractible and severance pay is constant with respect to outcome. Then $k^*(s) = 0$. Moreover, $r^*(s) = 1$ for $s \in [\frac{1}{2}, 1]$ and $r^*(s) = 0$ for $s \in [0, \frac{1}{2})$.*[15]

Allowing the board to contract on signals endows the board commitment power on retention decision at no cost. Severance pay as a costly commitment device, is no longer used in optimal contract as in the baseline model. As a result, $k^*(s) = 0$.

[14]Due to board's risk neutrality, randomization is not optimal except for the case that board is indifferent between retaining and firing the incumbent manager.

[15]It is assumed that the incumbent manager is retained with probability 1 if board is indifferent between replacement and retention.

For any effort level the board would like to induce, the board can design a contract without deviating from the socially optimal replacement cutoff. To see this, notice that incumbent manager is risk neutral and only cares about the expected wage. Thus, the board can incentivize the incumbent manger by increasing expected wage rate, which is determined by both the wage function $w(s)$ and the replacement policy $r(s)$. For given effort q and replacement policy $r(s)$, board can adjust the wage function $w(s)$ to induce q without changing $r(s)$. That is, the board can optimize effort and selection separately if signal is contractible, and the replacement cutoff is equal to $\frac{1}{2}$ in the optimal contract.

6.2 Performance-based severance pay

We maintain the assumption that signal is non-contractible and assume that the board can provide severance package on outcome after the incumbent's departure. Severance pay does not have to include a lump-sum payment. A contract is in the form of a triple (w_1, w_2, k). w_1 is the wage rate when the incumbent manager stays as in the baseline model. The tuple (w_2, k) constitutes a severance package. w_2 is the payment to the incumbent manager if he is forced out and $y = 1$. k is the constant severance pay as in the baseline model.

Proposition 4 *Suppose signal is non-contractible and the board can provide performance-based severance pay. Then* $k^* = 0$, $w_1^* = w_2^*$ *and* $\hat{s}^* = \frac{1}{2}$.

$k^* = 0$ under optimal contract. Constant severance pay is less effective to the board than performance-based severance pay when the incumbent manager is no longer in office because severance pay covers payment for low outcome and rewards failure. Thus, only performance-based severance pay is employed in the optimal contract.

Again the board has no incentive to deviate from the first best cutoff. Due to manager's risk neutrality, the effort choice of the incumbent manager is only determined by the expected wage. For a given effort level q the board wants to motivate, the expected wage is fixed, which is also the total cost to hire the incumbent manager. Since q is fixed, it remains to maximize the expected ability of the manager who stays in office in period 2. Hence the replacement cutoff stays at the first best in optimal contract.[16]

7 Extensions

Several assumptions that seem unrealistic are implicitly imposed to simplify the analysis in the baseline model. In this section, we modify the model to relax these assumptions separately. We show that the main result of optimal replacement policy is robust to these modifications. For ease of exposition, we assume $\mu = \frac{1}{2}$.

7.1 More effort vs. better selection

It is interesting to study if our main result on entrenchment (anti-entrenchment) remains correct when inducing effort becomes more important than selecting high ability manager. In the baseline model, the selection effect plays no role in determining entrenchment or anti-entrenchment at first glance since the selection effect always drives the optimal cutoff to $\frac{1}{2}$. This intuition is not correct for the reason that all three effects are nested and interact each other. When selection becomes less important, the benefit of inducing effort is becoming more important to the board's contractual problem. As a result, anti-entrenchment is less likely to emerge when selection becomes less important. We model this by either decreasing the variance of manager's pool or increasing the importance of effort relative to ability in the success probability.

To be specific, we assume the type space is $\theta \in \left\{\frac{1}{2} - \delta, \frac{1}{2} + \delta\right\}$, where $\delta \in \left(0, \frac{1}{2}\right]$ and the success probability is equal to $q^{1+\tau}\theta$. For manager's maximization problem to be well defined given quadratic cost function, we assume that $\tau \in (-1, 1)$. δ is a measure of the variance of manager's ability ex ante while τ is a measure of the relative importance of effort compared to selection. When $(\delta, \tau) = (0, 0)$, we are back to the baseline model.

Proposition 5 (Comparative static) *Suppose $\theta \in \left\{\frac{1}{2} - \delta, \frac{1}{2} + \delta\right\}$ and success probability is equal to $q^{1+\tau}\theta$, where $(\delta, \tau) \in (0, \frac{1}{2}] \times (-1, 1)$. Then*

[16] In practice, severance package usually comes in the form of combination of a lump-sum payment and stock option. This can be rationalized by the assumption that the manager is more risk averse than the board. If that is the case, optimal contract will involve some degree of lump-sum payment in response to risk sharing. Thanks to Navin Kartik for this useful comment.

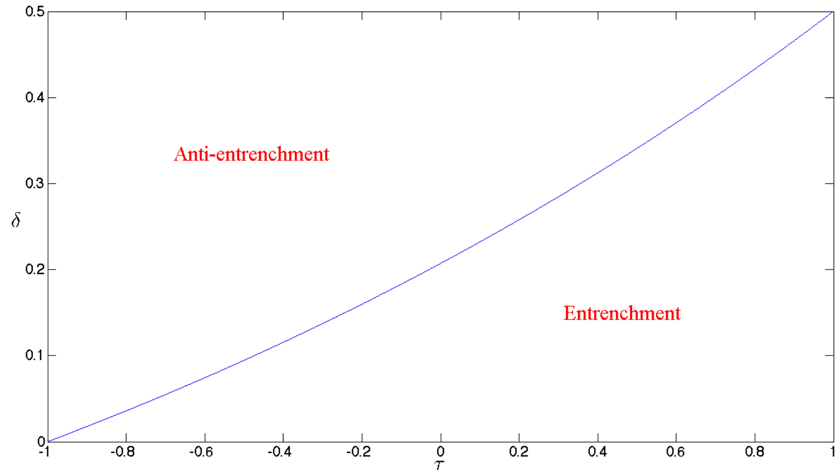

Figure 4: Optimal replacement policy when $\alpha \to \infty$

1. *if* $\delta > \frac{1}{2}^{\frac{1-\tau}{2}} - \frac{1}{2}$, *there exists* $\bar{\alpha}_A$ *such that anti-entrenchment is optimal for* $\alpha > \bar{\alpha}_A$*;*

2. *if* $\delta < \frac{1}{2}^{\frac{1-\tau}{2}} - \frac{1}{2}$, *there exists* $\bar{\alpha}_E$ *such that entrenchment is optimal for* $\alpha > \bar{\alpha}_E$*.*

7.2 Costly execution

When execution is costly, the outcome depends on the incumbent manager's effort q in period 1, effort e of the manager in period 2 as well as the ability of the manager in office in period 2. Effort q can be interpreted as project quality of the project selected by the incumbent manager and e can be interpreted as effort to execute the project.[17] Different from the baseline model, the board needs to offer a contract to the replacement manager after retention. Moreover, optimal contract to the incumbent manager balances the two dimensional moral hazard problem as well as the ability selection problem.

To formalize idea, we assume the success probability is equal to $\tilde{\theta}\left[(1-\lambda)q + \lambda\tilde{e}\right]$, where $\tilde{\theta} \in \{0,1\}$ is manager's ability in period 2, $q \in [0,1]$ is effort of the incumbent manager in period 1 and $\tilde{e} \in [0,1]$ is effort of the manager in period 2. By the specification, period 1 effort q and period 2 effort e are assumed to be substitutes. The parameter $\lambda \in [0,1]$ measures the importance of period 1 effort relative to period 2

[17] One can also interpret q as vision of the firm.

effort. When $\lambda = 0$ the model degenerates to the baseline model. After the board makes retention decision, the manager in office at the beginning of period 2 exerts effort e to execute the project. Cost function to the incumbent manger is assumed to be separable and quadratic, i.e., $C_i(q,e) = \frac{1}{2}q^2 + \frac{1}{2}e^2$. Cost function to the replacement is assumed to be $C_r(q,e) = C_i(0,e)$.

Lemma 4 (First best outcome with costly execution) *Suppose success probability is equal to $\tilde{\theta}\big[(1-\lambda)q + \lambda\tilde{e}\big]$ and the cost function to the incumbent and replacement manager is $C_i(q,e) = \frac{1}{2}q^2 + \frac{1}{2}e^2$ and $C_r(q,e) = C_i(0,e)$. Then the socially optimal cutoff is equal to $\frac{1}{2}$.*

The proof is similar to Lemma 1 and thus omitted. With costly execution, the first best replacement cutoff remains at $\frac{1}{2}$. In fact, this result is very general. The optimal cutoff is always $\frac{1}{2}$ as long as the marginal impact on manager's ability θ is positive.

The board provides two types of contracts, the one with severance package (w,k) to the incumbent manager and the one that only specifies wage w_r to the new manager. Denote variables with subscript r as the variables related to the replacement manager after retention.

We first calculate the optimal contract to the new manager w_r under replacement for a given belief of period 1 effort q. Given q and w_r, replacement manager chooses e_r to maximize:

$$\frac{1}{2}\big[(1-\lambda)q + \lambda e_r\big]w_r - \frac{1}{2}e_r^2 \Longrightarrow e_r = \frac{1}{2}\lambda w_r.$$

Noting that the new manager's effort on execution is independent of q. This is because q and e are assumed to be substitutes. The board chooses w_r to maximize:

$$\frac{1}{2}\big[(1-\lambda)q + \lambda e_r\big](1-w_r) \Longrightarrow w_r^* = \max\Big\{\frac{1}{2} - \frac{1-\lambda}{\lambda^2}q, 0\Big\}.$$

Since q and e are substitutes by assumption, the optimal wage to the new manager is weakly decreasing in the belief of q. When first period effort q is large or λ is small, board provides contract with $w_r = 0$ to the new manager.

Let $\underline{\pi}(q)$ be the board's expected profit under optimal contract after replacement. $\underline{\pi}(q)$ can be calculated as follows,

$$\underline{\pi}(q) = \begin{cases} \frac{1}{4}\left(\frac{1}{2}\lambda + \frac{1-\lambda}{\lambda}q\right)^2 & \text{for} \quad q \leq \frac{1}{2}\frac{\lambda^2}{1-\lambda} \\ \frac{1}{2}(1-\lambda)q & \text{for} \quad q > \frac{1}{2}\frac{\lambda^2}{1-\lambda} \end{cases}.$$

Next we calculate the equilibrium for a given contract (w, k) to the incumbent manager. For a fixed contract (w, k) and belief of cutoff $\hat{s}$, incumbent manager chooses (q, e) to maximize:

$$\frac{1}{2}[1-F_1(\hat{s})]\big[(1-\lambda)q+\lambda e\big]w+\frac{1}{2}[F_1(\hat{s})+F_0(\hat{s})]k-\frac{1}{2}q^2-\frac{1}{4}[(1-F_1(\hat{s}))+(1-F_0(\hat{s}))]e^2.$$

$$\Longrightarrow q = (1-\lambda)\frac{1-F_1(\hat{s})}{2}w \quad \text{and} \quad e = \lambda\frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}w.$$

q is decreasing in $\hat{s}$ while e is increasing in $\hat{s}$. A higher equilibrium replacement cutoff leads to lower first period effort q and higher second period effort e by the incumbent manager. First period effort q is decreasing in $\hat{s}$ for the reason that higher cutoff implies lower retaining probability in period 2 and dis-incentivizes the incumbent manager as in the baseline model. Conditional on the fact that the incumbent manager is retained, higher cutoff yields higher estimate on the incumbent's ability and thus the incumbent is willing to exert more effort in period 2 (θ and e are assumed to be compliments). As a result, e is increasing in equilibrium cutoff $\hat{s}$. Similar to Casamatta and Guembel (2010), the incumbent is easier to motivate, but for different reasons. Incumbent manager is easier to motivate in Casamatta and Guembel (2010) due to incumbent's reputational concern while in our model it is due to the incumbent's learning of his ability. For given wage rate w, the new manager chooses $e_r = \frac{1}{2}\lambda w$ while the incumbent manager chooses $e = \lambda\frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}w > \frac{1}{2}\lambda w$. The incumbent manager learns from stay that his ability is above average. Since ability and effort are assumed to be compliments, higher estimate of ability implies higher marginal return on effort. Thus, the incumbent manager exerts more effort in period 2 than the potential replacement manager given the same wage.

For a fixed contract (w, k) and belief on effort (q, e), the board chooses cutoff $\hat{s}$ to maximize:

$$\frac{1}{2}[1-F_1(\hat{s})]\big[(1-\lambda)q+\lambda e\big](1-w)+\frac{1}{2}[F_1(\hat{s})+F_0(\hat{s})]\left[\underline{\pi}(q)-k\right].$$

$$\Longrightarrow \frac{f_1(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})}\big[(1-\lambda)q+\lambda e\big](1-w)=\underline{\pi}(q)-k.$$

Note that not every $\hat{s}$ can be implemented. For instance, $\hat{s}$ very close to 1 cannot be induced by a contract. This is due to the limited liability assumption of severance pay. Extremely high cutoff can only be induced if severance pay is allowed to be negative. This is different from the baseline model. Using an aggressive replacement policy results in a small q, making board's outside option very unattractive. On the other hand, an aggressive replacement policy improves learning and makes e very high, increasing the value of keeping the current manager. Thus, unless the board is compensated by negative severance pay, a very aggressive replacement policy cannot be induced by a contract subject to limited liability constraint.

Similar to the baseline model, the expected profit can be written as a function of cutoff $\hat{s}$ alone assuming away limited liability constraint of k,[18]

$$\pi(\hat{s})=\frac{1}{8}\Bigg\{(1-\lambda)^2\underbrace{\frac{1-F_1(\hat{s})}{2}}_{\text{incentive effect}}+\lambda^2\underbrace{\frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}}_{\text{learning effect}}\Bigg\}$$
$$\cdot\Bigg\{\underbrace{\Big[[1-F_1(\hat{s})]+\frac{1}{2}\big[F_1(\hat{s})+F_0(\hat{s})\big]\Big]}_{\text{selection effect}}+\underbrace{\Big[F_1(\hat{s})+F_0(\hat{s})\Big]\Big(\frac{f_1(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})}-\frac{1}{2}\Big)}_{\text{commitment cost effect}}\Bigg\}.$$

Proposition 6 (Optimal replacement policy with costly execution) *Suppose success probability is equal to $\tilde{\theta}\big[(1-\lambda)q+\lambda\tilde{e}\big]$ and the cost function to the incumbent and replacement manager is $C_i(q,e)=\frac{1}{2}q^2+\frac{1}{2}e^2$ and $C_r(q,e)=C_i(0,e)$.*

1. *If $\lambda\in[0,\sqrt{2}-1)$, there exists $\overline{\alpha}$ such that $\hat{s}^*(\alpha)>\frac{1}{2}$ for $\alpha>\overline{\alpha}$;*
2. *For $\lambda\in[0,1]$, there exists $\underline{\alpha}$ such that $\hat{s}^*(\alpha)<\frac{1}{2}$ for $\alpha<\underline{\alpha}$.*

Learning effect enters into board's profit function besides the three aforementioned effects. When information structure is sufficiently noisy ($\alpha\to 0$), the incumbent's learning is very slow for all $s\in(0,1)$. Learning effect plays little role in determing the optimal replacement policy since $\frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}\lambda^2$ can be considered as a constant. Thus, entrenchment is expected to be optimal when α is sufficiently small independent

[18]The intuition can be clearly illustrated assuming that every $\hat{s}$ can be induced. The limited liability constraint of severance pay is taken into consideration in the proof of Proposition 6.

of the size of λ.

When information structure is sufficiently informative ($\alpha \to \infty$), the incumbent's learning becomes very fast. When execution becomes sufficiently important, entrenchment can be optimal to the board. When period 1 effort q is sufficiently important relative to period 2 effort e (i.e., $\lambda < \sqrt{2} - 1$), incentive effect is more important than learning effect in board's contractual problem. Thus, the main insight in the baseline model follows through and entrenchment is expected to emerge in optimal contract.

7.3 Signal of outcome instead of ability

In the baseline model, it is assumed that the board observes a signal of the incumbent manager's ability rather than the outcome under the incumbent's management. Since signal is not related to effort by the incumbent manager, the incumbent cannot increase the retaining probability by exerting more effort. When the board receives a signal related to effort, the incumbent manager is able to increasing the retaining probability by exerting more effort.

To formalize idea, for outcome $y \in \{0, 1\}$, signal s is drawn from distribution with density $h_y(\cdot)$ and cdf $H_y(\cdot)$. Similarly, we assume $\big\{h_1(\cdot), h_0(\cdot)\big\}$ satisfies assumption 1 - 4. The board is able to infer the expected outcome as well as incumbent manager's ability from the signal.

The social planner chooses $(\hat{s}, q)$ to maximize,

$$\max_{\{\hat{s},q\}} \ \frac{1}{2}q\big[1 - H_1(\hat{s})\big] + \frac{1}{2}q\Big[\frac{1}{2}qH_1(\hat{s}) + \big(1 - \frac{1}{2}q\big)H_0(\hat{s})\Big] - \frac{1}{2}q^2.$$

Lemma 5 (First best outcome when signal is of outcome) *If board observes signal of outcome instead of incumbent's ability, $\hat{s}^{\mathcal{FB}} = \frac{1}{2}$ and $q^{\mathcal{FB}} = \frac{1+H_0(\frac{1}{2})-H_1(\frac{1}{2})}{2+H_0(\frac{1}{2})-H_1(\frac{1}{2})}$ in the first best outcome.*[19]

Given effort level q, the Bayesian update of the incumbent manager's ability is,

$$\varphi(\hat{s}, q) = \frac{\frac{1}{2}qh_1(\hat{s}) + \frac{1}{2}(1-q)h_0(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \Big(1 - \frac{1}{2}q\Big)\,h_0(\hat{s})}.$$

[19]The proof is similar to Lemma 1 and omitted.

It can be verified that $\varphi(\frac{1}{2}, q) = \frac{1}{2}$ independent of q and α. Thus it is socially optimal to replace the incumbent manger if and only if the posterior belief of incumbent's ability falls below the prior.

Given a contract (w, k) and belief of replacement cutoff $\hat{s}$, manager chooses q to maximize:

$$\frac{1}{2}q[1 - H_1(\hat{s})]w + \left\{\frac{1}{2}qH_1(\hat{s}) + \left(1 - \frac{1}{2}q\right) H_0(\hat{s})\right\} k - \frac{1}{2}q^2.$$

The best response of manager is:

$$q = \max\left\{\frac{1}{2}[1 - H_1(\hat{s})]w - \frac{1}{2}[H_0(\hat{s}) - H_1(\hat{s})]k, 0\right\}.$$

Different from the baseline model, the incumbent manager is directly dis-incentivized by severance pay. An increase in severance pay increases the opportunity cost of exerting effort and leads to a decrease in effort directly. If the severance pay is high enough, the incumbent manager exerts no effort at all and is willing to be fired. Under this extension, severance pay is a double-edged sword. By the direct effect (better outside option if the incumbent manager is replaced), effort decreases. By the indirect effect (better job security with lower equilibrium replacement cutoff), effort increases. The design of the optimal contract should take this non-trivial incentive of k on q into consideration.

Fixed (w, k) and q, the board chooses $\hat{s}$ to maximize:

$$\frac{1}{2}q\big[1 - H_1(\hat{s})\big](1 - w) + \left\{\frac{1}{2}qH_1(\hat{s}) + \left(1 - \frac{1}{2}q\right) H_0(\hat{s})\right\} \left(\frac{1}{2}q - k\right).$$

Board's indifference condition is:

$$\frac{\frac{1}{2}qh_1(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \big(1 - \frac{1}{2}q\big)h_0(\hat{s})}(1 - w) = \frac{1}{2}q - k.$$

$\hat{s}$ is the solution to $\zeta(\hat{s}, q) = \max\left\{ \min\left\{\frac{\frac{1}{2}q-k}{1-w}, 1\right\}, 0\right\}$, where $\zeta(\hat{s}, q)$ is the estimate of the outcome under incumbent's management at $\hat{s}$ given q,

$$\zeta(\hat{s}, q) \equiv \frac{\frac{1}{2}qh_1(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \left(1 - \frac{1}{2}q\right) h_0(\hat{s})}.$$

It is difficult to write down the expected profit as a function of $\hat{s}$ alone since now q is affected by (w, k) directly and equilibrium cutoff $\hat{s}$ indirectly. However, we can still discuss the optimal replacement policy under extreme information structure.

Multiple equilibria may exist for some contract (w, k) since incentive on effort q is not monotone in k as in the baseline model. For the same reason, equilibria may not be Pareto ranked. We further assume that the equilibrium most favorable to the board is selected when multiple equilibria exist.

Proposition 7 (Optimal replacement policy when signal is of outcome) *Suppose the board observes a signal of outcome rather than the incumbent's ability. Then there exists* $\overline{\alpha}$ *and* $\underline{\alpha}$ *such that,*

1. $\hat{s}^*(\alpha) > \frac{1}{2}$ *for* $\alpha > \overline{\alpha}$*;*
2. $\hat{s}^*(\alpha) < \frac{1}{2}$ *for* $\alpha < \underline{\alpha}$*.*

When information structure is sufficiently noisy ($\alpha \to 0$), $H_0(s)$ is very close to $H_1(s)$ for $s \in [0, 1]$. The direct negative effect of severance pay on effort is small and the model is back to the baseline in the limit. Knowing that the board has noisy monitoring technology, the incumbent manager has little incentive to manipulate the realization of the signal. Entrenchment is expected to be optimal when α is sufficiently small.

When information structure is sufficiently informative ($\alpha \to \infty$), the magnitude of the direct negative effect of k (i.e., $\frac{1}{2}[H_0(\hat{s}) - H_1(\hat{s})]$) is very large. Under this scenario, the board can simply avoid the disadvantage of k by setting $k = 0$. Moreover, this does not contradict with obtaining net commitment gain. In fact, we can construct a contract with high wage and zero severance pay that yields anti-entrenchment and dominates all possible contracts that yield entrenchment. Thus, anti-entrenchment emerges in optimal contract as board's information structure is sufficiently informative.

8 Conclusion

This paper explores how the problem of motivating the incumbent manager to exert effort and keeping the flexibility on selecting the high ability manager interacts with equilibrium replacement policy. We focus on the situation where board observes a non-contractable signal after the incumbent manager exerts effort and solve for the optimal contract. We show that the boards information technology of assessing the incumbent managers ability is an important determinant of the optimal contract and managerial turnover. Different from the existing literature on managerial turnover, which aims to rationalize entrenchment, we show that anti-entrenchment can also be optimal for shareholders under some situations. This result is robust to allowing costly execution and the possibility that the board observes a signal of outcome rather than incumbent manager's ability. The model highlights the board's monitoring technology as an important determinants of managerial turnover.

There are several interesting questions that can be pursued using the stylized model introduced in this paper. For future research, it would be interesting to endogenize the informativeness of board's monitoring technology. In practice, informativeness is often the choice of the board. Some boards actively monitor their CEOs while some tends to be passive monitors. Endogenizing board's monitoring technology can help us better understand the differences of monitoring intensities across industries.

Another intriguing research avenue would be to incorporate voluntary departure into the model by allowing the possibility that manager possesses private information of firm's profit. As Inderst and Mueller (2010) point out, manager sometimes have private information about firm's performance. In such scenario, optimal contract needs to provide incentives for the incumbent manager to step down voluntarily. It would be interesting to build a unified model with both forced departure and voluntary departure and study the interactions in between.

Reference

Almazan, Andres and Javier Suarez (2003), "Entrenchment and severance pay in optimal governance structures." *Journal of Finance*, 58, 519–547.

Anderson, Simon P and Régis Renault (2003), "Efficiency and surplus bounds in cournot competition." *Journal of Economic Theory*, 113, 253–264.

Casamatta, Catherine and Alexander Guembel (2010), "Managerial legacies, entrenchment, and strategic inertia." *The Journal of Finance*, 65, 2403–2436.

Hermalin, Benjamin E and Michael S Weisbach (1998), "Endogenously chosen boards of directors and their monitoring of the ceo." *American Economic Review*, 88, 96–118.

Huang, Peggy (2011), "Marital prenups? a look at ceo severance agreements." *Available at SSRN 1928264*.

Inderst, Roman and Holger M Mueller (2010), "Ceo replacement under private information." *Review of Financial Studies*, 23, 2935–2969.

Johnson, Justin P and David P Myatt (2006), "On the simple economics of advertising, marketing, and product design." *The American Economic Review*, 96, 756–784.

Kaplan, Steven N and Bernadette A Minton (2012), "How has ceo turnover changed?" *International Review of Finance*, 12, 57–87.

Laux, Volker (2008), "Board independence and ceo turnover." *Journal of Accounting Research*, 46, 137–171.

Laux, Volker (2014), "Corporate governance, board oversight, and ceo turnover." *Foundations and Trends in Accounting, Forthcoming.*

Manso, Gustavo (2011), "Motivating innovation." *The Journal of Finance*, 66, 1823–1860.

Mares, Vlad and Jeroen M Swinkels (2014), "On the analysis of asymmetric first price auctions." *Journal of Economic Theory*, 152, 1–40.

Rusticus, Tjomme O (2006), *Executive severance agreements*. Ph.D. thesis, University of Pennsylvania.

Shleifer, Andrei and Robert W Vishny (1989), "Management entrenchment: The case of manager-specific investments." *Journal of financial economics*, 25, 123–139.

Taylor, Curtis R and Huseyin Yildirim (2011), "Subjective performance and the value of blind evaluation." *The Review of Economic Studies*, 78, 762–794.

Taylor, Lucian A (2010), "Why are ceos rarely fired? evidence from structural estimation." *The Journal of Finance*, 65, 2051–2087.

Weyl, E Glen and Michal Fabinger (2013), "Pass-through as an economic tool: principles of incidence under imperfect competition." *Journal of Political Economy*, 121, 528–583.

Appendix A: Proofs of the propositions

Proof of Lemma 1. The first best outcome is the solution to the following maximization problem,

$$\max_{\{\hat{s},q\}} \mu\big[1 - F_1(\hat{s})\big]q + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\mu q - \frac{1}{2}q^2.$$

First order condition with respect to $\hat{s}$ yields:

$$f_1(\hat{s}) = f_0(\hat{s}) \Longrightarrow \hat{s}^{\mathcal{FB}} = \frac{1}{2}.$$

First order condition with respect to q yields:

$$q^{\mathcal{FB}} = \mu\Big\{1 + (1-\mu)\big[F_0(\hat{s}^{\mathcal{FB}}) - F_1(\hat{s}^{\mathcal{FB}})\big]\Big\}.$$

■

Proof of Proposition 1. It is useful to first prove the two lemmas.

Lemma A1 (Uniform convergence of $F_1(\cdot)$ when $\alpha \to \infty$) *For given $\epsilon > 0$, there exists N such that for $\alpha > N$, $F_1(s;\alpha) < \epsilon$ for $s \in [0, \frac{1}{2}]$ and $F_1(s;\alpha) < (2s-1) + \epsilon$ for $s \in [\frac{1}{2}, 1]$.*

Proof. By the definition of completely informative information structure, given $\epsilon' = \frac{1}{2}\epsilon$ and $\Delta \in (0, \frac{1}{2})$, there exist N such that $f_1(\Delta;\alpha) < \epsilon'$ for $\alpha > N$. Thus,

$$F_1\Big(\frac{1}{2};\alpha\Big) = \int_0^{\frac{1}{2}} f_1(t;\alpha)dt = \int_0^{\Delta} f_1(t;\alpha)dt + \int_{\Delta}^{\frac{1}{2}} f_1(t;\alpha)dt \le \Delta\epsilon' + \Big(\frac{1}{2} - \Delta\Big).$$

Let $\Delta = \frac{1}{2} - \epsilon'$. $F_1\Big(\frac{1}{2};\alpha\Big)$ can be bounded from above by,

$$F_1(s;\alpha) \le F_1\Big(\frac{1}{2};\alpha\Big) \le \epsilon'\Big(\frac{1}{2} - \epsilon'\Big) + \epsilon' < 2\epsilon' = \epsilon \text{ for } s \in [0, \frac{1}{2}].$$

Similarly, for all $s \in [\frac{1}{2}, 1]$,

$$F_1(s;\alpha) = 2s - F_0(s;\alpha) = (2s-1) + F_1(1-s;\alpha) < (2s-1) + \epsilon \text{ for } \alpha > N.$$

■

Lemma A2 (Uniform convergence of $F_1(\cdot)$ when $\alpha \to 0$) *For given $\epsilon > 0$, there exists N such that for $\alpha < N$, $F_1(s;\alpha) > s - \epsilon$ for all $s \in [0,1]$.*

Proof. By the definition of completely uninformative information structure, for given ϵ and $\Delta \in (0, \frac{1}{2})$, there exists N such that $f_1(\delta;\alpha) > 1-\epsilon$. Thus,

$$\begin{aligned} s - F_1(s;\alpha) = \int_0^s \big[1 - f_1(t;\alpha)\big] dt &\leq \int_0^{\frac{1}{2}} \big[1 - f_1(t;\alpha)\big] dt \\ &= \int_0^{\Delta} \big[1 - f_1(t;\alpha)\big] dt + \int_{\Delta}^{\frac{1}{2}} \big[1 - f_1(t;\alpha)\big] dt \\ &\leq \Delta + \epsilon(\frac{1}{2} - \Delta) \text{ for } s \in [0, \frac{1}{2}]. \end{aligned}$$

Let $\Delta = \frac{1}{2}\epsilon$. $s - F_1(s;\alpha)$ can be bounded from above by,

$$s - F_1(s;\alpha) \leq \Delta + \epsilon(\frac{1}{2} - \Delta) < \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon \text{ for } s \in [0, \frac{1}{2}].$$

Similarly, for all $s \in [\frac{1}{2}, 1]$,

$$s - F_1(s;\alpha) = s - \big[2s - F_0(s;\alpha)\big] = (1-s) - F_1(1-s;\alpha) < \epsilon.$$

■

Recall that the expected profit function is,

$$\pi(\hat{s}) = \frac{1}{4}\mu\big[1 - F_1(\hat{s})\big] \left\{ \mu\big[1 - F_1(\hat{s})\big] + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big] \frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})} \right\}.$$

By Assumption 3, $f_0(s) = f_1(1-s)$ and $F_0(s) = 1 - F_1(1-s)$. The expected profit function can be written as,

$$\pi(\hat{s}) = \frac{1}{4}\mu\big[1 - F_1(\hat{s})\big] \left\{ \mu\big[1 - F_1(\hat{s})\big] + \big[(2\mu-1)F_1(\hat{s}) + 2(1-\mu)\hat{s}\big] \frac{\mu f_1(\hat{s})}{2(1-\mu) + (2\mu-1) f_1(\hat{s})} \right\}.$$

1. Anti-entrenchment

   (a) $\mu \geq \frac{1}{2}$.

k is non-negative for all $\hat{s} \in [0,1]$ given $w = \frac{1}{2}$. Too see this, note that

$$k(\hat{s}, q) = \mu q - \frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})} q(1-w) \geq \mu q - \frac{1}{2} q \geq 0.$$

By Lemma A1, for any ϵ, there exists N such that for $\alpha > N$,

$$\pi(\hat{s}; \alpha) < \frac{1}{4}\mu\Big(\mu + \mu\big[|2\mu - 1|\,\epsilon + 2(1-\mu)\hat{s}\big]\Big) < \frac{1}{4}\mu^2(2-\mu) + \frac{1}{4}\mu^2|2\mu-1|\,\epsilon \text{ for all } \hat{s} \in [0, \frac{1}{2}].$$

Meanwhile, given $\hat{s} \in (\frac{1}{2}, 1)$ and ϵ', there exists N' such that $\frac{\mu f_1(\hat{s})}{2(1-\mu)+(2\mu-1)f_1(\hat{s})} > 1 - \epsilon'$ for $\alpha > N'$.

Let $\overline{\alpha} = \max\{N, N'\}$. Then for $\alpha > \overline{\alpha}$,

$$\pi(\hat{s}; \alpha) > \frac{1}{4}\mu^2\big(2 - 2\hat{s} - \epsilon\big)\Big\{(2 - 2\hat{s} - \epsilon) + 2\frac{1-\mu}{\mu}\hat{s}(1-\epsilon')\Big\}.$$

Let $\hat{s} = \frac{1}{2}(1+\epsilon)$ and $\epsilon' = \frac{\epsilon}{1+\epsilon}$. Then,

$$\begin{aligned}\pi\Big(\frac{1}{2}(1+\epsilon); \alpha\Big) &> \frac{1}{4}\mu^2(1-2\epsilon)\Big\{(1-2\epsilon) + \frac{1-\mu}{\mu}(1+\epsilon)(1-\epsilon')\Big\} \\ &= \frac{1}{4}\mu^2(1-2\epsilon)\Big\{(1-2\epsilon) + \frac{1-\mu}{\mu}\Big\}.\end{aligned}$$

To prove the proposition, it suffices to find ϵ such that

$$\frac{1}{4}\mu^2(1-2\epsilon)\Big\{(1-2\epsilon) + \frac{1-\mu}{\mu}\Big\} \geq \frac{1}{4}\mu^2(2-\mu) + \frac{1}{4}\mu^2|2\mu-1|\,\epsilon.$$

Noting that the left hand side is decreasing in ϵ while the right hand side is increasing in ϵ. Thus it suffices to prove,

$$\frac{1}{\mu} > (2-\mu) \Longleftrightarrow (1-\mu)^2 > 0.$$

(b) $\mu < \frac{1}{2}$.

$\hat{s} \leq \frac{1}{2}$ can be induced by a contract with $w = \frac{1}{2}$. Thus, the analysis for $\hat{s} \in [0, \frac{1}{2}]$ is the same as $\mu \geq \frac{1}{2}$.

For $\hat{s} > \frac{1}{2}$, consider a specific contract $(w', k') = (\frac{1}{2}, 0)$. The equilibrium cutoff $\hat{s}(w', k')$ is determined by

$$\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})} = 2\mu.$$

Since $\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s})+(1-\mu)f_0(\hat{s})} > \mu$, $\hat{s}(w', k')$ is strictly above $\frac{1}{2}$ and anti-entrenchment emerges under this contract.

Lemma A3 *For $\Delta \in (0, \frac{1}{2})$, there exists N such that for $\alpha > N$, $\hat{s}(\alpha) < \frac{1}{2} + \Delta$ with contract $(w', k') = (\frac{1}{2}, 0)$.*

Proof. It suffices to prove that for $\Delta \in (0, \frac{1}{2})$, there exists N such that for $\alpha > N$, board's indifference condition never holds for all $\hat{s} \in [\frac{1}{2} + \Delta, 1]$ with contract $(w', k') = (\frac{1}{2}, 0)$.

Board's indifference condition can be simplified as,

$$\frac{f_1(\hat{s})}{f_0(\hat{s})} = \frac{2(1-\mu)}{1-2\mu}.$$

$\frac{f_1(\hat{s};\alpha)}{f_0(\hat{s};\alpha)}$ is approaching infinity as $\alpha \to \infty$ while $\frac{2(1-\mu)}{1-2\mu}$ is bounded, which is a contradiction. ■

Board's expected profit under this contract is,

$$\begin{aligned}
&\pi(\hat{s}(w', k'); \alpha) \\
&= \frac{1}{4}\mu\big[1 - F_1(\hat{s})\big] \left\{\mu\big[1 - F_1(\hat{s})\big] + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big] \frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu) f_0(\hat{s})}\right\} \\
&= \frac{1}{4}\mu^2\big[1 - F_1(\hat{s})\big] \left\{\big[1 - F_1(\hat{s})\big] + 2\big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\right\}.
\end{aligned}$$

It can be verified that $\pi(\hat{s}(w', k'); \alpha)$ is approaching $\frac{1}{4}\mu^2(3 - 2\mu)$ as $\alpha \to \infty$.

To prove anti-entrenchment, it suffices to prove,

$$\frac{1}{4}\mu^2(3 - 2\mu) > \frac{1}{4}\mu^2(2 - \mu) \Longleftrightarrow \mu < 1.$$

2. Entrenchment

   It suffices to prove that there exists $\underline{\alpha}$ such that for $\alpha < \underline{\alpha}$ and, $\pi(\hat{s}) < \pi(0) = \frac{1}{4}\mu^2$ for all $\hat{s} \in [\frac{1}{2}, 1]$.

   Since $f_1(s) < 2$ for $s \in [0, 1)$ by normalization, it directly follows that $1 - F_1(s) <$

$2(1-s)$. Thus, $1 - F_1(1-\Delta) < 2\Delta$ for $\Delta \in (0, \frac{1}{2})$.

$$\Longrightarrow \pi(\hat{s}) < \frac{1}{2}\mu\Delta(2\mu\Delta + 1) \text{ for } \hat{s} \in [1-\Delta, 1].$$

For Δ to be sufficently small, $\frac{1}{2}\mu\Delta(2\mu\Delta+1) < \frac{1}{4}\mu^2$. In particular, let $\Delta(\mu) = \frac{\sqrt{4\mu^2+1}-1}{4\mu}$. Then $\hat{s} \in [1-\Delta(\mu), 1]$ can not be optimal replacement policy.

It remains to prove that there extist $\underline{\alpha}$ such that for $\alpha < \underline{\alpha}$, $\pi(\hat{s}) < \pi(0)$ for all $\hat{s} \in [\frac{1}{2}, 1-\Delta(\mu)]$. By the definition of completely uninformative information structure, for any ϵ', there exists N' such that $\frac{\mu f_1(\hat{s})}{2(1-\mu)+(2\mu-1)f_1(\hat{s})} < \mu + \epsilon'$ for $\hat{s} \in [\frac{1}{2}, 1-\Delta]$ and $\alpha < N'$.

By Lemma A2, for any ϵ, there exists N such that for $\alpha < N$, $F_1(s;\alpha) > s - \epsilon$ for $s \in [0,1]$. Thus,

$$\pi(\hat{s}) < \frac{1}{4}\mu(1-\hat{s}+\epsilon)\Big[\mu(1-\hat{s}+\epsilon) + (\mu+\epsilon')\Big].$$

Let $\epsilon = \frac{\sqrt{3}}{3} - \frac{1}{2}$, $\epsilon' = 2\mu\epsilon$ and $\underline{\alpha} = \max\{N, N'\}$. Then for $\alpha < \underline{\alpha}$,

$$\pi(\hat{s}) < \frac{1}{4}\mu^2(1-\hat{s}+\epsilon)(2-\hat{s}+3\epsilon) \leq \frac{3}{4}\mu^2\big(\frac{1}{2}+\epsilon\big)^2 = \frac{1}{4}\mu^2.$$

■

Proof of Lemma 2. For existence, it suffices to construct an example. Suppose two conditional density functions $\{\tilde{f}_1(x), \tilde{f}_0(x)\}$ with support $x \in \mathcal{X} = [0,1]$ induce $g(\cdot)$. We impose the following assumption on $\{\tilde{f}_1(\cdot), \tilde{f}_0(\cdot)\}$,

$$\mu\tilde{F}_1(x) + (1-\mu)\tilde{F}_0(x) = x \quad \text{for } x \in [0,1] \Longleftrightarrow \mu\tilde{f}_1(x) + (1-\mu)\tilde{f}_0(x) = 1 \quad \text{for } x \in [0,1].$$

It simply requires that unconditional distribution of x is a uniform distribution on $[0,1]$ given prior μ. Meanwhile, we have

$$g(p) = \big[\mu\tilde{f}_1(\varphi^{-1}(p)) + (1-\mu)\tilde{f}_0(\varphi^{-1}(p)))\big]\frac{d\varphi^{-1}(p)}{dp}.$$

Thus, $g(p)dp = d\varphi^{-1}(p) \Longrightarrow \varphi(G(p)) = p \Longrightarrow \tilde{f}_1(x) = \frac{1}{\mu}G^{-1}(x)$ and $\tilde{f}_0(x) = \frac{1}{1-\mu}[1-G^{-1}(x)]$.

Next, define a new signal s as $s = \frac{1}{2}\tilde{F}_1(x) + \frac{1}{2}\tilde{F}_0(x)$ with conditional density functions $\{f_1(\cdot), f_0(\cdot)\}$. It follows directly that $\frac{1}{2}F_1(s) + \frac{1}{2}F_0(s) = s$, which satisfies the assumption on information structure. This finishes the proof of existence.

For uniqueness, suppose two information structures $\{f_1(s), f_0(s)\}$ and $\{f_1^\dagger(s), f_1^\dagger(s)\}$ induce the same $g(p,\mu)$. By the definition of the information structure,

$$\frac{1}{2}f_1^\dagger(s) + \frac{1}{2}f_0^\dagger(s) = 1 = \frac{1}{2}f_1(s) + \frac{1}{2}f_0(s).$$

By the definition of p,

$$\mu f_1(s) + (1-\mu)f_0(s) = \frac{\mu f_1(s)}{p} = \frac{(1-\mu)f_0(s)}{1-p}.$$

By the derivation of $g(p,\mu)$,

$$g(p,\mu) = \Big[\mu f_1(\varphi^{-1}(p,\mu)) + (1-\mu)f_0(\varphi^{-1}(p,\mu))\Big]\frac{\partial\varphi^{-1}(p,\mu)}{\partial p}.$$

$$\Longrightarrow pg(p,\mu) = \mu f_1(\varphi^{-1}(p,\mu))\frac{\partial\varphi^{-1}(p,\mu)}{\partial p}.$$

$$\Longrightarrow \int_0^p tg(t,\mu)dt = \mu F_1(\varphi^{-1}(p,\mu)) = \mu F_1^\dagger(\varphi^{\dagger-1}(p,\mu)).$$

Similarly,

$$(1-p)g(p,\mu) = (1-\mu)f_0(\varphi^{-1}(p,\mu))\frac{\partial\varphi^{-1}(p,\mu)}{\partial p}.$$

$$\Longrightarrow \int_0^p (1-t)g(t,\mu)dt = (1-\mu)F_0(\varphi^{-1}(p,\mu)) = (1-\mu)F_0^\dagger(\varphi^{\dagger-1}(p,\mu)).$$

Thus,

$$\frac{1}{2}F_1^\dagger(\varphi^{\dagger-1}(p,\mu)) + \frac{1}{2}F_0^\dagger(\varphi^{\dagger-1}(p,\mu)) = \frac{1}{2}F_1(\varphi^{-1}(p,\mu)) + \frac{1}{2}F_0(\varphi^{-1}(p,\mu)).$$

$$\Longrightarrow \varphi^{\dagger-1}(p,\mu) = \varphi^{-1}(p,\mu) \Longrightarrow f_1^\dagger(s) = f_1(s).$$

Since $\frac{1}{2}f_1^\dagger(s) + \frac{1}{2}f_0^\dagger(s) = \frac{1}{2}f_1(s) + \frac{1}{2}f_0(s)$, it follows directly that $f_0^\dagger(s) = f_0(s)$. This finishes the proof of uniqueness. ■

Proof of Lemma 3. By definition $\underline{\rho} \leq 1 - \frac{G(p)g'(p)}{g^2(p)} \leq \overline{\rho}$. Integrating both sides from

0 to p yields,

$$\underline{\rho} p \leq \frac{G(p)}{g(p)} - \frac{G(0)}{g(0)} \leq \overline{\rho} p.$$

$$\Longrightarrow \frac{1}{\overline{\rho}} \frac{1}{p} \leq \frac{g(p)}{G(p)} \leq \frac{1}{\underline{\rho}} \frac{1}{p} \Longleftrightarrow \frac{1}{\overline{\rho}} \leq \frac{pg(p)}{G(p)} \leq \frac{1}{\underline{\rho}}.$$

Integrating both sides from p to $\frac{1}{2}$ yields,

$$\frac{1}{2}(2p)^{\frac{1}{\underline{\rho}}} \leq G(p) \leq \frac{1}{2}(2p)^{\frac{1}{\overline{\rho}}}.$$

■

Proof of Proposition 2.

Lemma A4 *If $G(p) \leq p$ for $p \in [0, \frac{1}{2}]$, entrenchment is optimal to the board. Moreover, if $G(p)$ is convex in p for $p \in [0, \frac{1}{2}]$, $\hat{p}^* = 0$.*

Proof. We finish the proof by two steps:

1. $\tilde{\pi}(1 - \hat{p}) < \tilde{\pi}(0)$ for $\hat{p} \in [0, \frac{1}{2}]$.

   It is equivalent to prove,

$$\int_{1-\hat{p}}^{1} tg(t)dt \bigg(1 - \int_{1-\hat{p}}^{1} G(t)dt\bigg) < \int_{0}^{1} tg(t)dt \bigg(1 - \int_{0}^{1} G(t)dt\bigg).$$

   Since $G(1 - \hat{p}) = 1 - G(\hat{p})$, $\int_0^1 G(t)dt = \frac{1}{2}$. Thus, right hand side can be further simplified as,

$$\int_{0}^{1} tg(t)dt \bigg(1 - \int_{0}^{1} G(t)dt\bigg) = \bigg(1 - \int_{0}^{1} G(t)dt\bigg)^2 = \frac{1}{4}.$$

For the left hand side,

$$
\begin{aligned}
&\int_{1-\hat{p}}^{1} tg(t)dt\Big(1-\int_{1-\hat{p}}^{1} G(t)dt\Big)\\
&=\Big(1-\int_{1-\hat{p}}^{1} G(t)dt-(1-\hat{p})G(1-\hat{p})\Big)\Big(1-\int_{1-\hat{p}}^{1} G(t)dt\Big)\\
&<\Big(1-\int_{0}^{\hat{p}}(1-G(t))dt-\frac{1}{2}(1-\hat{p})[1-G(\hat{p})]\Big)^2\\
&=\Big(\frac{1-\hat{p}}{2}(1+G(\hat{p}))+\int_{0}^{\hat{p}} G(t)dt\Big)^2\\
&\leq\Big(\frac{1-\hat{p}}{2}(1+\hat{p})+\int_{0}^{\hat{p}} tdt\Big)^2=\frac{1}{4}.
\end{aligned}
$$

2. $\tilde{\pi}(\hat{p})$ is strictly decreasing in $\hat{p}$ for $\hat{p}\in[0,\frac{1}{2}]$ if $G(p)$ is convex in p for $p\in[0,\frac{1}{2}]$.

   First, notice that,

$$
\int_{\hat{p}}^{1} tg(t)dt=\int_{\hat{p}}^{1} tdG(t)=1-\int_{\hat{p}}^{1} G(t)dt-\hat{p}G(\hat{p})<1-\int_{\hat{p}}^{1} G(t)dt \text{ for } \hat{p}\in(0,\frac{1}{2}].
$$

   Second, when $g(\hat{p})$ is increasing in $\hat{p}$, we have,

$$
G(\hat{p})=\int_{0}^{\hat{p}} g(t)dt\leq\int_{0}^{\hat{p}} g(t)dt=\hat{p}g(\hat{p}).
$$

   Thus, $\tilde{\pi}'(p)<0$ for $p\in(0,\frac{1}{2}]$.

■

It directly follows that $\hat{p}^*=0$ for $\alpha\leq\alpha_1$ by Lemma A4. For $\alpha>\alpha_1$, it is useful to first prove the following two lemmas.

Lemma A5 *If $\rho(p;\alpha)$ is weakly decreasing in p, $\frac{G(p)}{pg(\hat{p})}$ is weakly decreasing in p for $p\in[0,\frac{1}{2}]$.*

Proof. By the definition of ρ-concavity,

$$
\rho(t)=1-\frac{G(t)g'(t)}{g^2(t)}.
$$

Integrating both sides from 0 to p yields,

$$
\int_{0}^{p}\rho(t)dt=\frac{G(p)}{g(p)}\Longrightarrow\frac{G(p)}{pg(p)}=\frac{\int_0^p\rho(t)dt}{p}.
$$

$$\Longrightarrow \Big(\frac{\int_0^p \rho(t)dt}{p}\Big)' = \frac{\rho(p)p - \int_0^p \rho(t)dt}{p^2} = \frac{\int_0^p \big[\rho(p) - \rho(t)\big]dt}{p^2} \leq 0.$$

■

Lemma A6 *For $\alpha_1 > \alpha_2$, $G(p;\alpha_1) > G(p;\alpha_2)$ for $p \in (0, \frac{1}{2})$.*

Proof. By Lemma A5,

$$\int_0^p \rho(t)dt = \frac{G(p)}{g(p)} \Longrightarrow \ln\big(\frac{1}{2}\big) - \ln G(p;\alpha) = \int_p^{\frac{1}{2}} \frac{1}{\int_0^\omega \rho(t;\alpha)dt} d\omega.$$

It can be verified that $\int_p^{\frac{1}{2}} \frac{1}{\int_0^\omega \rho(t;\alpha)dt} d\omega$ is decreasing in α by the definition of ρ-concave order. Thus, $G(p;\alpha)$ is increasing in α. ■

Rearranging the first order condition with respect to $\hat{p}$ yields,

$$\tilde{\pi}'(\hat{p}) \gtrless 0 \Longleftrightarrow \frac{G(\hat{p})}{\hat{p}g(\hat{p})} \gtrless \frac{1 - \int_{\hat{p}}^1 G(t)dt}{\int_{\hat{p}}^1 tg(t)dt} = \frac{\frac{1}{2} + \int_0^{\hat{p}} G(t)dt}{\frac{1}{2} + \int_0^{\hat{p}} G(t)dt - pG(p)}.$$

By Lemma A5, the LHS is decreasing in $\hat{p}$. It is can be verified that the RHS is increasing in $\hat{p}$. Thus, board's profit maximization problem for $\hat{p} \in [0, \frac{1}{2}]$ is well-behaved.

Notice that $\lim_{p\to 0} \frac{G(\hat{p})}{\hat{p}g(\hat{p})} = \rho(0) > 1$ for $\alpha > \alpha_1$, and $\lim_{p\to 0} \frac{1-\int_{\hat{p}}^1 G(t)dt}{\int_{\hat{p}}^1 tg(t)dt} = 1$. It suffices to compare the end points of the two curves.

If $2\int_0^{\frac{1}{2}} \rho(t;\alpha)dt > \frac{\frac{1}{2}+\int_0^{\frac{1}{2}} G(t)dt}{\frac{1}{4}+\int_0^{\frac{1}{2}} G(t)dt}$, $\tilde{\pi}(\hat{p})$ is increasing in $\hat{p} \in [0, \frac{1}{2}]$ and the optimal cutoff $\hat{p}^*$ lies between $\frac{1}{2}$ and 1.

If $2\int_0^{\frac{1}{2}} \rho(t;\alpha)dt < \frac{\frac{1}{2}+\int_0^{\frac{1}{2}} G(t)dt}{\frac{1}{4}+\int_0^{\frac{1}{2}} G(t)dt}$, $\tilde{\pi}(\hat{p})$ is first increasing and then decreasing in $\hat{p} \in [0, \frac{1}{2}]$. The maximal can be pinned down by the first order condition for $\hat{p} \in [0, \frac{1}{2}]$. We further argue that this local maximal is indeed the global maximal for $\hat{p} \in [0, 1]$. To see this, notice that second order derivative of the profit function with respect to $\hat{p}$ is,

$$\tilde{\pi}''(\hat{p}) = \frac{1}{4}\Big[-\hat{p}g'(\hat{p})\Big(1 - \int_{\hat{p}}^1 G(t)dt\Big) - 3\hat{p}g(\hat{p})G(\hat{p})\Big].$$

Since $G(p)$ is concave for $p \in [0, \frac{1}{2}]$ for $\alpha > \alpha_1$, $G(p)$ is convex for $p \in [\frac{1}{2}, 1]$. This directly implies $g'(p) > 0$ for $p \in [\frac{1}{2}, 1]$. Thus $\pi''(\hat{p}) < 0$ for $p \in [\frac{1}{2}, 1]$. Since $\tilde{\pi}'(\frac{1}{2}) < 0$, profit is decreasing in $\hat{p}$ for $\hat{p} \in [\frac{1}{2}, 1]$.

By definition of ρ-concavity, $2\int_0^{\frac{1}{2}} \rho(t;\alpha)dt$ is increasing in α. By Lemma A6, $\int_0^{\frac{1}{2}} G(t;\alpha)dt$ is increasing in $\alpha \implies \frac{\frac{1}{2}+\int_0^{\frac{1}{2}} G(t)dt}{\frac{1}{4}+\int_0^{\frac{1}{2}} G(t)dt}$ is decreasing in α. By Assumption 8 and 9,

$$\lim_{\alpha \to \alpha_1} 2\int_0^{\frac{1}{2}} \rho(t;\alpha)dt = 1 \text{ and } \lim_{\alpha \to \infty} 2\int_0^{\frac{1}{2}} \rho(t;\alpha)dt = \infty.$$

Thus, there exists $\alpha_2 > \alpha_1$ such that for $\alpha > \alpha_2$, anti-entrenchment is optimal; for $\alpha < \alpha_2$, entrenchment is optimal. ∎

Proof of Example 1. Given the functional form of $g(\cdot)$, it can be verified that board's profit function is,

$$\tilde{\pi}(\hat{p};\alpha) = \begin{cases} \frac{1}{4}\left[\frac{1}{2} + \frac{1}{4}\frac{\alpha}{\alpha+1}(2\hat{p})^{\frac{\alpha+1}{\alpha}}\right]\left[\frac{1}{2} - \frac{1}{4}\frac{1}{1+\alpha}(2\hat{p})^{\frac{\alpha+1}{\alpha}}\right] & \text{for} \quad \hat{p} \in [0, \frac{1}{2}] \\ \frac{1}{16}[2(1-\hat{p})]^{\frac{\alpha+1}{\alpha}}\left[\frac{\alpha}{\alpha+1} + \frac{\hat{p}}{1-\hat{p}}\right]\left[\hat{p} + \frac{1}{4}\frac{\alpha}{\alpha+1}[2(1-\hat{p})]^{\frac{\alpha+1}{\alpha}}\right] & \text{for} \quad \hat{p} \in (\frac{1}{2}, 1] \end{cases}.$$

It is obvious that $\alpha_1 = 1$. By the proof of Proposition 2, α_2 is the informativeness such that the first derivative of profit at $\hat{p} = \frac{1}{2}$ is equal to 0. $\implies \alpha_2 = \frac{\sqrt{5}+1}{2}$. ∎

Proof of Proposition 3. Designing a contact based on signal is equivalent to designing a contract based on posterior belief of incumbent's ability $p \in [0,1]$. By abuse of notation, denote $\{w(p), r(p), k(p)\}$ as the contract based on board's posterior belief. It suffices to prove that $r^*(p) = 1$ for $p \in [\mu, 1]$ and $r^*(p) = 0$ for $p \in [0, \mu)$ in optimal contract.

For a given contract $\{w(p), r(p), k(p)\}$, the incumbent manager chooses q to maximize:

$$\int_0^1 \{r(p)qpw(p) + [1-r(p)]k(p)\}g(p)dp - C(q).$$

First order condition yields,

$$C'(q) = \int_0^1 r(p)qw(p)g(p)dp.$$

Note that $k(p)$ can not provide incentive on effort level. Since the incumbent manager is protected by limited liability, $k^*(p) = 0$ in optimal contract.

The board chooses $\{w(p), r(p)\}$ to maximize:

$$\int_0^1 \Big\{\big[r(p)qp(1-w(p))\big] + \mu q(1-r(p))\Big\}\, g(p)dp$$
$$= q\bigg(\int_0^1 \big[r(p)p + \mu(1-r(p))\big]\, g(p)dp - \int_0^1 r(p)pw(p)g(p)dp\bigg).$$

Given effort level q, $\int_0^1 r(p)pw(p)g(p)dp = C'(q)$ is a constant by the incumbent manager's first order condition. It is equivalent to maximize:

$$\int_0^1 \big[r(p)p + \mu(1-r(p))\big]\, g(p)dp.$$

The integral is maximized by setting $r(p) = 1$ for $p \in [\mu, 1]$ and $r(p) = 0$ for $p \in [0, \mu)$. ■

Proof of Proposition 4. We first proved that $k^* = 0$ in optimal contract. Given (w_1, w_2, k) and belief of cutoff $\hat{s}$, the incumbent manager chooses q to maximize:

$$\mu[1 - F_1(\hat{s})]qw_1 + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\big(\mu q w_2 + k\big) - C(q).$$

$$\Longrightarrow q = \mu[1 - F_1(\hat{s})]w_1 + \mu\big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]w_2.$$

Given (w_1, w_2, k) and belief of project quality q, firm's indifference condition yields:

$$\mu q(1-w_2) - k = \frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu)f_0(\hat{s})} q(1-w_1).$$

The board chooses (w_1, w_2, k) to maximize expected profit,

$$\mu\big[1 - F_1(\hat{s})\big]q(1-w_1) + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\big[\mu q(1-w_2) - k\big]$$
$$= q(1-w_1)\left\{\mu\big[1 - F_1(\hat{s})\big] + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\frac{\mu f_1(\hat{s})}{\mu f_1(\hat{s}) + (1-\mu)f_0(\hat{s})}\right\}.$$

Given $(q, \hat{s})$ the board would like to induce, it is obvious that profit is decreasing in w_1. By the two equilibrium conditions, it is easy to verify that $w_1(k)$ is increasing in k and $w_2(k)$ is decreasing in k. Thus, $k^* = 0$.

The maximization problem of the board can be written as,

$$\max_{\{w_1,w_2,q,\hat{s}\}} \pi(w_1,w_2,q,\hat{s}) \equiv \mu\big[1-F_1(\hat{s})\big]q(1-w_1) + \big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]\mu q(1-w_2)$$

s.t.

$$q - \Big(\mu[1-F_1(\hat{s})]w_1 + \mu\big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big]w_2\Big) = 0$$

and

$$\mu(1-w_2) - \varphi(\hat{s})(1-w_1) = 0.$$

Let $\mathcal{L}$ be the Lagrangian and denote λ_1 and λ_2 as Lagrangian multipliers on the two constraints respectively.

$$\frac{\partial\mathcal{L}(w_1,w_2,q,\hat{s},\lambda_1,\lambda_2)}{\partial w_1} = 0 \Longrightarrow -\mu\big(q+\lambda_1\big)\big[1-F_1(\hat{s})\big] + \varphi(\hat{s})\lambda_2 = 0.$$

$$\frac{\partial\mathcal{L}(w_1,w_2,q,\hat{s},\lambda_1,\lambda_2)}{\partial w_2} = 0 \Longrightarrow -(q+\lambda_1)\big[\mu F_1(\hat{s}) + (1-\mu)F_0(\hat{s})\big] - \lambda_2 = 0.$$

It can be verified that $\hat{s}=0$ is never optimal. Thus, $\varphi(\hat{x})>0$. Then $\lambda_2=0$ and $\lambda_1=-q$ must be true. The first order condition of the Lagrangian with respect to $\hat{s}$ yields,

$$\frac{\partial\mathcal{L}(w_1,w_2,q,\hat{s},\lambda_1,\lambda_2)}{\partial\hat{s}} = 0.$$

$$\begin{aligned}
\Longrightarrow & -q(1-w_1)f_1(\hat{s}) + \big[\mu f_1(\hat{s}) + (1-\mu)f_0(\hat{s})\big]q(1-w_2) \\
& \qquad + \lambda_1\Big(f_1(\hat{s})w_1 - \big[\mu f_1(\hat{s}) + (1-\mu)f_0(\hat{s})\big]w_2\Big) = 0 \\
\Longrightarrow & \ \varphi(\hat{s}) = \mu \Longrightarrow \hat{s}^* = \frac{1}{2}.
\end{aligned}$$

Since $\mu(1-w_2)-\varphi(\hat{s})(1-w_1)=0$, it follows directly that $w_1^*=w_2^*$. ■

Proof of Proposition 5

Proof. Given contract (w,k) and $\hat{s}$, the incumbent manager's best response is,

$$q = \Big\{(1+\tau)\Big[\frac{1}{2}\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big] + \frac{1}{2}\big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big]w\Big\}^{\frac{1}{1-\tau}}.$$

Similarly, the board's indifference condition is,

$$\frac{1}{2}q^{1+\tau} - k = \left[\frac{1}{2} + \frac{f_1(\hat{s}) - f_0(\hat{s})}{f_1(\hat{s}) + f_0(\hat{s})}\delta\right] q^{1+\tau}(1-w).$$

Plugging the two equilibrium conditions into board's profit function yields,

$$\pi(\hat{s}) = M\Big[\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big] + \big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big]^{\frac{1+\tau}{1-\tau}}$$
$$\cdot\Big\{\Big[\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big] + \big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big] + \big[F_1(\hat{s})+F_0(\hat{s})\big]\Big[\frac{1}{2} + \frac{f_1(\hat{s}) - f_0(\hat{s})}{f_1(\hat{s}) + f_0(\hat{s})}\delta\Big]\Big\},$$

where $M = \frac{1}{4}^{\frac{2}{1-\tau}}(1-\tau)(1+\tau)^{\frac{2+2\tau}{1-\tau}}$.

Next, we calculate board's expected profit for a given cutoff $\hat{s}$ as $\alpha \to \infty$,

$$\lim_{\alpha\to\infty} \pi(\hat{s};\alpha) = \begin{cases} M\Big[1-(1-2\delta)\hat{s}\Big]^{\frac{1+\tau}{1-\tau}} & \text{for} \quad \hat{s} \in [0,\frac{1}{2}) \\ M\Big[\delta+\frac{1}{2}\Big]^{\frac{1+\tau}{1-\tau}}(1+\delta) & \text{for} \quad \hat{s} = \frac{1}{2} \\ M\Big[(1+2\delta)(1-\hat{s})\Big]^{\frac{1+\tau}{1-\tau}}(1+2\delta) & \text{for} \quad \hat{s} \in (\frac{1}{2},1] \end{cases}.$$

1. Entrenchment as $\alpha \to \infty$

   Notice that $\frac{f_1(\hat{s})-f_0(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})} \le 1$. Then $\pi(\hat{s};\alpha)$ can be bounded from above by,

$$\pi(\hat{s};\alpha) \le M\Big[\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big] + \big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big]^{\frac{1+\tau}{1-\tau}}$$
$$\cdot\Big\{\Big[\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big] + \big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big] + \big[F_1(\hat{s})+F_0(\hat{s})\big]\Big[\frac{1}{2}+\delta\Big]\Big\}.$$

   Denote the right hand side as $\pi_E(\hat{s};\alpha)$. By Lemma A1, $F_1(\hat{s};\alpha)$ converges uniformly to $\max\{0, 2\hat{s}-1\}$ as $\alpha \to \infty$. Thus, $\pi_E(\hat{s};\alpha)$ converges uniformly to $M\Big[(1+2\delta)(1-\hat{s})\Big]^{\frac{1+\tau}{1-\tau}}(1+2\delta)$ for $\hat{s} \in [\frac{1}{2},1]$ as $\alpha \to \infty$. Since $\pi(0;\alpha) = M$, entrenchment is optimal for sufficiently large α if,

$$M > \max_{\hat{s}\in[\frac{1}{2},1]} \Big\{M\big[(1+2\delta)(1-\hat{s})\big]^{\frac{1+\tau}{1-\tau}}\big(1+2\delta\big)\Big\} \Longrightarrow \delta < \frac{1}{2}^{\frac{1-\tau}{2}} - \frac{1}{2}.$$

2. Anti-entrenchment as $\alpha \to \infty$

   Notice that $\frac{f_1(\hat{s})-f_0(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})} \le 0$ for $\hat{s} \in [0,\frac{1}{2}]$. Then $\pi(\hat{s};\alpha)$ can be bounded from above by,

$$\pi(\hat{s};\alpha) \le M\Big[\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big]+\big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big]^{\frac{1+\tau}{1-\tau}} \cdot \Big\{\Big[\big(\frac{1}{2}+\delta\big)\big[1-F_1(\hat{s})\big]+\big(\frac{1}{2}-\delta\big)\big[1-F_0(\hat{s})\big]\Big]+\frac{1}{2}\big[F_1(\hat{s})+F_0(\hat{s})\big]\Big\}.$$

Denote the right hand side as $\pi_A(\hat{s};\alpha)$. By Lemma A1, $F_1(\hat{s};\alpha)$ converges uniformly to $\max\{0, 2\hat{s}-1\}$. Thus, $\pi_A(\hat{s};\alpha)$ converges uniformly to $\xi(\hat{s};\delta,\tau) = M\Big[1-(1-2\delta)\hat{s}\Big]^{\frac{1+\tau}{1-\tau}}\big(1+2\delta\hat{s}\big)$ for $\hat{s}\in[0,\frac{1}{2}]$ as $\alpha\to\infty$.

Given $(\delta,\tau)\in(0,\frac{1}{2})\times(-1,1)$, it can be verified that there exists $\nu(\delta,\tau)<\frac{1}{2}$ such that $\xi(\hat{s};\delta,\tau) < M\big[\delta+\frac{1}{2}\big]^{\frac{1+\tau}{1-\tau}}(1+2\delta)$ for $\hat{s}\in[\nu(\delta,\tau),\frac{1}{2}]$. Thus, $\hat{s}\in[\nu(\delta,\tau),\frac{1}{2}]$ can never be optimal for sufficiently large α.

Since $f_1(\hat{s})$ is strictly increasing in $\hat{s}$ and $\lim_{\alpha\to\infty} f_1(\hat{s};\alpha)=0$ for all $\hat{s}\in[0,\frac{1}{2})$, $f_1(\hat{s};\alpha)$ converges uniformly to 0 for $\hat{s}\in[0,\nu(\delta,\tau)]$ as $\alpha\to\infty$. Then $\pi(\hat{s};\alpha)$ converges uniformly to $M\Big[1-(1-2\delta)\hat{s}\Big]^{\frac{1+\tau}{1-\tau}}$ for $\hat{s}\in[0,\nu(\delta,\tau)]$ as $\alpha\to\infty$. Thus, entrenchment is optimal for sufficiently large α if,

$$\max_{\hat{s}\in[0,\nu(\delta,\tau)]} M\Big[1-(1-2\delta)\hat{s}\Big]^{\frac{1+\tau}{1-\tau}} < M\Big[\delta+\frac{1}{2}\Big]^{\frac{1+\tau}{1-\tau}}(1+2\delta) \Longrightarrow \delta > \frac{1}{2}^{\frac{1-\tau}{2}} - \frac{1}{2}.$$

■

Proof of Proposition 6. Given $\hat{s}$, a contract can always be constructed to induce $\hat{s}$. However, it is not necessarily $w=\frac{1}{2}$. $k\ge 0$ does not hold for all $\hat{s}\in[0,1]$ when $w=\frac{1}{2}$. To see this, notice that the severance pay k is,

$$k(\hat{s},w) = \underline{\pi}\big(q(\hat{s},w)\big) - \frac{f_1(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})}\big[(1-\lambda)q(\hat{s},w)+\lambda e(\hat{s},w)\big](1-w).$$

Given $\hat{s}$, letting w be sufficiently close to 1 generates a positive severance pay k.

Define $\hat{\mathcal{S}} = \Big\{\hat{s}\,\big|k(\hat{s},\frac{1}{2})\ge 0 \,\&\, \hat{s}\in[0,1]\Big\}$, which is the the set of cutoffs that can be induced by contracts that satisfy $w=\frac{1}{2}$ and $k\ge 0$.

If $\hat{s}\in\hat{\mathcal{S}}$, board's expected profit can be written as,

$$\overline{\pi}(\hat{s}) = \frac{1}{16}\Big[1-F_1(\hat{s})\Big]\Big[(1-\lambda)^2+\lambda^2\frac{1}{1-\hat{s}}\Big]\Big[[1-F_1(\hat{s})]+\hat{s}f_1(\hat{s})\Big].$$

If $\hat{s}\notin\hat{\mathcal{S}}$, $w=\frac{1}{2}$ cannot be sustained. Define $\mathcal{W}(\hat{s}) = \Big\{w\,\big|k(\hat{s},w)\ge 0 \,\&\, w\in[0,1]\Big\}$, which is the set of wages that can induce $\hat{s}$ without violating limited liability constraint of k.

1. Entrenchment

   It suffices to prove that $\pi(\hat{s};\alpha) < \pi(0;\alpha)$ for all $\hat{s} \in [\frac{1}{2}, 1]$ for sufficiently small α. $\pi(0;\alpha)$ is independent of α and can be calculated as,

$$\pi(0;\alpha) = \frac{1}{16}\big[(1-\lambda)^2 + \lambda^2\big].$$

   Lemma A7 *There exists $\Delta \in (0, \frac{1}{2})$ and N such that for $\alpha < N$, $\pi(\hat{s};\alpha) < \pi(0;\alpha)$ for all $\hat{s} \in [1-\Delta, 1]$.*

   Proof. By Lemma A2, for any $\epsilon > 0$ there exists N such that for $\alpha < N$, $1 - F_1(1-\Delta) < \Delta + \epsilon$ for all $\Delta \in [0,1]$. Note that $\frac{1}{4}\Big(\frac{1}{2}\lambda + \frac{1-\lambda}{\lambda}q\Big)^2 \geq \frac{1}{2}(1-\lambda)q$ for all q.

   The expected profit of replacement can be bounded from above by,

$$\begin{aligned}\underline{\pi}(q) - k &\leq \frac{1}{4}\Big(\frac{1}{2}\lambda + \frac{1-\lambda}{\lambda}q\Big)^2 = \frac{1}{16}\Big(\lambda + \frac{(1-\lambda)^2}{\lambda}\big[1-F_1(\hat{s})\big]w\Big)^2 \\ &\leq \frac{1}{16}\Big(\lambda + \frac{(1-\lambda)^2}{\lambda}\big[1-F_1(\hat{s})\big]\Big)^2.\end{aligned}$$

   Thus, board's expected profit can be bounded for $\hat{s} \in [1-\Delta, 1]$,

$$\begin{aligned}\pi(\hat{s}) &\leq \frac{1}{2}\big[1-F_1(\hat{s})\big]\left\{\frac{1-F_1(\hat{s})}{2}(1-\lambda)^2 + \frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}\lambda^2\right\}w(1-w) \\ &\qquad + \frac{1}{32}\big[F_1(\hat{s}) + F_0(\hat{s})\big]\Big(\lambda + \frac{(1-\lambda)^2}{\lambda}\big[1-F_1(\hat{s})\big]\Big)^2 \\ &\leq \frac{1}{8}\big[1-F_1(\hat{s})\big]\left\{\frac{1-F_1(\hat{s})}{2}(1-\lambda)^2 + \frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}\lambda^2\right\} \\ &\qquad + \frac{1}{16}\Big(\lambda + \frac{(1-\lambda)^2}{\lambda}\big[1-F_1(\hat{s})\big]\Big)^2 \\ &< \frac{1}{8}(\Delta+\epsilon)\Big[\frac{\Delta+\epsilon}{2}(1-\lambda)^2 + \lambda^2\Big] + \frac{1}{16}\Big[\lambda + \frac{(1-\lambda)^2}{\lambda}(\Delta+\epsilon)\Big]^2.\end{aligned}$$

   Note that the last expression is increasing in $\Delta + \epsilon$. It suffices to prove,

$$\frac{1}{16}\lambda^2 < \frac{1}{16}\big[(1-\lambda)^2 + \lambda^2\big].$$

Consequently, we can always find sufficiently small Δ and ϵ such that,

$$\frac{1}{8}(\Delta+\epsilon)\Big[\frac{\Delta+\epsilon}{2}(1-\lambda)^2+\lambda^2\Big]+\frac{1}{16}\Big[\lambda+\frac{(1-\lambda)^2}{\lambda}(\Delta+\epsilon)\Big]^2<\frac{1}{16}\big[(1-\lambda)^2+\lambda^2\big].$$

■

Next, notice that $\overline{\pi}(\hat{s})$ is the maximum expected profit without limited liability constraint of k. Thus $\pi(\hat{s})\leq\overline{\pi}(\hat{s})$ for all $\hat{s}\in[0,1]$.

Given ϵ, the expected profit for $\alpha<N$ for all $\hat{s}\in[\frac{1}{2},1-\Delta]$ is bounded from above by,

$$\begin{aligned}\pi(\hat{s})&\leq\frac{1}{16}\Big[1-F_1(\hat{s})\Big]\Big[(1-\lambda)^2+\lambda^2\frac{1}{1-\hat{s}}\Big]\Big[[1-F_1(\hat{s})]+\hat{x}f_1(\hat{s})\Big]\\&\leq\frac{1}{16}(1-\hat{s}+\epsilon)\Big[(1-\lambda)^2+\lambda^2\frac{1}{1-\hat{s}}\Big]\big[(1-\hat{s}+\epsilon)+\hat{s}(1+\epsilon)\big]\\&\leq\frac{1}{16}\Big(1+\frac{\epsilon}{\delta}\Big)\Big[\frac{1}{2}(1-\lambda)^2+\lambda^2\Big](1+2\epsilon).\end{aligned}$$

It remains to prove that $\frac{1}{16}\Big[\frac{1}{2}(1-\lambda)^2+\lambda^2\Big]<\frac{1}{16}\big[(1-\lambda)^2+\lambda^2\big]$, which is obvious.

2. Anti-entrenchment

   For $\hat{s}\in[0,\frac{1}{2}]$, the expected profit can be bounded from above by,

$$\begin{aligned}\pi(\hat{s};\alpha)&\leq\frac{1}{16}\Big[1-F_1(\hat{s};\alpha)\Big]\Big[(1-\lambda)^2+\lambda^2\frac{1}{1-\hat{s}}\Big]\Big[[1-F_1(\hat{s};\alpha)]+\hat{x}f_1(\hat{s};\alpha)\Big]\\&<\frac{3}{32}\Big[(1-\lambda)^2+2\lambda^2\Big].\end{aligned}$$

   It remains to find $\hat{s}>\frac{1}{2}$ that yields profit no less than $\frac{3}{32}[(1-\lambda)^2+2\lambda^2]$. For notational convenience, let $\psi=\big(\frac{\lambda}{1-\lambda}\big)^2$. $\lambda<\sqrt{2}-1$ directly implies that $\psi<\frac{1}{2}$. By limited liability constraint of k, we have,

$$\frac{f_1(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})}\big[(1-\lambda)q+\lambda e\big](1-w)\leq\underline{\pi}(q).$$

   Notice that $\frac{f_1(\hat{s})}{f_1(\hat{s})+f_0(\hat{s})}<1$, it suffices to satisfy,

$$\frac{1-F_1(\hat{s})}{4}(1-\lambda)^2w\geq\left\{\frac{1-F_1(\hat{s})}{2}(1-\lambda)^2+\frac{1-F_1(\hat{s})}{[1-F_1(\hat{s})]+[1-F_0(\hat{s})]}\lambda^2\right\}w(1-w),$$

and

$$\frac{1-F_1(\hat{s};\alpha)}{2}w \geq \frac{1}{2}\psi.$$

The second inequality comes from the construction that board will not induce effort to the new manager after replacement. Let $\hat{s} = \frac{1}{2} + \kappa(\lambda)$ and $w = \frac{1}{2} + \iota(\lambda)$, it suffices to find (κ, ι) that yields higher expected profit given λ. Noting that the first inequality is independent of α. By Lemma A1, $\frac{1-F_1(\hat{s};\alpha)}{2}$ can be arbitrarily close to $1 - \hat{s}$ when α is sufficiently large. Thus, these two conditions can be further simplified as,

$$\frac{1}{2}\Big(\frac{1}{2} - \kappa\Big) \geq \Big[\Big(\frac{1}{2} - \kappa\Big) + \psi\Big]\Big(\frac{1}{2} - \iota\Big),$$

and

$$2\Big(\frac{1}{2} + \iota\Big)\Big(\frac{1}{2} - \kappa\Big) \geq \psi.$$

$$\Longrightarrow \iota \geq \max\Big\{\frac{\frac{1}{2}\psi}{\frac{1}{2} - \kappa + \psi}, \frac{\psi}{1-2\kappa} - \frac{1}{2}\Big\}.$$

Let $\iota = \frac{\frac{1}{2}\psi}{\frac{1}{2}-\kappa+\psi}$. It can be verified that $\psi < \frac{1}{4}$ if $\kappa < \frac{1}{2} - \psi$. Board's expected profit of contract to the incumbent manager (w, k) that induces $\hat{s} = \frac{1}{2} + \kappa$ with wage $w = \frac{1}{2} + \iota$ as $\alpha \to \infty$ is,

$$\lim_{\alpha\to\infty} \pi(\hat{s};\alpha) = \Big[(1-\lambda)^2\big(\frac{1}{2} - \kappa\big) + \lambda^2\Big]\Big(\frac{1}{4} - \frac{\frac{1}{2}\psi}{\frac{1}{2} - \kappa + \psi}^2\Big).$$

Note that,

$$\begin{aligned}
\lim_{\kappa\to 0}\lim_{\alpha\to\infty} \pi(\hat{s};\alpha) &= \lim_{\kappa\to 0}\Big\{\Big[(1-\lambda)^2\big(\frac{1}{2} - \kappa\big) + \lambda^2\Big]\Big[\frac{1}{4} - \Big(\frac{\frac{1}{2}\psi}{\frac{1}{2} - \kappa + \psi}\Big)^2\Big]\Big\} \\
&= \frac{1}{2}\big[(1-\lambda)^2 + 2\lambda^2\big]\Big[\frac{1}{4} - \Big(\frac{\frac{1}{2}\psi}{\frac{1}{2} + \psi}\Big)^2\Big] \\
&> \frac{3}{32}\big[(1-\lambda)^2 + 2\lambda^2\big].
\end{aligned}$$

Thus, we can find sufficiently small κ such that $\lim_{\alpha\to\infty} \pi(\hat{s};\alpha) > \frac{3}{32}\big[(1-\lambda)^2 + 2\lambda^2\big]$. That is, anti-entrenchment is optimal to board when α is sufficiently large and $\lambda < \sqrt{2} - 1$.

■

Proof of Proposition 7.

1. Entrenchment

   It can be verified that $\pi(0;\alpha) = \frac{1}{16}$. Similarly, $\pi(1;\alpha) = 0$. Thus, $\hat{s} = 1$ is never optimal. It suffices to prove that there exists N such that for $\alpha < N$, $\pi(\hat{s}) < \pi(0)$ for all $\hat{s} \in [\frac{1}{2}, 1]$.

   Lemma A8 *There exists* $\Delta \in (0, \frac{1}{2})$ *and* N *such that for* $\alpha < N$, $\pi(\hat{s};\alpha) < \pi(0;\alpha)$ *for all* $\hat{s} \in [1-\Delta, 1]$.

   Proof. Since $q = \max\left\{\frac{1}{2}[1 - H_1(\hat{s})]w - \frac{1}{2}[H_0(\hat{s}) - H_1(\hat{s})]k, 0\right\}$, the effort level of the incumbent manager can be bounded from above by,

   $$q \leq \frac{1}{2}[1 - H_1(\hat{s})]w.$$

   Thus, the expected profit can be bounded from above by,

   $$\begin{aligned}\pi(\hat{s}, q) &\leq \frac{1}{2}q[1 - H_1(\hat{s})](1 - w) + \left\{\frac{1}{2}qH_1(\hat{s}) + \left(1 - \frac{1}{2}q\right)H_0(\hat{s})\right\}\left(\frac{1}{2}q - k\right) \\ &\leq \frac{1}{2}q[1 - H_1(\hat{s})](1 - w) + \frac{1}{2}q\left\{\frac{1}{2}qH_1(\hat{s}) + \left(1 - \frac{1}{2}q\right)H_0(\hat{s})\right\} \\ &\leq \frac{1}{2}q\big[[1 - H_1(\hat{s})] + 1\big] \leq q < 1 - H_1(\hat{s}).\end{aligned}$$

   Let $\Delta = \frac{1}{32}$. By lemma A2, for $\epsilon' = \frac{1}{32}$, there exists N such that for $\alpha < N$, $H_1(\hat{s}) \geq \hat{s} - \epsilon'$ for all $\hat{s} \in [0, 1]$. Since $\hat{s} \geq 1 - \Delta$, we have,

   $$\pi(\hat{s}, q) < 1 - H_1(\hat{s}) \leq 1 - \hat{s} + \epsilon' \leq \Delta + \epsilon' = \frac{1}{16} = \pi(0;\alpha).$$

   ■

   Lemma A9 *Given* $\Delta \in (0, \frac{1}{2})$ *and* $q \in [0, 1]$, *for any* $\epsilon > 0$, *there exists* N' *such that for* $\alpha < N'$, $\frac{\frac{1}{2}qh_1(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \left(1 - \frac{1}{2}q\right)h_0(\hat{s})} \leq \frac{1}{2}q + \epsilon$ *for* $\hat{s} \in [\frac{1}{2}, 1 - \Delta]$.

   Proof. For any ϵ, let $\epsilon' = \frac{\epsilon}{1+\epsilon}$. By definition of completely uninformative information structure, there exists N' such that for $\alpha < N'$, $h_1(1-\delta;\alpha) < 1+\epsilon'$.

$$\frac{\frac{1}{2}qh_1(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \left(1-\frac{1}{2}q\right)h_0(\hat{s})} - \frac{1}{2}q = \frac{1}{2}q\left(1-\frac{1}{2}q\right)\frac{h_1(\hat{s}) - h_0(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \left(1-\frac{1}{2}q\right)h_0(\hat{s})}$$
$$\leq \frac{1}{2}\frac{h_1(\hat{s}) - h_0(\hat{s})}{h_0(\hat{s})} = \frac{h_1(\hat{s}) - 1}{2 - h_1(\hat{s})}$$
$$\leq \frac{h_1(1-\delta;\alpha) - 1}{2 - h_1(1-\delta;\alpha)} \leq \frac{\epsilon'}{1-\epsilon'} = \epsilon.$$

■

By Lemma A9, for all $\hat{s} \in [\frac{1}{2}, 1-\Delta]$, $\pi(\hat{s}, q)$ can be bounded from above by,

$$\pi(\hat{s}, q) \leq \frac{1}{2}q[1 - H_1(\hat{s})](1-w) + \left\{\frac{1}{2}qH_1(\hat{s}) + \left(1-\frac{1}{2}q\right)H_0(\hat{s})\right\}\left(\frac{1}{2}q + \epsilon\right)(1-w)$$
$$\leq \frac{1}{2}q(1-w)\left[\left[1 - H_1(\hat{s})\right] + \frac{1}{2}qH_1(\hat{s}) + \left(1-\frac{1}{2}q\right)H_0(\hat{s})\right] + \epsilon$$
$$\leq \frac{1}{4}[1 - H_1(\hat{s})][2 - H_1(\hat{s})]w(1-w) + \epsilon$$
$$\leq \frac{1}{16}(1 - \hat{s} + \epsilon)(2 - \hat{s} + \epsilon) + \epsilon = \frac{1}{16}\left(\frac{1}{2} + \epsilon\right)\left(\frac{3}{2} + \epsilon\right) + \epsilon.$$

The last expression is strictly less than $\frac{1}{16}$ for sufficiently small ϵ.

2. Anti-entrenchment

   For $\hat{s} \in [0, \frac{1}{2}]$, $\zeta(\hat{s}, q) \leq \frac{1}{2}q$. Thus,

$$\pi(\hat{s}, q) = \frac{1}{2}q[1 - H_1(\hat{s})](1-w) + \left\{\frac{1}{2}qH_1(\hat{s}) + \left(1-\frac{1}{2}q\right)H_0(\hat{s})\right\}\left(\frac{1}{2}q - k\right)$$
$$\leq \frac{1}{2}q[1 - H_1(\hat{s})](1-w) + \zeta(\hat{s}, q)(1-w)H_0(\hat{s})$$
$$\leq \frac{1}{2}q(1-w)\left[1 + H_0(\hat{s}) - H_1(\hat{s})\right]$$
$$\leq \frac{1}{4}w(1-w)\left[1 - H_1(\hat{s})\right]\left[1 + H_0(\hat{s}) - H_1(\hat{s})\right] \leq \frac{1}{8}.$$

Next, we consider a fixed contract $(w', k') = (\frac{4}{5}, 0)$. It can be verified that this contract will not yield equilibrium with replacement cutoff $\hat{s}$ below $\frac{1}{2}$. To see this, notice that the effort level under this contract is,

$$q = \frac{2}{5}[1 - H_1(\hat{s})].$$

The expected profit of replacement is,

$$\frac{1}{2}q - k = \frac{1}{5}[1 - H_1(\hat{s})].$$

The expected profit created by the manager on the margin is,

$$\frac{\frac{1}{2}qh_1(\hat{s})}{\frac{1}{2}qh_1(\hat{s}) + \left(1 - \frac{1}{2}q\right)h_0(\hat{s})}(1-w) \leq \frac{1}{2}q(1-w) = \frac{1}{25}[1 - H_1(\hat{s})], \text{ for } \hat{s} \in [0, \frac{1}{2}].$$

The indifference condition of the board never holds for $\hat{s} \in [0, \frac{1}{2}]$. Thus, the only possible equilibrium replacement policy under this contract is $\hat{s} > \frac{1}{2}$. It remains to prove that the profit of the contract is above $\frac{1}{8}$ for sufficiently large α.

Lemma A10 *For $\Delta \in (0, \frac{1}{2})$, there exists N such that for $\alpha > N$, $\hat{s}(\alpha) < \frac{1}{2} + \Delta$ with contract $(w', k') = (\frac{4}{5}, 0)$.*

Proof. It suffices to prove that for $\Delta \in (0, \frac{1}{2})$, there exists N such that for $\alpha > N$, board's indifference condition never holds for all $\hat{s} \in [\frac{1}{2} + \Delta, 1]$ with contract $(w', k') = (\frac{4}{5}, 0)$.

Board's indifference condition can be simplified as,

$$\frac{h_1(\hat{s};\alpha)}{h_0(\hat{s};\alpha)} = 1 + \frac{4}{H_1(\hat{s};\alpha)}.$$

Since $H_1(\hat{s}) \geq 2\hat{s} - 1$,

$$1 + \frac{4}{H_1(\hat{s};\alpha)} \leq 1 + \frac{4}{H_1(\frac{1}{2} + \Delta;\alpha)} \leq 1 + \frac{2}{\Delta}.$$

$\frac{h_1(\hat{s};\alpha)}{h_0(\hat{s};\alpha)}$ is approaching infinity as $\alpha \to \infty$ while $1 + \frac{4}{H_1(\hat{s};\alpha)}$ is bounded, which is a contradiction. ■

For notational convenience, define $\Lambda(\hat{s};\alpha) = 1 - H_1(\hat{s})$. Board's expected profit can be written as,

$$
\begin{aligned}
&\pi\big(\hat{s}(w',k';\alpha),q(w',k';\alpha)\big)\\
=&\frac{1}{5}\Lambda^2(\hat{s};\alpha)\Big(\frac{7}{5}-\frac{2}{5}\Lambda(\hat{s};\alpha)\Big)+\frac{1}{5}\Lambda(\hat{s};\alpha)\Big(1-\frac{1}{5}\Lambda(\hat{s};\alpha)\Big)(2\hat{s}-1)\\
\geq&\frac{1}{5}\Lambda^2(\hat{s};\alpha)\Big(\frac{7}{5}-\frac{2}{5}\Lambda(\hat{s};\alpha)\Big)\geq\frac{1}{5}\Lambda^2(\hat{s};\alpha).
\end{aligned}
$$

By Lemma A1, given $\epsilon > 0$, there exists N such that for $\alpha > N$, $\Lambda(\hat{s};\alpha) > 2(1-\hat{s}) - \epsilon$ for all $\hat{s} \in [\frac{1}{2}, 1]$. Thus,

$$
\pi\big(\hat{s}(w',k';\alpha),q(w',k';\alpha)\big)\geq\frac{1}{5}\big[2(1-\hat{s})-\epsilon\big]^2\geq\frac{1}{5}(1-2\Delta-\epsilon)^2.
$$

Let $\Delta = \epsilon = \frac{1}{24}$. Then,

$$
\pi\big(\hat{s}(w',k';\alpha),q(w',k';\alpha)\big)\geq\frac{1}{5}(1-2\Delta-\epsilon)^2=\frac{49}{320}\geq\frac{1}{8}.
$$

■

Appendix B: Normalization of information structure

In this section we first show that normalizing the signal space $\mathcal{S}$ to $[0,1]$ and assuming $\frac{1}{2}F_1(s)+\frac{1}{2}F_0(s) = s$ are without loss of generality. Next we show that the three assumptions imposed on $\{f_1(\cdot), f_0(\cdot)\}$ can be derived from similar assumptions on information structures without such normalization.

Suppose instead the board receives a noisy signal $x \in \mathcal{X}$ about incumbent manager's ability θ_i. x is drawn from distribution with cdf $\tilde{F}_{\theta_i}(\cdot)$ and pdf $\tilde{f}_{\theta_i}(\cdot)$ for $\theta_i \in \{0,1\}$ with support $\mathcal{X} = [\underline{x}, \overline{x}]$, where $-\infty \leq \underline{x} < \overline{x} \leq \infty$. The two conditional distributions together with the signal space $\{\tilde{f}_1(\cdot), \tilde{f}_0(\cdot), \mathcal{X}\}$ define an information structure.

Given an information structure $\{\tilde{f}_1(\cdot), \tilde{f}_0(\cdot), \mathcal{X}\}$, define a new signal x by applying the probability integral transformation to $x = \frac{1}{2}\tilde{F}_1(x) + \frac{1}{2}\tilde{F}_0(x)$. Then the unconditional distribution of s is uniform on $[0,1]$ with $\mu = \frac{1}{2}$. Let $F_\theta(s)$ and $f_\theta(s)$ be the corresponding conditional cdf and pdf for $\theta_i \in \{0,1\}$ respectively. It can be verified that $\frac{1}{2}F_1(s) + \frac{1}{2}F_0(s) = s$ for all $s \in [0,1]$.

Assumption 10 *Monotone likelihood ratio property (MLRP):* $\frac{\tilde{f}_1(x)}{\tilde{f}_0(x)}$ *is strictly increasing in* $x \in [\underline{x}, \overline{x}]$.

Assumption 10 directly implies Assumption 1. For binary states, MLRP assumption is without loss of generality since it can always be satisfied by relabeling signals according to the likelihood ratio.

Lemma B1 *Suppose two information structures* $\{\tilde{f}_1(\cdot), \tilde{f}_0(\cdot), \mathcal{X}\}$ *and* $\{\tilde{f}_1^\dagger(\cdot), \tilde{f}_0^\dagger(\cdot), \mathcal{X}^\dagger\}$ *generate the same distributions of posterior beliefs for* μ. *Then they yield the same distributions of posterior beliefs for* $\mu' \neq \mu$.

The proof of Lemma B1 is similar to Lemma 2 and thus omitted. Since entrenchment (anti-entrenchment) is defined by comparing the expected ability of the incumbent manager with the replacement manager, only the posterior belief of the incumbent manager matters. By Lemma B1, we can restrict attention on information structures that satisfy $\frac{1}{2}F_1(s) + \frac{1}{2}F_0(s) = s$ for $s \in [0,1]$ without loss of generality.

Assumption 11 *Perfectly informative at extreme signals:* $\lim_{x \to \underline{x}} \frac{\tilde{f}_1(x)}{\tilde{f}_0(x)} = 0$ *and* $\lim_{x \to \overline{x}} \frac{\tilde{f}_1(x)}{\tilde{f}_0(x)} = +\infty$.

Assumption 12 *Mirror symmetry: There exists* $\check{x} \in (\underline{x}, \overline{x})$ *such that* $\tilde{f}_0(x) = \tilde{f}_1(2\check{x} - x)$.

Similarly, Assumption 11 and 12 directly imply Assumption 2 and 3 respectively. We close this section by introducing two indexed families of information structures that satisfy Assumption 10 - 12. Moreover, the corresponding normalized signals after probability integral transformation satisfy Definition 4.

Example 2 (Normal Distribution) *Suppose* $x = \theta_i + \epsilon$ *for* $\theta_i \in \{0, 1\}$, *where* $\epsilon \sim \mathcal{N}(0, \alpha^{-1})$. *Then* $x|\theta \sim \mathcal{N}(\theta, \alpha^{-1})$.

Example 3 (Beta Distribution) *Suppose* $\tilde{f}_1(x;\alpha) = (1+\alpha)x^\alpha$ *and* $\tilde{f}_0(x;\alpha) = (1+\alpha)(1-x)^\alpha$ *for* $x \in [0,1]$. *Then* $\tilde{F}_1(x;\alpha) = x^{1+\alpha}$ *and* $\tilde{F}_0(s;\alpha) = 1 - (1-x)^{1+\alpha}$. *This example is borrowed from Taylor and Yildirim (2011).*

For both examples, $\alpha \in (0, \infty)$ is interpreted as the informativeness of the information structure.

Appendix C: Properties of the ρ-concave order

By Lemma A6, concave order implies rotation order first introduced by Johnson and Myatt (2006) when $\mu = \frac{1}{2}$. It can be verified that for $\mu \neq \frac{1}{2}$, rotation order does not remain. Intuitively, if information structure becomes more informative, more densities concentrate on $p = 0$ and $p = 1$, and the distribution becomes more disperse.

Lemma C1 (Bayesian update) *Suppose $G_1(\cdot)$ is more informative than $G_2(\cdot)$ in the ρ-concave order. Then $\varphi(s, \mu|G_1) \geq \varphi(s, \mu|G_2)$ for $s \in (\frac{1}{2}, 1]$ and $\varphi(s, \mu|G_1) \leq \varphi(s, \mu|G_2)$ for $s \in (0, \frac{1}{2}]$.*

Proof. Since $G_1(0) = G_1(0) = 0$ and $G_1(\frac{1}{2}) = G_2(\frac{1}{2}) = \frac{1}{2}$ and $G_1(p) \geq G_2(p)$ for $p \in [0, \frac{1}{2}]$ by Lemma A6, $G_1^{-1}(s) \leq G_2^{-1}(s)$ for $s \in [0, \frac{1}{2}]$. Thus $\varphi(s, \mu|G_1) = \frac{\mu G_1^{-1}(s)}{\mu G_1^{-1}(s)+(1-\mu)[1-G_1^{-1}(s)]} \leq \frac{\mu G_2^{-1}(s)}{\mu G_2^{-1}(s)+(1-\mu)[1-G_2^{-1}(s)]} = \varphi(s, \mu|G_2)$. The proof for $s \in (\frac{1}{2}, 1]$ is similar. ■

Lemma C1 shows the implication of the ρ-concave order on the Bayesian update of the incumbent manager's ability. The posterior belief $\varphi(x, \mu; \alpha)$ rotates counter-clockwise via $(\frac{1}{2}, \mu)$ the as information structure becomes more informative. In other words, a fixed signal x has more information value to the board as information structure becomes more informative.

Lemma C2 (Comparison with Blackwell's sufficiency) *If $G_1(\cdot)$ is more informative than $G_2(\cdot)$ in the ρ-concave order, $G_1(\cdot)$ is more informative than $G_2(\cdot)$ in the sense of Blackwell.*

Proof.

Lemma C3 *$F_1(s|G_1) \leq F_1(s|G_2)$ and $F_0(s|G_1) \geq F_0(s|G_2)$ for $s \in [0, 1]$.*

Proof. From the proof of Lemma C1, $G_1^{-1}(s) \leq G_2^{-1}(s)$ for $s \in [0, \frac{1}{2}]$.

1. For $s \in [0, \frac{1}{2}]$,

$$F_1(s|G_1) = \int_0^s f_1(t|G_1)dt = \int_0^s 2G_1^{-1}(t)dt \leq \int_0^s 2G_2^{-1}(t)dt = F_1(s|G_2).$$

2. For $s \in (\frac{1}{2}, 1]$,

$$
\begin{aligned}
F_1(s|G_1) = \int_0^s f_1(t|G_1)dt &= \int_0^{1-s} f_1(t|G_1)dt + \int_{1-s}^s f_1(t|G_1)dt \\
&= \int_0^{1-s} f_1(t|G_1)dt + \frac{1}{2}(2s-1) \\
&\leq \int_0^{1-s} f_1(t|G_2)dt + \int_{1-s}^s f_1(t|G_2)dt = F_1(s|G_2).
\end{aligned}
$$

Thus, $F_1(s|G_1) \leq F_1(s|G_2)$ for $s \in [0,1]$. Similarly, $F_0(s|G_1) \geq F_0(s|G_2)$. ■

Note that for binary states, Blackwell's order is equivalent to Lehmann's order. Thus, it suffices to prove that for $\omega \in (0,1)$,

$$F_1(F_0^{-1}(\omega|G_1)|G_1) \leq F_1(F_0^{-1}(\omega|G_2)|G_2).$$

Suppose we have the contrary, then there exist ω' such that,

$$F_1(F_0^{-1}(\omega'|G_1)|G_1) > F_1(F_0^{-1}(\omega'|G_2)|G_2).$$

By Lemma C3, it follows directly that $F_0^{-1}(\omega'|G_1) > F_0^{-1}(\omega'|G_2)$. However, $F_0^{-1}(\omega'|G_1) > F_0^{-1}(\omega'|G_2)$ can not be true. To see this, let $s_1 = F_0^{-1}(\omega'|G_1)$ and $s_2 = F_0^{-1}(\omega'|G_2)$. Then $s_1 > s_2$ and $F_0(s_1|G_1) = F_0(s_2|G_1) = \omega'$. Again by Lemma C3, we have $F_0(s_1|G_1) > F_0(s_2|G_1) \geq F_0(s_2|G_2)$, which is a contradiction. ■