

Optimal Long-term Executive Contracts

Robert A. Miller* Anh Nguyen†

PRELIMINARY AND INCOMPLETE

Abstract

This paper studies the optimal long-term contract when firms and executives learn about their match qualities through realized performances, which are affected by executives' hidden effort choices. Our model is motivated by the empirical observation that both executive retention rates and compensation in SP1500 companies are influenced not only by current performance but also by past performance. Due to the interaction between learning and moral hazard, the optimal contract involves involuntary executive termination by the firm as a disciplinary device against moral hazard. We then identify and estimate the learning process of match quality and the cost of executive moral hazard for SP1500 companies. We also quantify the firm's profit gain from its ability to commit to long-term contracts.

*Tepper School of Business, Carnegie Mellon University. Email: ramiller@andrew.cmu.edu

†Tepper School of Business, Carnegie Mellon University. Email: anhnguyen@cmu.edu

1 Introduction

In executive compensation, short-termism is one of the key incentive misalignments between the executive and the firm. To resolve this conflict, firms frequently structure executive compensation packages with diverse vesting periods, essentially forming a long-term contract. In this paper, we explore the use of executive dismissals as another disciplinary tool besides compensation for the firm.

2 Literature Review

Our paper contributes to the literature on job search ([Jovanovic, 1979](#); [Miller, 1984](#); [Antonovics and Golan, 2012](#); [Pastorino, Forthcoming](#)). Similar to the setting in [Jovanovic \(1979\)](#), our model features ex-ante homogeneous jobs, and executives and firms learn about match-specific executive quality as executive performances are realized. Our model differs from the existing models in this literature in two aspects. First, in our model, an executive's private effort choice affects the realized performance. This implies that the learning process is endogenous to the hidden action of the executive. Second, whereas in most models, the separation between a firm and a worker is always a voluntary action made by the worker, in our model, separation can be involuntary.

More broadly, involuntary termination in the optimal contract between an agent and a principal can arise for several reasons: the agent has limited liability ([Clementi and Hopenhayn \(2006\)](#); [DeMarzo and Sannikov \(2006\)](#)), the agent is risk-averse and providing insurance through the contract is too costly ([Spear and Wang \(2005\)](#)), or the executive's realized quality is lower than that of a potential replacement (([Hermalin and Weisbach, 1998](#))). We consider a setting in which agents are risk averse and have access to a sufficiently complete market to smooth their outside wealth. The incorporation of risk aversion is empirically motivated by extensive literature that studies the role of risk in compensation contracts, such as [Demsetz and Lehn \(1985\)](#) and [Peters and Wagner \(2014\)](#). In our model, a firm also uses termination to deter an executive from shirking. This is facilitated by the firm's ability to commit to a long-term contract and the interaction between match quality uncertainty and moral hazard. This interaction is a central theme in a large body of literature on principal-agent models featuring symmetric uncertainty and imperfect monitoring (([Holmström, 1999](#); [Prat and Jovanovic, 2014](#); [DeMarzo and Sannikov, 2016](#); [Cisternas, 2018](#); [Bhaskar and Mailath, 2019](#); [Lyman, 2023](#))). Our paper differs from the existing literature in two key ways: First, we do not assume a parametric learning process. Second, whereas existing models often require

either risk-neutral executives or firms that can directly control consumption decisions, our model does not allow firms to control the executive’s consumption, a choice motivated by both institutional settings and the fact that executives’ consumption decisions are typically unobservable in data.

In addition, this paper complements the literature on quantifying executive turnover (Parrino (1997); Taylor (2010); Peters and Wagner (2014); Jenter and Lewellen (2021)). A central issue in this field is that it is often not observable whether an executive is fired. The most common approach in the literature, following Parrino (1997), relies on press reports along with age and other observed characteristics as criteria. In our model, the firing decision does not need to be deterministic. Our approach allows us to quantify the probability of an executive being fired as part of the optimal contract.

Our paper also contributes to the empirical literature on dynamic moral hazard models (Margiotta and Miller, 2000; Gayle and Miller, 2009, 2015). The main difference between our paper and the existing papers is that we consider long-term contracts instead of short-term contracts. Our reasons are two-fold. First, in the aforementioned papers, because the current effort choice is always a sufficient statistic for the distribution of output produced by the agent, the optimal contract can always be constructed as a sequence of short-term contracts (Fudenberg et al. (1990); Malcomson and Spinnewyn (1988); Rey and Salanie (1990)). In contrast, in our model, the distribution of an executive’s performance depends not only on her contemporaneous effort choice but also on her past effort choices. Therefore, a sequence of short-term contracts is no longer optimal. In addition, we also add to the existing literature on long-term contracting, which often does not consider moral hazard and instead focuses on persistent private information. For example, Hendel and Lizzeri (2003); Hofmann and Browne (2013); Atal et al. (2020) and Ghili et al. (2021) study the welfare effects of long-term contracts in the insurance market in which the benefits from using a long-term contract is to insure consumers against reclassification risks—i.e., the risk of receiving adverse information in the future.

3 Setting and Data

We construct our data from the Compustat Execucomp database in the period between 1992 and 2022. The Compustat database includes details on executive compensations at SP500 companies in 1992 and 1993 and at SP1500 companies from 1994 to 2022. The dataset also includes the stock performance of the firms in the SP1500 index. For our analysis, we

restricted attention to chief executive officers only. The two key variables of interest are (1) each executive’s annual compensation, and (2) firm’s performance.

To calculate an executive’s compensation, we follow the approach outlined by [Margiotta and Miller \(2000\)](#) and [Gayle et al. \(2015\)](#) which is to include the change in wealth from previously held options and stocks in the compensation measure. Therefore, an executive’s total compensation is comprised of two components. The first component, direct compensation, is observed in the Compustat database and includes salary, bonuses, other annual payments, restricted stock and option grants, and long-term incentive plan payouts. The second component is equal to the value of options and stocks at the beginning of the period, multiplied by the firm’s abnormal return. In our benchmark analysis, we compute the abnormal return as the difference between the firm’s stock return and the market portfolio’s return, the latter measured by the S&P 500 return.

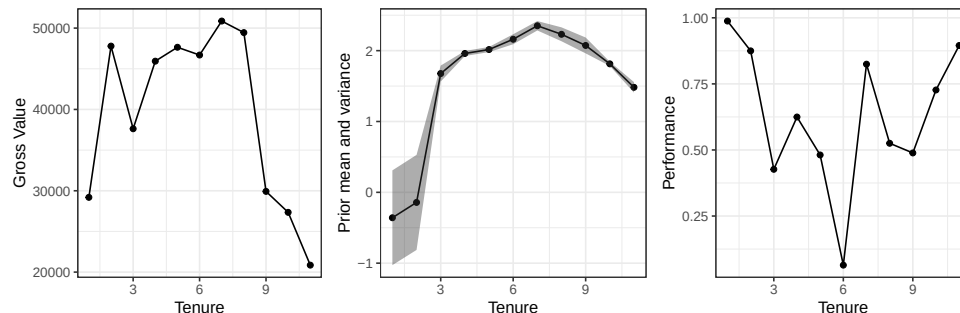
To compute the firm’s performance, we use its gross market value, which is equal to the sum of the market value of the firm and dividends paid prior to executive compensation. We denote the gross market value at the beginning of period t as $\tilde{V}_{t,j}$. Because this measure does not reflect the potential differences across industries, we also construct an additional measure that reflects the public belief about a firm’s performance. Conditional on the realized gross market value in $t + 1$ being v , this measure is constructed as follows:

$$y := \Pr \left(\tilde{V}_{t+1,j} < v | \{\tilde{V}_{\tau,j}\}_{\tau \leq t}, X_{t,j} \right)$$

where $X_{t,j}$ denote the set of observed characteristics of a firm, which include industry-fixed effects, firm size, and the calendar year. Therefore, conditional on the history of the realized firm performance up to the tenure year t , the higher y , the better the firm performs relative to the shareholders’ expectations based on past performances. Because the dimension of the conditioning variables increases as t increases, when constructing y , we assume that the conditional distribution of $\tilde{V}_{t+1,j} | \tilde{V}^{t,j}, X_{t,j}$ can be characterized by its first two central moments. We denote these central moments as $(M_{1,t,j}, M_{2,t,j})$ and henceforth refer to them as the prior mean and prior variance of the shareholders’ belief at t . [Figure 3.1](#) illustrates the construction of this performance index.

[Table 3.2](#) presents the summary statistics of key variables in the data. In this table, an external executive is defined as one who was not employed by the firm one year prior to becoming an executive, and a large firm as one with a market value greater than the median value in the data. Most executives in our dataset are hired internally, and approximately 40% of executives separate within the first three years. The value of stocks and options

Figure 3.1: Example of the constructed prior mean, prior standard deviation, and performance



Note: Constructed data for John F. Smith, Jr, CEO at GENERAL MOTORS CO from 1992 to 2002.

represents a significant component of executives' total compensation; for example, the average salary and bonus only account for approximately 17% of the total compensation for an external executive. Firms with short-tenure executives tend to perform better on average than those with long-tenure executives; however, the former group experiences significantly higher variations in abnormal returns compared to the latter. Additionally, the compensation variance for executives with short tenures is smaller than that for those with longer tenures, suggesting that pay for longer-tenure executives is more heavily linked to stocks and options. Finally, the data show significant variations in compensation across executives.

Table 3.2: Summary statistics

	All						Across-CEOs
	External	Internal	Short tenure (≤ 3)	Long tenure (> 3)	Large firms	Small firms	
Salary and bonus	1.02 (1.19)	0.92 (1.2)	0.71 (0.93)	0.97 (1.24)	1.09 (1.38)	0.53 (0.37)	(0.74)
Total compensation	4.81 (24.14)	4.35 (21.51)	3.9 (14.81)	4.48 (22.77)	5.55 (25.82)	1.64 (3.36)	(20.29)
Abnormal return	0.26 (7.49)	0.1 (8.17)	0.32 (12.99)	0.09 (6.98)	0.15 (9.5)	0.04 (2.51)	(6.67)
Performance index	0.5 (0.25)	0.49 (0.23)	0.45 (0.25)	0.5 (0.23)	0.57 (0.21)	0.32 (0.19)	(0.15)
Number of obs.	4272	32643	5185	31730	26030	10885	
Number of firm- executive pairs	735	5905	2592	4048	4211	2429	

Note: Compensation figures are expressed in millions of 2000 dollars. An external executive is defined as one who has not worked for the firm in the year prior to becoming an executive. We classify large firms as those whose market value exceeds the median firm value in the data.

Table 3.3: Coefficients of the regressions of total compensation and retention on performance index and prior distribution statistics

	Total Compensation	Retention
	(1)	(2)
Performance index	4.69 (0.06)	1.13 (0.09)
External executives	-0.43 (0.09)	-0.16 (0.13)
Prior mean	1.58 (0.02)	0.02 (0.03)
Prior variance	-0.63 (0.03)	0.27 (0.05)
Performance index x External	0.3 (0.17)	-0.29 (0.24)
Prior mean x External	-0.37 (0.07)	0.1 (0.09)
Prior variance x External	0.94 (0.13)	0.55 (0.17)

Note: Total compensation is measured in millions of 2000 USD. Column (1) reports the coefficients of a quantile regression that estimates the median value of total compensation, and column (2) reports the coefficients of the logistic regression that predicts the probability of retention. The variables included but not reported are executive gender, age, industry-fixed effects, and year-fixed effects. The standard errors are in parentheses.

In Table 3.3, we explore how past and current performance affect executive pay and retention. As previously explained, we capture the effect of past performances using the prior mean and variance of the gross market value, conditional on past market values and observed characteristics. Because the distribution of total compensation has large tails, we employ quantile regression on total compensation to minimize the influence of outliers on our regression estimates. The coefficient estimates from this regression are reported in Column (1) of Table 3.3. Higher performance and more optimistic beliefs about future performance are associated with higher compensation. Furthermore, executives are paid more when the belief about the firm's performance is less uncertain, as indicated by the coefficient on the prior variance. However, compared to internal executives, external executives are paid more amid greater uncertainty about the firm's performance. Additionally, the pay of external executives is less sensitive to the prior mean than that of internal executives.

In Column (2) of Table 3.3, we examine the correlation between retention (both voluntary

and involuntary) and past and current performance. Executives are more likely to stay with the firm when its performance is higher and when the uncertainty in prior beliefs has not yet been resolved. The retention likelihood of external executives is less affected by performance; however, this relationship is not statistically significant. Conversely, the influence of prior beliefs on retention is more pronounced for external executives compared to internal executives. For example, external executives are more likely to remain with the firm when the belief about the firm's performance is more favorable or when there is greater uncertainty in these prior beliefs.

4 Model

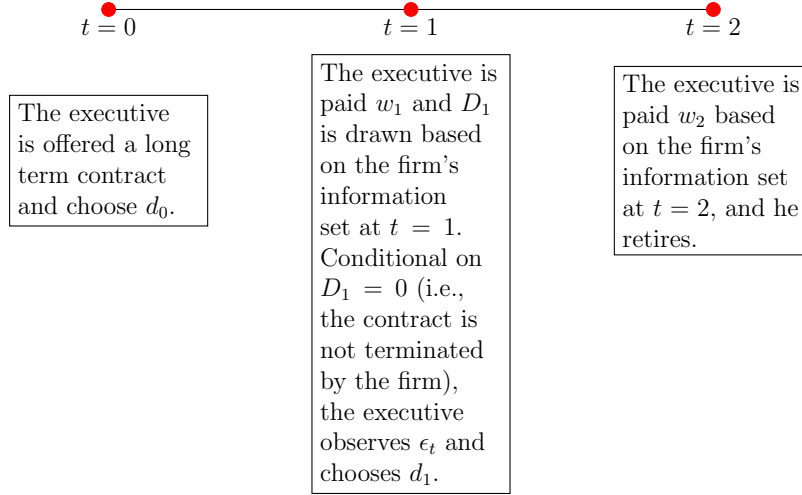
A firm is looking to hire an executive from a pool of potential executives. The executives are ex-ante identical to the firm; however, executives have heterogeneous match qualities, which are firm-specific. The match quality between the firm and an executive is unknown to both parties. Importantly, the match quality positively affects the firm's gross market value. We assume that the firm and all executives are infinitely lived. Once hired, let T denote the duration of the employment, and $t \in \{0, 1, 2, \dots, T\}$ denote a tenure year. The firm can employ an executive for at most $\bar{T}$ years, that is, $T < \bar{T}$, after which the executive retires.

At the beginning of period $t = 0$, the executive receives a contract offer from the firm specifying wages $\{w_1, w_2, \dots, w_T\}$, where w_t is the wage received at the beginning of period t for work performed in period $t - 1$. We assume that there is a lower bound on expected wages, denoted by $\underline{w}$. The contract also specifies a termination probability $\{q_1, q_2, \dots, q_T\}$. The contract $\{w_t, q_t\}_{t=1, \dots, \bar{T}}$ can depend on the firm's information set at the beginning of each period, which we will specify in more detail later.

4.1 Executive preferences and consumption choices

The executive derives utility from consumption c_t . In any period t , conditional on the firm not terminating the contract ($D_t = 0$), the executive can make one of three choices: (1) stay employed but not exert effort ($d_{0,t} = 1$), (2) stay employed and exert effort ($d_{1,t} = 1$), and (3) leave the firm ($d_{2,t} = 1$). Note that $d_t = (d_{0,t}, d_{1,t}, d_{2,t})$ are mutually exclusive. The firm does not observe the executive's past and current choices. If the executive is not employed by the firm, his per-period utility is normalized to 1. Figure 4.1 shows a summary of the timeline and executive decisions when $\bar{T} = 2$.

Figure 4.1: Summary of the timeline for $\bar{T} = 2$



The executive's lifetime utility is given by

$$\sum_{t=0}^{T-1} -\delta^t \exp(-\rho c_t) [(d_{0,t}\alpha_0 + d_{1,t}\alpha_1) \exp(-\epsilon_t) + d_{2,t}] + \sum_T^{\infty} -\delta^t \exp(-\rho c_t) \quad (1)$$

where δ is the discount factor, ρ is the absolute risk aversion coefficient, and T is the realized tenure duration. ϵ_t represents an idiosyncratic preference to work for the firm. We assume that $\epsilon_0 = 0$, and $\{\epsilon_t\}_{t \geq 1}$ follow a logistic distribution and are independent over time. α_0 represents the utility cost of choosing $d_{0,t} = 1$, and α_1 represents the utility cost of choosing $d_{1,t} = 1$. We assume that exerting effort is more costly, i.e., $\alpha_1 > \alpha_0$; however, executive effort has a positive impact on the firm's payoff.

The executive has access to a financial market and can choose to smooth consumption; however, he cannot hedge against the volatility of w_{t+1} by purchasing a contingent plan in t . Here, we follow the framework of [citehere], which shows that it suffices to consider a finite set of financial securities to characterize the optimal financial portfolio of the executive. In particular, let λ_t denote the price at time 0 of consumption at time t . Let b_t denote the price of a bond that pays a unit of consumption from period t onward, and a_t denote the price in period t of a security that pays the random quantity $(\ln \lambda_s - s \ln \delta)$ of consumption in each period s from period t onward.¹ We further assume that the executive and the firm have perfect foresight about $\{a_t, b_t\}_{t \leq \bar{T}}$.

Let ξ_t denote the executive's wealth at t . Let $v_t = \exp(-\rho w_t/b_t)$, which is a measure of

¹i.e., $a_t = \mathbb{E}_t \left[\sum_{s=t}^{\infty} \frac{\lambda_s}{\lambda_t} (\ln \lambda_s - s \ln \delta) \right]$

compensation from working annuitized for consumption smoothing purposes. Let $\mathcal{S}_t^*$ denote the executive's information set in t . Because past effort decisions and past output realizations affect the executive's belief about θ , $\mathcal{S}_t^*$ includes the information derived from his privately observed decisions $\{d_\tau\}_{\tau < t}$. When the relationship is not terminated by the firm — that is, conditional on $D_t = 0$ — let p_t denote the probability that the executive chooses to stay with the firm before realizing ϵ_t . Under optimal consumption decisions, the executive value function prior to ϵ_t being realized is given by

$$- \lambda_t b_t \exp \left(-(a_t + \rho \xi_t) b_t^{-1} \right) (A_t | \mathcal{S}_t^*) \quad (2)$$

where

$$A_t = (1 - p_t) + p_t \mathbb{E}_t \left[\exp(-\epsilon_t b_t^{-1} | d_{2,t} = 0) \right] \min \left\{ \alpha_0^{b_t^{-1}} \mathbb{E}_t \left(v_{t+1} [q_{t+1} + (1 - q_{t+1}) A_{t+1} | d_{0,t} = 1, \mathcal{S}_t^*] \right)^{1-b_t^{-1}}, \right. \\ \left. \alpha_1^{b_t^{-1}} \mathbb{E}_t \left(v_{t+1} [q_{t+1} + (1 - q_{t+1}) A_{t+1} | d_{1,t} = 1, \mathcal{S}_t^*] \right)^{1-b_t^{-1}} \right\}$$

Intuitively, the first component of A_t is the continuation value from not being employed by the firm, and the second component of A_t is the continuation value from being employed by the firm at the optimal effort level. Henceforth, we refer to A_t as the continuation value in period t .

4.2 Contracting problem

We assume that the firm wants to implement a contract that encourages the executive to work in all periods prior to the termination of the contract. Because the effort decision is private and, therefore, not contractible, the contract set by the firm must satisfy an incentive compatibility constraint. Let $\mathcal{S}_t$ denote the firm's information set at t . The incentive compatibility constraint in period t is given by:

$$\alpha_1^{(b_t-1)^{-1}} \mathbb{E}_t \left(v_{t+1} [q_{t+1} + (1 - q_{t+1}) A_{t+1} | d_{1,t} = 1, \mathcal{S}_t] \right) \\ \leq \alpha_0^{(b_t-1)^{-1}} \mathbb{E}_t \left(v_{t+1} [q_{t+1} + (1 - q_{t+1}) A_{t+1} | d_{0,t} = 1, \mathcal{S}_t] \right) \quad (IC)$$

In addition, the firm needs to meet the executive's participation constraint. In period 0, there is no privately observed preference shock. Therefore, the participation constraint in

period 0 is given by

$$\alpha_1^{(b_t-1)^{-1}} \mathbb{E}_t (v_{t+1} [q_{t+1} + (1 - q_{t+1})A_{t+1}]) \leq 1 \quad (IR_0)$$

In subsequent periods, because ϵ_t is not observed by the firm and is unbounded, the participation constraint is set such that the executive will stay with the firm with a probability p_t . Let ϵ_t^* denote the threshold at which the executive, under the equilibrium strategies $(d_{1,\tau})_{\tau < t} = 1$, is indifferent between staying with the firm and quitting.

$$\exp(-\epsilon_t^* b_t^{-1} | d_{2,t} = 0) \alpha_1^{b_t^{-1}} \mathbb{E}_t (v_{t+1} [q_{t+1} + (1 - q_{t+1})A_{t+1}])^{1-b_t^{-1}} = 1$$

Therefore, $p_t = \Pr(\epsilon_t \geq \epsilon_t^*)$. Let $F_\epsilon^{-1}(\cdot)$ be the quantile function of ϵ . Define $\underline{u}_t(p) := \exp(F_\epsilon^{-1}(1-p)(b_t-1)^{-1})$. The participation constraint in $t \geq 1$ is given by

$$\alpha_1^{(b_t-1)^{-1}} \mathbb{E}_t (v_{t+1} [q_{t+1} + (1 - q_{t+1})A_{t+1} | d_{1,t} = 1, \mathcal{S}_t]) \leq \underline{u}_t(p_t | d_{1,t} = 1, \mathcal{S}_t) \quad (IR)$$

Let $\tilde{V}_{t+1}$ denote the firm's gross market value at the beginning of $t+1$ before paying executive compensation, and let V_{t+1} denote the net market value. The distribution of $\tilde{V}_{t+1}$ depends on the information set of the executive $\mathcal{S}_t$ and his private decisions d_t . At $t = 0$, V_0 is the firm's initial market value before hiring the executive. Let $V_{T=t}$ denote the market value of the firm if the executive is replaced at the beginning of period t . Because there is a 1-1 mapping between w_t and v_t , we formulate the firm optimization problem as follows:

$$V_0 = \max_{(p_t, v_{t+1}, q_{t+1} | \mathcal{S}_t)} \mathbb{E}_0 \left[\sum_{t=0}^T \lambda_t \left[\prod_{t'=0}^{t-1} (1 - q_{t'}) p_{t'} \right] \left[(1 - q_t) p_t (\tilde{V}_{t+1} - V_t - w_{t+1}) | \mathcal{S}_t \right] + [1 - (1 - q_t) p_t] V_{T=t} \right] \quad (\mathcal{P})$$

subject to (IR_0) , (IR) , and (IC)

4.2.1 Relaxing the IC constraint

As shown in (IC) , in deciding whether to exert effort in period t , the executive must take into account the continuation value of each decision. Because the optimal contract satisfies (IC) for all future periods, if an executive chooses the equilibrium action $d_{1,t} = 1$, in computing $A_{t+1} | (d_{1,t} = 1)$, it is without loss to assume that the executive will always choose $\{d_{0,\tau}\}_{\tau > t} = 0$. On the contrary, if the executive chooses not to exert effort ($d_{0,t} = 1$), his information set is different from that of the firm. Therefore, the IC constraints in future periods are not

necessarily satisfied, and the optimal future path of actions might involve the executive choosing $d_{0,\tau} = 1$ for some $\tau > t$.

Therefore, to simplify the problem, we consider a relaxed problem for the firm in which the IC constraint is replaced by the following constraint.

$$\begin{aligned} & \alpha_1^{1/(b_t-1)} \left(\mathbb{E}_t \left(v_{t+1} [q_{t+1} + (1 - q_{t+1}) A_{t+1}] \mid d_{1,t} = 1, \mathcal{S}_t \right) \right) \\ & \leq \alpha_0^{1/(b_t-1)} \mathbb{E}_t \left[v_{t+1} \left(q_{t+1} + (1 - q_{t+1}) (A_{t+1} \mid d_{0,t} = 1, \{d_{0,\tau}\}_{\tau>t} = 0, \mathcal{S}_t) \right) \right] \end{aligned} \quad (\tilde{IC})$$

Intuitively, $(\tilde{IC})$ is weaker than (IC) because the RHS of $(\tilde{IC})$ assumes that, conditional on choosing to shirk in period t , the executive cannot choose to shirk in any future periods. Therefore, the RHS of $(\tilde{IC})$ is always weakly greater than that of (IC) , and constraint (IC) implies constraint $(\tilde{IC})$. Figure 4.2 illustrates the difference between the relaxed IC constraint and the original IC constraint for the firm. In the relaxed problem, we restrict attention to only one-shot deviations and omit the green portions of the decision tree when considering the executive's incentive constraints. In Appendix ??, we show that it is without loss to consider the relaxed problem under a sufficient set of conditions on how the latent match value affects the distribution of the gross market value $\tilde{V}$. Essentially, these conditions are satisfied if: (1) an executive who previously shirked always possesses a belief regarding the match value—and consequently, the firm's future payoff—that is superior in a first-order stochastic sense compared to the firm's own belief; and (2) motivating an executive to exert effort becomes easier when their match value is higher. In section 4.2.4, we also illustrate this result using a parametric example.

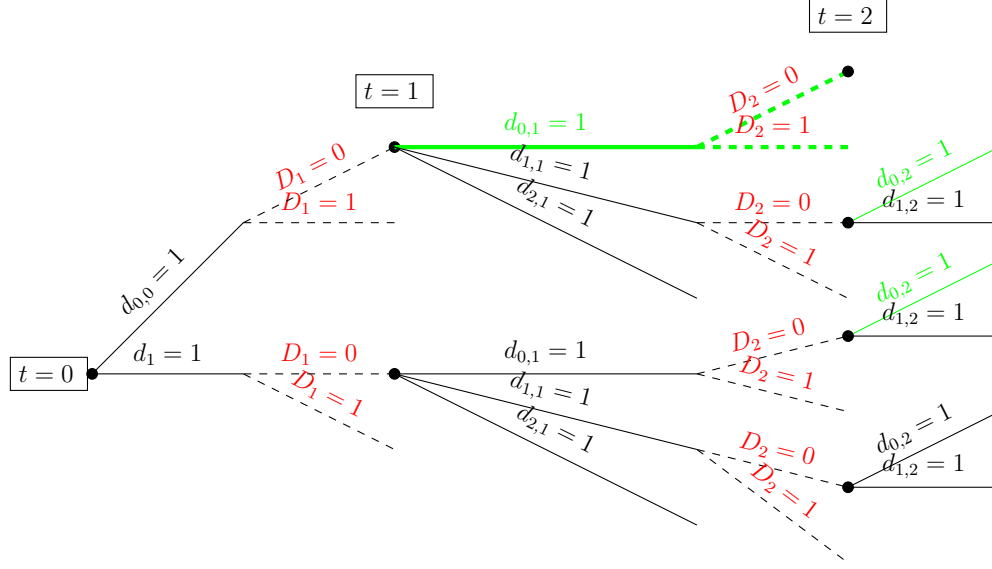
4.2.2 Necessary conditions

Let $\lambda_t^{IC} | \mathcal{S}_t$ and $\lambda_t^{IR} | \mathcal{S}_t$ denote the Lagrangian multipliers associated with the IC and IR constraints in period t , respectively. Let $\lambda_t^{q,0} | \mathcal{S}_t$ and $\lambda_t^{q,1} | \mathcal{S}_T$ denote the Lagrangian multipliers associated with the constraint that $q_t \geq 0$ and $q_t \leq 1$. Let $\lambda_t^w | \mathcal{S}_t$ be the Lagrangian multiplier of constraint $w_t \geq \underline{w}$. Because the idiosyncratic preference shock ϵ_t is unbounded, p_t is always an interior solution (i.e., $p_t \in (0, 1)$). For notational convenience, we normalize λ_t^{IC} and λ_t^{IR} by dividing them by the probability of reaching $\mathcal{S}_t$, i.e., $\prod_{\tau=0}^{t-1} p_\tau (1 - q_\tau) (1 - q_t)$.

For $\tau \leq t$, define

$$\mathcal{S}_t^\tau := (\mathcal{S}_t, \{d_{1,t'}\}_{t' \neq \tau, t' \leq t} = 1, d_{1,\tau} = 0)$$

Figure 4.2: Illustration of the difference between the full problem and the relaxed problem for the firm



$$g_t^\tau := \frac{f_{\tilde{V}}(\tilde{V}_{t+1} | \mathcal{S}_t^\tau)}{f_{\tilde{V}}(\tilde{V}_{t+1} | \mathcal{S}_t^\tau)}$$

$$A_t^\tau := A_t | S_t^* = S_t^\tau$$

$$p_t^\tau := p_t | S_t^* = s_t^\tau$$

$\mathcal{S}_t^\tau$ is the information set of an executive who shirked only in period τ at the beginning of t . g_t^τ is the likelihood ratio of gross market value in t when an executive shirks in period τ relative to when an executive exerts effort in all periods including t . A_t^τ is the continuation value of an executive whose information set is S_t^τ .

Define:

$$K_{t,0} := [\lambda_t^{IR} + \lambda_t^{IC}] \alpha_1^{1/(b_t-1)} v_{t+1} A_{t+1} - \lambda_t^{IC} \alpha_0^{1/(b_t-1)} v_{t+1} A_{t+1}^t g_t \quad (3)$$

$$K_{t,1} := [\lambda_t^{IR} + \lambda_t^{IC}] \alpha_1^{1/(b_t-1)} v_{t+1} - \lambda_t^{IC} \alpha_0^{1/(b_t-1)} v_{t+1} g_t \quad (4)$$

$K_{t,0}$ represents the marginal change in the IC and IR constraints in t with respect to v_{t+1} multiplied by v_{t+1} when the executive is not fired in period $t+1$, and $K_{t,1}$ represents the same marginal change when the executive is fired in period $t+1$.

The optimal contract must satisfy the following first-order conditions.

$$0 = p_t \lambda_t - [(1 - q_{t+1})K_{t,0} + q_{t+1}K_{t,1}] - (1 - q_t)^{-1} \sum_{t' < t} \left[\prod_{\tau=t'}^t (1 - q_\tau) p_\tau \right] \frac{\partial}{\partial v_{t+1}} ((1 - q_{t'+1})K_{t',0}) \quad (FOC_v)$$

$$0 = -\lambda_t \left(p_t \mathbb{E}_t \left[\tilde{V}_{t+1} - w_{t+1} - V_t + (\lambda_{t+1}/\lambda_t V_{t+1} - V_{T,T=t}) \right] \right) - \sum_{t' < t} \left[\prod_{\tau=t'}^t (1 - q_\tau) p_\tau \right] \frac{\partial}{\partial q_t} ((1 - q_{t'+1})K_{t',0} + q_{t'+1}K_{t',1}) - \lambda_t^{q,1} + \lambda_t^{q,0} \quad (FOC_q)$$

$$0 = \lambda_t \mathbb{E}_t \left[\tilde{V}_{t+1} - w_{t+1} - V_t + (\lambda_{t+1}/\lambda_t V_{t+1} - V_{T,T=t}) \right] + \lambda_t^{IR} \underline{u}'(p_t) \quad (FOC_p)$$

(FOC_v) is the FOC w.r.t v_{t+1} multiplied by v_{t+1} . The first term in the LHS of (FOC_v) is the marginal reduction in wage expenses by offering a higher v_{t+1} and thus a lower w_{t+1} . The second term is the marginal effect of increasing v_{t+1} on the IC and IR constraints in period t . The last summation is the aggregate marginal effect of increasing v_{t+1} on previous periods' IC and IR constraints.

(FOC_q) is the FOC w.r.t q_t . The first term in the LHS of (FOC_q) is the difference in the expected profit of the firm when the probability of firing increases. By retaining the executive, conditional on the executive subsequently choosing to stay, the discounted profit of the firm is the expected dividend in period t , represented by $\tilde{V}_{t+1} - V_t - w_{t+1}$, and the market value of the firm in $t + 1$, which is V_{t+1} . If the executive is replaced, the firm's discounted profit is its market value under a new executive $V_{T,T=t}$. In addition, by firing the executive, the IC and IR constraints of previous periods are also affected, which are the remaining expressions in (FOC_q).

Finally, (FOC_p) is the FOC w.r.t p_t . A decrease in p_t has a direct effect on firm profit similar to that of an increase in q_t . In addition, lowering p_t reduces the cost of meeting the executive's IR constraint, which shows up as the second expression in (FOC_p).

4.2.3 Characterizations of the optimal contract

In the optimal contract, involuntary termination occurs for two reasons. First, the firm will terminate the contract with an executive who generates a lower expected payoff for the firm

than that of a newly hired executive by choosing $q_t = 1$. Second, for an executive whose performance is in an intermediate range, the company will terminate the contract with some positive probability $q_t \in (0, 1)$ to discourage the executive from shirking. In our model, an executive has the incentive to choose low effort ($d_{0,t-1} = 1$) in period $t - 1$ to (1) save the non-pecuniary cost of effort and (2) earn a potentially more favorable contract in the future after a low-performance realization. When setting the contract for an executive with a low past performance, the firm might choose a high-power contract—i.e., rewarding more for good performance—in anticipation that the executive has a low match type. This future compensation path provides a higher payoff for an executive who has previously shirked than for an executive who did not. Therefore, a positive firing probability in period t provides a greater punishment for an executive who has shirked in $t - 1$ than for an executive who has not previously shirked. On the other hand, an executive whose performance is sufficiently high is unlikely to have shirked and, therefore, will not be terminated by the firm. The following proposition formalizes these arguments.

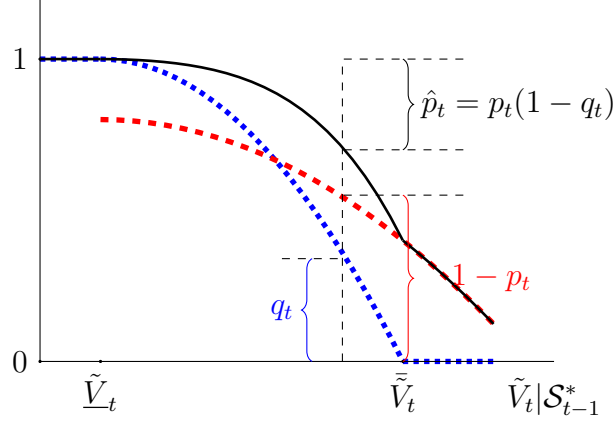
Proposition 1. *Conditional on the firm's information set at beginning of $t - 1$, $\mathcal{S}_{t-1}$, there exist thresholds of gross market value $\underline{\tilde{V}}_t$ and $\bar{\tilde{V}}_t$ such that the contract is terminated with certainty in t if and only if $\tilde{V}_t \leq \underline{\tilde{V}}_t$, and the contract is terminated with some probability $q_t \in (0, 1)$ if and only if $\tilde{V}_{t+1} \in (\underline{\tilde{V}}_t, \bar{\tilde{V}}_t)$. When $\tilde{V}_t > \bar{\tilde{V}}_t$, the executive is never fired — i.e., $q_t = 0$.*

Under the assumption that the distribution of $\tilde{V}_t | \mathcal{S}_{t-1}$ has full support with a continuous and differentiable density function, the necessary conditions $((FOC_p), (FOC_v), (FOC_p))$ are all continuous and differentiable functions of $\tilde{V}_t$. Because p_t is always interior, it is also differentiable and its slope at $\bar{\tilde{V}}_t$ is always positive. Similarly, q_t is also differentiable when $q_t \in (0, 1)$. However, the left-derivative of $q_t(\cdot | \mathcal{S}_{t-1}^*)$ at $\bar{\tilde{V}}_t$ is negative whereas its right-derivative is equal to 0. Therefore, the retention probability, defined as $\hat{p}_t := (1 - q_t)p_t$ exhibits a kink at $\bar{\tilde{V}}_t$. The following Proposition summarizes this argument.

Proposition 2. *Conditional on the information set $\mathcal{S}_t^*$, q_t and p_t are continuous functions of $\tilde{V}_t$. Furthermore, the retention probability $\hat{p}_t := (1 - q_t)p_t$ exhibits a kink only at $\bar{\tilde{V}}_t$.*

Figure 4.3 provides an illustration of Proposition (2). Here, we show the voluntary termination probability $(1 - p_t)$ in red, the firing probability q_t in blue, and the separation probability $1 - \hat{p}_t$ in black. The probability of separation has two components: (1) the probability that the executive is fired q_t and (2) the probability that the executive leaves the firm voluntarily if not being fired $(1 - q_t)(1 - p_t)$. At $\tilde{V}_t \leq \underline{\tilde{V}}_t$, the firing probability is

Figure 4.3: Illustration of the optimal separation probability



Note: The dashed red curve represents the quitting probability $1 - p_t$. This probability is not defined when an executive is fired with certainty (below $\underline{\tilde{V}}_t$). The dashed blue curve represents the firing probability q_t , which is equal to 1 for low outcomes $\tilde{V}_t \leq \underline{\tilde{V}}_t$ and equal to 0 when the executive's performance is sufficiently good ($\tilde{V}_t \geq \tilde{\tilde{V}}_t$). The solid black line represents the separation probability in the data, which is equal to $1 - p_t(1 - q_t)$.

equal to 1, and at $\tilde{V}_t \geq \tilde{\tilde{V}}_t$, the firing probability is equal to 0. The kink on the separation probability happens at $\tilde{\tilde{V}}_t$.

4.2.4 A 2-period parametric example

To make the exposition clearer, let θ denote the match value between the firm and the executive. Let the prior belief about θ be $U[0, 1]$. Let the performance of an executive be denoted by $x_t \in \{0, 1\}$, where $x_t | (d_{1,t} = 1, \theta)$ is equal to 1 with probability θ and equal to 0 with probability $1 - \theta$. Conditional on $d_{0,t} = 1$, x_t is 0 with certainty. The market value of the firm prior to hiring the executive is $V_0 = 0$. Once the executive is hired, the gross value of the firm is $\tilde{V}_{t+1} = V_t + \bar{V}x_t$. When the executive is replaced, the firm value is V_T . We assume that the firm cannot pay a negative wage, that is, $\underline{w} = 0$.

Using Bayes updating, we show that conditional on $x_0 = 1$ and $d_{1,0} = 1$, the posterior belief about θ has density equal to 2θ . Conditional on $x_0 = 0$ and $d_{1,0} = 1$, the posterior belief about θ has density equal to $2(1 - \theta)$. Therefore, $\Pr(x_1 = 1 | (x_0 = 1, d_{1,0} = 1)) = \frac{2}{3}$, and $\Pr(x_1 = 1 | (x_0 = 0, d_{1,0} = 1)) = \frac{1}{3}$. On the other hand, conditional on $d_{0,0} = 1$, the observed output does not contain any additional information on θ . Thus, conditional on the executive having bad (good) performance in period $t = 0$, the expected gross market value of the firm in period $t = 1$ is $1/3\bar{V}$ ($2/3\bar{V}$). We focus on the case in which the firm might

want to retain the executive despite having an initial bad performance by assuming that the firm's gross market value in this case is still higher than the net market value from hiring a replacement, i.e., $1/3\bar{V}$ is higher than $V_{T,T=1} - \lambda_2/\lambda_1 V_{T,T=2}$.

We make two observations. First, conditional on $x_t = 0$, $(w_t|\mathcal{S}_t) = \underline{w}$ for both periods ($t = 1, 2$). Because setting $w_2|(x_1 = 1) = \underline{w}$ is not incentive compatible, it is without loss to ignore the constraint $w_2|(x_1 = 1) \geq \underline{w}$. Second, we focus on the case in which α_0 is sufficiently high that an executive will not accept an offer with a constant wage $\underline{w}$. This implies that both the IC and the IR constraints in period $t = 0$ hold with equality.

Consider $t = 1$. Conditional on x_0 , (IC) simplifies to:

$$\alpha_1^{(b_1-1)^{-1}} [v_2 \Pr(x_1 = 1|d_{1,1} = 1, \mathcal{S}_1^*) + \Pr(x_1 = 0|d_{1,1} = 1, \mathcal{S}_1^*)] \leq \alpha_0^{(b_1-1)^{-1}} \quad (5)$$

When $x_0 = 1$, the IC constraint is equivalent to

$$v_2 \leq 1.5 (\alpha_0/\alpha_1)^{(b_1-1)^{-1}} - 0.5 \quad (6)$$

Whereas, conditional on $x_0 = 0$, the IC constraint becomes

$$v_2 \leq 3 (\alpha_0/\alpha_1)^{(b_1-1)^{-1}} - 2 \quad (7)$$

Now, consider $t = 0$, (IC) can be rewritten as:

$$\begin{aligned} \alpha_1^{(b_0-1)^{-1}} [0.5v_1 (q_1 + (1 - q_1)A_1|x_1 = 1, d_{1,0} = 1) + 0.5 (q_1 + (1 - q_1)A_1|x_1 = 0, d_{1,0} = 1)] \\ \leq \alpha_0^{(b_0-1)^{-1}} (q_1 + (1 - q_1)A_1|x_1 = 0, d_{0,0} = 1) \end{aligned} \quad (8)$$

Our first result shows that it is without loss to consider the relaxed IC constraint ($\tilde{IC}$), which imposes that the executive never finds it optimal to shirk more than once.

Proposition 3. *Consider a contract that sets $v_2|x_1 = 1$ such that it satisfies (7), then the same contract is incentive compatible for an executive who had previously chosen $d_{0,0} = 1$.*

Proof:

Conditional on $d_{0,0} = 1$ and $x_1 = 0$, the contract at $t = 1$ satisfies the IC constraint in period 1 if

$$v_2 \leq 2(\alpha_0/\alpha_1)^{(b_1-1)^{-1}} - 1 \quad (9)$$

Because $(\alpha_0/\alpha_1)^{(b_1-1)^{-1}} < 1$, the inequality in (7) implies (9). ■

The following results establish that it is optimal not to fire an executive after a good performance ($x_0 = 1$). However, when the cost of incentivizing the executive against shirking is high (i.e., when α_1/α_0 is large), it is optimal to fire an executive after a poor performance ($x_0 = 0$).

Proposition 4. *The optimal contract has the following properties:*

- *The executive will not be fired after a good performance, that is, $q_0|(x_0 = 1) = 1$.*
- *When the cost of effort, represented by α_1/α_0 , is sufficiently large, the executive is fired with probability 1 ($q_1|(x_0 = 0) = 1$).*
- *When the cost of effort is in an intermediate region, the executive is fired with a strictly positive probability. That is, $0 < q_1(x_0 = 0) < 1$.*

Proof:

The proof is provided in Appendix A. ■

Intuitively, setting a positive firing probability is useful only if it helps to satisfy the IC constraint in period 0. However, because the executive is risk-averse, setting $q_1 > 0$ causes the participation constraint to be more stringent, requiring the firm to increase wages w_1 and/or w_2 to meet the participation constraint. Following a good performance $x_0 = 1$, the firm is certain that the executive did not shirk in period $t = 0$; therefore, it is optimal to set $q_1|(x_1 = 1) = 0$.

On the other hand, $x_0 = 0$ is likely to occur when an executive shirks in period $t = 0$. Following this deviation, the firm's belief about the executive becomes more pessimistic than the executive's own belief. Specifically, the firm believes that, conditional on the executive putting in effort (choosing $d_{1,1} = 0$), the probability of achieving a high output in $t = 1$ is only $1/3$, in contrast to the actual probability of $1/2$. Consequently, the executive anticipates a higher expected utility in period $t = 1$ if he chooses to shirk in $t = 0$. In this scenario, choosing a positive probability of firing acts as an effective deterrent against moral hazard, as the utility loss from being terminated is now greater for an executive who has shirked compared to one who has not. However, since implementing a positive firing probability is also costly for the firm, it is more inclined to resort to choosing $q_1 > 0$ when motivating the executive through wages alone becomes more difficult, i.e., when the ratio α_1/α_0 is sufficiently large.

5 Identification and Estimation

5.1 Identification

In this section, we assume that we observe a sample of homogeneous firms and ex-ante homogeneous executives.² The primitive objects of our model are the parameters that govern the preferences of the executive $\{\alpha_0, \alpha_1, \beta, \rho\}$ and the likelihood ratio of the firm's gross market value when the executive deviates from the equilibrium actions ($d_{1,t} = 1$), which are $\{g_t^\tau\}_{\tau \leq t} | \mathcal{S}_t$. We assume that $\mathcal{S}_t = \{\tilde{V}_\tau\}_{0 \leq \tau \leq t}$, and is order irrelevant. In addition, because we do not observe bunching in the lower support of w_{t+1} in the data, we assume that the constraint $w_{t+1} \geq \underline{w}$ does not bind. Consequently, this implies that $q_1 \in [0, 1)$.

Specifically, we observe $\mathcal{S}_t$, $\tilde{V}_{t+1}$, V_{t+1} , w_{t+1} , whether an executive stays with the firm, and the prices of financial securities (a_t, b_t, λ_t) . We do not observe whether an executive was fired. From the data, the distribution of $(\tilde{V}_{t+1} | \mathcal{S}_t, d_{1,t} = 1)$ and $(w_{t+1} | \mathcal{S}_t)$, and the retention probability $\hat{p}_t$ are identified.

From Proposition 1, there exists a change point in the graph of $\hat{p}_{t+1} | (\cdot, \mathcal{S}_t)$ at $\tilde{V}_{t+1}$, which is also identified from the observed retention data. Therefore, the indicator variable $1_{q_{t+1}=0} | \mathcal{S}_t$ is identified.

Our identification argument proceeds as follows. We first identify α_1 and ρ using the information on compensation in the period immediately before mandatory retirement ($t = \bar{T} - 1$). In this period, the firm does not set a positive firing probability; therefore, $p_t = \hat{p}_t$. In addition, the continuation value A_{t+1} is always equal to 1. Therefore, at $t = \bar{T} - 1$, the identification problem is similar to that of a risk-averse executive facing a lottery of compensation and prefers the lottery over an outside option with probability p_t . Because compensation depends on $\mathcal{S}_t$, there is variation in the distribution of the lotteries' payoffs across different $\mathcal{S}_t$. The following assumption requires that this variation is sufficiently rich.

Assumption 1. *There exist realizations s and s' of $\mathcal{S}_t$ such that the likelihood ratio between the distribution of compensation $w_{t+1} | (\mathcal{S}_{t-1} = s)$ and that of $w_{t+1} | (\mathcal{S}_{t-1} = s')$ is monotone in w_{t+1} .*

Assumption 1 states that there exists two realizations s and s' of $\mathcal{S}_t$ such that the likelihood ratio between the distribution of compensation in $t + 1$ (before $\tilde{V}_{t+1}$ is realized) conditional on s and that conditional on s' is monotone. For example, if the distribution

²In Section ??, we explain how our estimation procedure takes into account observed heterogeneity across firms and across executives.

of $w_{t+1}|s$ first-order stochastic dominates $w_{t+1}|s'$, Assumption 1 is satisfied. In addition, because w_{t+1} and $\mathcal{S}_t$ are observed, this assumption can be tested.

Proposition 5. *Under Assumptions 1, α_1 and ρ are identified.*

Next, we identify $A_1|\mathcal{S}_0, \tilde{V}_1, A_1^0|\mathcal{S}_0, \tilde{V}_1, g_0^0|\mathcal{S}_0, q_1|\mathcal{S}_0$, and α_0 using the data at $t = 0$. For notational convenience, we drop the conditioning variables whenever it does not give rise to confusion.

Proposition 6. *A_1, A_1^0, g_0^0, q_1 , and α_0 are identified.*

Proof:

From the FOCs w.r.t v_1 and q_1 and the expected value of the FOCs w.r.t v_2 , we obtain the following equations:

$$0 = \lambda_0 - q_1 K_{0,1} - (1 - q_1) K_{0,0} \quad (10)$$

$$0 = -p_1 \lambda_1 \left[\mathbb{E} \left(\tilde{V}_2 - w_2 - V_1 + \lambda_2 / \lambda_1 V_2 | \mathcal{S}_1^* \right) - V_{T,T=1} \right] - K_{0,1} + K_{0,0} - \lambda_0^{q,1} + \lambda_1^{q,0} \quad (11)$$

$$0 = \lambda_1 p_1 - \lambda_1^{IR} \underline{u}(p_1) - (\lambda_0^{IR} + \lambda_0^{IC})(A_1 - (1 - p_1)) + \lambda_0^{IC}(A_1^0 - (1 - p_1^0))g_0^0 \quad (12)$$

Taking expectation of (10) over v_1 yields $0 = \lambda_0 - \lambda_0^{IR}$. therefore, $\lambda_0^{IR} = \lambda_0$. We then consider $w_1 \rightarrow \bar{w}$, which implies $g_0 \rightarrow 0$. Therefore, $\lambda_0^{IR} + \lambda_0^{IC}$ is identified and is equal to:

$$\lambda_0^{IR} + \lambda_0^{IC} = \lambda_0 \alpha_1^{-1/(b_0-1)} \left[\lim_{w_1 \rightarrow \bar{w}} v_1 A_1 \right]^{-1}$$

Thus, both λ_0^{IR} and λ_0^{IC} are identified. Because there is an identified 1-1 mapping between A_1 and p_1 and between A_1^0 and p_1^0 , the remaining objects to be identified in (10)-(12) are $\{g_0^0, A_1, A_1^0\}$ and $\{\lambda_1^{IR}, \lambda_1^{IC}\}$. Using similar arguments as in the identification argument for λ_0^{IR} and λ_0^{IC} , $\{\lambda_1^{IR}, \lambda_1^{IC}\}$ are also identified if A_1 is identified.

Now, we show that $\{g_0^0, A_1^0, A_1\}$ are identified. Because $1_{q_1=0}$ is identified, we consider two separate cases. In the first case, $q_1 \in (0, 1)$, and $\lambda_1^{q,0} = 0$. Multiply (11) by $(1 - q_1)$ and add to (10):

$$0 = \lambda_0 - \hat{p}_1 \lambda_1 \left[\mathbb{E}(\tilde{V}_2 - w_2 - V_1 + \lambda_2 / \lambda_1 V_2 | \mathcal{S}_1^*) - V_{T,T=1} \right] - K_{0,1} \quad (13)$$

$K_{0,1}$ is thus identified from (13); therefore, g_0^0 is also identified. For a given value of A_1 , q_1 is identified and is equal to $1 - \hat{p}_1 / p_1$. In addition, (10) identifies A_1^0 as an increasing function

of A_1 . In the last step, we show that (12) is a monotonic function of A_1 using Lemma B.3 in the Appendix. Therefore, A_1 is also identified.

In the second case, $q_1 = 0$, therefore, A_1 is identified from $p_1 = \hat{p}_1$. (12) then pins down $(A_1^0 - (1 - p_1^0))g_0^0$, and $A_1^0 g_0^0$ is identified from (10). From Lemma B.3, $\frac{A_1^0 - (1 - p_1^0)}{A_1^0}$ is monotonic in p_1^0 , therefore, both p_1^0 (and thus A_1^0) and g_0^0 are identified. ■

In subsequent periods $t > 0$, the identification argument follows a strong induction argument.

Proposition 7. *If $\{\alpha_1, \alpha_0, \rho\}$ and $\{A_{t+1}^\tau, A_{t+1}, q_{\tau+1}, g_t^\tau, \lambda_\tau^{IC}, \lambda_\tau^{IC}\}_{\tau \leq t}$ are identified, then $A_{t+2}^\tau, q_{t+2}, \lambda_{t+1}^{IC}, \lambda_{t+1}^{IR}, (g_{t+1}^\tau)_{\tau \leq t+1}$ are identified.*

The proof is included in Appendix B.1. Here, we utilize the observations that there exist t different values of $\mathcal{S}_t$ that produce the same set of likelihood ratios $\{g_t^\tau\}_{\tau \leq t}$ and continuation values $(A_{t+1}, \{A_{t+1}^\tau\}_{\tau \leq t})$. For example, at $t = 3$, $\mathcal{S}_t = (\tilde{V}_1, \tilde{V}_2, \tilde{V}_3)$ and $\mathcal{S}'_t = (\tilde{V}_3, \tilde{V}_1, \tilde{V}_2)$ yield the same $\{g_3^\tau\}_{\tau \leq 3}$ and $(A_4, \{A_4^\tau\}_{\tau \leq 3})$. Therefore, we have a system of $2t + 1$ equations, which include t FOC w.r.t q_{t+1} and v_{t+1} and an expected FOC w.r.t v_{t+2} . This system of equations have $2t + 1$ unknowns, which are $\{A_{t+1}, \{A_{t+1}^\tau\}_{\tau \leq t}, \{g_t^\tau\}_{\tau \leq t}\}$. In the proof, we show that this system of equations has a unique solution.

A Proof of Section 4

A.1 Proof of Proposition 4

Proof of statement 1 of Proposition 4:

With some algebra, (FOC_q) can be expressed as

$$0 = -p_1(1 - q_1)\lambda_1 [\mathbb{E}(\bar{V}x_1 - w_2|\mathcal{S}_1^*) - (V_1 - \lambda_2/\lambda_1 V_2)] + \lambda_0 - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{b_0^{-1}}v_1 + \lambda_0^{IC}\alpha_0^{b_0^{-1}}v_1g_0^0 + \lambda^{q_1,0} \quad (14)$$

The first term in (14) represents the marginal loss in profit for the firm from firing the executive. The next three terms $(\lambda_0 - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{b_0^{-1}}v_1 + \lambda_0^{IC}\alpha_0^{b_0^{-1}}v_1g_0)$ represent the impact of q_1 on the incentive compatibility constraint and the participation constraint of the executive in period 1.

Conditional on $x_0 = 1$, $g_0^0 = 0$. Therefore, (14) simplifies to

$$0 = -p_1(1 - q_1)\lambda_1 [\mathbb{E}(\bar{V}x_1 - w_2|\mathcal{S}_1^*) - (V_1 - \lambda_2/\lambda_1 V_2)] + \lambda_0 - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{b_0^{-1}}v_1 + \lambda^{q_1,0} \quad (15)$$

Conditional on $x_0 = 1$, (FOC_v) simplifies to

$$0 = \lambda_0 v_1^{-1} - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{b_0^{-1}}[q_1 + (1 - q_1)A_1] \quad (16)$$

Because $A_1 \leq 1$, $(\lambda_0 - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{b_0^{-1}}v_1 + \lambda_0^{IC}\alpha_0^{b_0^{-1}}v_1g_0) < 0$. Combining (16) and (14) implies $\lambda^{q_1,0} > 0$. Therefore, $q_1|(x_1 = 1) = 0$. ■

Proof of statement 2 of Proposition 4:

To incentivize the agent to work, conditional on $x_0 = 0$ and $x_1 = 1$, (9) must hold. However, as $\alpha_0/\alpha_1 \rightarrow (0.5)^{b_1-1}$, the RHS of (9) approaches 0, which implies $w_2 \rightarrow \infty$. Therefore, the firm cannot induce the executive to work, and thus $\mathbb{E}(x_1|x_0 = 0) = 0$. This implies that the firm is better off replacing the executive by setting $q_1 = 1$. ■

Proof of statement 3 of Proposition 4:

Conditional on choosing $d_0 = 1$ and obtaining $x_0 = 0$, the executive's continuation value if he stays with the firm at the beginning of period 1 is given by:

$$\mathbb{E}(\exp(-\epsilon_1 b_1^{-1}|d_{2,1} = 0))\alpha_1^{b_1^{-1}}(0.5v_2 + 0.5)^{1-b_1^{-1}} \quad (17)$$

Conditional on choosing $d_1 = 0$ and obtaining $x_0 = 0$, the executive's continuation value if

he stays with the firm at the beginning of period 1 is given by

$$\mathbb{E} \left(\exp(-\epsilon_1 b_1^{-1} | d_{2,1} = 0) \alpha_1^{b_1^{-1}} (1/3 v_2 + 2/3)^{1-b_1^{-1}} \right) \quad (18)$$

In period $t = 0$, the IC and IR constraints both bind. Therefore:

$$\alpha_0^{b_0^{-1}} (q_1 + (1 - q_1) A_1(x_0 = 0, d_{0,0} = 1)) = 1 \quad (19)$$

$$\alpha_1^{b_0^{-1}} [0.5(v_1 A_1 | x_0 = 1) + 0.5(q_1 + (1 - q_1) A_1(x_0 = 0, d_{1,0} = 1))] = 1 \quad (20)$$

Therefore,

$$A_1(x_0 = 0, d_{0,0} = 1) = \left[\alpha_0^{-b_0^{-1}} - q_1 \right] (1 - q_1)^{-1} \quad (21)$$

Combining the FOC w.r.t $v_1 | (x_0 = 1)$ and (20)

$$(\lambda_0^{IC} + \lambda_0^{IR}) \alpha_1^{b_0^{-1}} = \lambda_0 v_1^{-1} A_1(x_0 = 1)^{-1} \quad (22)$$

$$= \lambda_0 \left[2\alpha_1^{-b_0^{-1}} - (q_1 + (1 - q_1) A_1(x_0 = 0, d_{1,0} = 1)) \right]^{-1} \quad (23)$$

Combining the above equation and the FOC w.r.t q_1 yields

$$\begin{aligned} 0 &= -p_1(1 - q_1) \lambda_1 \left[\mathbb{E}(\bar{V} x_1 - w_2 | \mathcal{S}_1^*) - (V_1 - \lambda_2/\lambda_1 V_2) \right] + \lambda_0 \\ &\quad - (\lambda_0^{IC} + \lambda_0^{IR}) \alpha_1^{b_0^{-1}} + 2\lambda_0^{IC} \alpha_0^{b_0^{-1}} + \lambda^{q_1,0} \end{aligned} \quad (24)$$

Therefore, if $q_1 | (x_0 = 0) = 0$ is optimal, the following must be true at $q_1 = 0$

$$-p_1 \lambda_1 \left[\mathbb{E}(\bar{V} x_1 - w_2 | \mathcal{S}_1^*) - (V_1 - \lambda_2/\lambda_1 V_2) \right] + \lambda_0 - (\lambda_0^{IC} + \lambda_0^{IR}) \alpha_1^{b_0^{-1}} \leq 0 \quad (25)$$

Note that

$$\begin{aligned} 0 &\geq LHS(25) \geq -\lambda_1 (1/3 \bar{V} - (V_{T,T=1} - \lambda_2/\lambda_1 V_{T,T=2})) + \lambda_0 - \lambda_0 \left[2\alpha_1^{-b_0^{-1}} - A_1(x_0 = 0, d_{1,0} = 1) \right] \\ &\geq -\lambda_1 (1/3 \bar{V} - (V_{T,T=1} - \lambda_2/\lambda_1 V_{T,T=2})) + \lambda_0 - \lambda_0 \left[2\alpha_1^{-b_0^{-1}} - A_1(x_0 = 0, d_{0,0} = 1) \right] \\ &= -\lambda_1 (1/3 \bar{V} - (V_{T,T=1} - \lambda_2/\lambda_1 V_{T,T=2})) + \lambda_0 - \lambda_0 \left[2\alpha_1^{-b_0^{-1}} - \alpha_0^{-b_0^{-1}} \right] \end{aligned} \quad (26)$$

However, for $\alpha_0/\alpha_1 < \left[2 - (-\lambda_1/\lambda_0(1/3\bar{V} - (V_{T,T=1} - \lambda_2/\lambda_1 V_{T,T=2})) + 1) \alpha_1^{b_0^{-1}}\right]^{-b_0}$, the last expression in (26) is positive. Therefore, this is a contradiction. ■

A.2 Proof of Proposition 5

Proof of Proposition 5:

Let $t = \bar{T} - 1$. Suppose that there exist (α_1, ρ) and $(\tilde{\alpha}_1, \tilde{\rho})$ that can rationalize the data, where $\alpha \neq \tilde{\alpha}$ or $\rho \neq \tilde{\rho}$ or both. Because $A_{\bar{T}} = 1$, from (IR), the following equality must hold for both pairs (α_1, ρ) and $(\tilde{\alpha}_1, \tilde{\rho})$ at any $(\mathcal{S}_t^*)$.

$$\underline{u}(p_t|\mathcal{S}_t^*) = \mathbb{E} \left(\alpha_1^{(b_t-1)^{-1}} \exp(-\rho w_{t+1}/b_{t+1}) \right) \quad (27)$$

$$\underline{u}(p_t|\mathcal{S}_t^*) = \mathbb{E} \left(\tilde{\alpha}_1^{(b_t-1)^{-1}} \exp(-\tilde{\rho} w_{t+1}/b_{t+1}) \right) \quad (28)$$

Without loss of generality, we assume that $\rho > \tilde{\rho}$. Thus, $\alpha_1 < \tilde{\alpha}_1$. Taking the difference between the RHS of the previous equations:

$$\mathbb{E} \left[\alpha_1^{(b_t-1)^{-1}} \exp(-\rho w_{t+1}/b_{t+1}) - \tilde{\alpha}_1^{(b_t-1)^{-1}} \exp(-\tilde{\rho} w_{t+1}/b_{t+1}) \mid \mathcal{S}_{t-1}^*, \tilde{V}_t \right] = 0 \quad (29)$$

Denote $\Delta(w) = \alpha_1^{(b_t-1)^{-1}} \exp(-\rho w/b_{t+1}) - \tilde{\alpha}_1^{(b_t-1)^{-1}} \exp(-\tilde{\rho} w/b_{t+1})$. With some algebra, it can be shown that $\Delta(w)$ crosses 0 only once from below.

Let $f_w(\cdot | (\mathcal{S}_{t-1}^*, \tilde{V}_t))$ denote the distribution of compensation. Under Assumption 1, there exists s, s' such that $\frac{f_w(\cdot | (\mathcal{S}_t^* = s'))}{f_w(\cdot | (\mathcal{S}_t^* = s))}$ is monotone. Let $\bar{w}$ denote the upper bound of the support of $w_{t+1} | \mathcal{S}_{t+1} \in \{s, s'\}$. WLOG, we assume that the likelihood ratio is decreasing, which implies that $\bar{w}$ is the upper bound of the support of $w_{t+1} | \mathcal{S}_{t+1} = s'$. Consider (29) at $(\mathcal{S}_t^* = s')$:

$$0 = \int_{\underline{w}}^{\bar{w}} \Delta(w) f_w(w | (\mathcal{S}_{t-1}^*, \tilde{V}_t)) \frac{f_w(w | (\mathcal{S}_{t-1}^*, \tilde{V}_t'))}{f_w(w | (\mathcal{S}_{t-1}^*, \tilde{V}_t))} dw \quad (30)$$

$$\implies 0 = - \int_{\underline{w}}^{\bar{w}} \int_{\underline{w}}^w \Delta(w') f_w(w' | (\mathcal{S}_{t-1}^*, \tilde{V}_t)) dw' \frac{\partial}{\partial w} \frac{f_w(w | (\mathcal{S}_{t-1}^*, \tilde{V}_t'))}{f_w(w | (\mathcal{S}_{t-1}^*, \tilde{V}_t))} dw \quad (31)$$

where (31) uses integration by parts and (29). Because $\Delta(w)$ crosses 0 only once from below, $\int_{\underline{w}}^w \Delta(w') f_w(w' | (\mathcal{S}_{t-1}^*, \tilde{V}_t)) dw' < 0$ for any $w' < \bar{w}$. The likelihood ratio is decreasing; therefore, the LHS of (31) is strictly positive. This is a contradiction. ■

B Proofs of Section 5

First, we establish the following lemmas. The proofs of B.1 and B.2 follow directly from the definition of $K_{t-1,0}$ and are therefore omitted.

Lemma B.1.

$$(1 - q_{t+1}) \frac{\partial}{\partial q_{t+1}} K_{t-1,0} + v_{t+1} \frac{\partial}{\partial v_{t+1}} K_{t-1,0} = (\lambda_{t-1}^{IR} + \lambda_{t-1}^{IC}) v_t \frac{A_{t,0} - (1 - p_t)}{\mathbb{E}[v_{t+1}(q_{t+1} + A_{t+1}(1 - q_{t+1}))]} v_{t+1} \\ - \lambda_{t-1}^{IC} v_t \frac{A_{t,0}^{t-1} - (1 - p_t^{t-1})}{\mathbb{E}[v_{t+1}(q_{t+1} + A_{t+1}^{t-1}(1 - q_{t+1}))]} v_{t+1} g_t^{t-1}$$

Lemma B.2.

$$\mathbb{E}_t \left[v_{t+1} \frac{\partial}{\partial v_{t+1}} K_{t-1,0} \right] = (\lambda_{t-1}^{IR} + \lambda_{t-1}^{IC}) v_t (A_{t,0} - (1 - p_t)) - \lambda_{t-1}^{IC} v_t (A_{t,0}^{t-1} - (1 - p_t^{t-1})) g_{t-1}^{t-1}$$

Lemma B.3. *For any t , $A_t - (1 - p_t)$ is increasing in p_t .*

Proof of Lemma B.3:

Let ϵ_t^* denote the threshold of ϵ_t at which the executive is indifferent between staying and (voluntarily) leaving. Therefore, $\epsilon_t^* = F_\epsilon^{-1}(1 - p_t)$.

$$\frac{\partial}{\partial p_t} (A_t - (1 - p_t)) = \exp(-\epsilon_t^*) f_\epsilon(\epsilon_t^*) - \mathbb{E} [\exp(-\epsilon_t b_t^{-1}) | \epsilon_t > \epsilon_t^*] p_t \exp(-\epsilon_t^* b_t^{-1}) b_t^{-1} f_\epsilon(\epsilon_t^*) \\ \geq \exp(-\epsilon_t^* b_t^{-1}) f_\epsilon(\epsilon_t^*) - \int_{\epsilon_t^*}^{\infty} \exp(\epsilon_t b_t^{-1}) f_\epsilon(\epsilon_t) d\epsilon_t \\ \geq 0$$

where the first inequality is because $b_t > 1$. ■

B.1 Proof of Proposition 7

Proof of Proposition 7:

When p_t is identified, λ_t^{IR} is identified from the FOC w.r.t p_t . In addition, as $w_{t+1} \rightarrow \bar{w}$, $g_t^\tau \rightarrow 0$ and $q_{t+1} \rightarrow 0$. Therefore, $\lim_{w_{t+1} \rightarrow \bar{w}} A_{t+1}$ and $\lim_{w_{t+1} \rightarrow \bar{w}} \frac{\partial}{\partial v_{t+1}} K_{\tau,0}(\mathcal{S}_\tau^* | \mathcal{S}_t^*, \tilde{V}_{t+1})$ are identified for any $\tau < t$. From the FOC of v_{t+1} as $w_{t+1} \rightarrow \bar{w}$:

$$0 = \lambda_t - (\lambda_t^{IR} + \lambda_t^{IC}) \alpha_1^{1/(b_{t+1}-1)} \left(\lim_{w_{t+1} \rightarrow \bar{w}} v_{t+1} A_{t+1} \right)$$

$$-\sum_{\tau=0}^{t-1} \lim_{w_{t+1} \rightarrow \infty} v_{t+1} \mathbb{E}_{t+1} \lim_{w_{t+1} \rightarrow \bar{w}} \frac{\partial}{\partial v_{t+1}} K_{\tau,0} \quad (32)$$

Therefore, λ_t^{IC} is identified from (32).

To shorten notation, let $\Delta\Pi_t = \mathbb{E}_t \left[\tilde{V}_{t+1} - w_{t+1} - V_t + (\lambda_{t+1}/\lambda_t V_{t+1} - V_{T,T=t}) \right]$. Adding the FOC w.r.t v_{t+1} and the FOC w.r.t q_{t+1} multiplied by $(1 - q_{t+1})$ yields:

$$\begin{aligned} 0 = & \lambda_t - (\lambda_t^{IC} + \lambda_t^{IR}) \alpha_1^{1/(b_t-1)} v_{t+1} + \lambda_t^{IC} \alpha_0^{1/(b_t-1)} v_{t+1} g_t^t - \lambda_t \Delta\Pi_t + \lambda^{q_{t+1},0} \\ & - v_{t+1} \mathbb{E}_t \sum_{\tau=0}^{t-1} \left[(\lambda_\tau^{IR} + \lambda_\tau^{IC}) \prod_{t'=\tau+1}^{t-1} \frac{\partial A_{t'}}{\partial A_{t'+1}} \frac{A_t - (1 - p_t)}{\mathbb{E}[v_{t+1}(q_{t+1} + A_{t+1}(1 - q_{t+1}))]} \right. \\ & \left. - \lambda_\tau^{IC} \prod_{t'=\tau+1}^{t-1} \frac{\partial A_{t'}^\tau}{\partial A_{t'+1}^\tau} \frac{A_t^\tau - (1 - p_t^\tau)}{\mathbb{E}[v_{t+1}(q_{t+1} + A_{t+1}^\tau(1 - q_{t+1}))]} g_{t+1}^\tau \right] \end{aligned} \quad (33)$$

In (33), the terms that have not yet been identified are $\lambda^{q_{t+1},0}$ and $(g_t^\tau)_{\tau \leq t}$.

As noted in the main text, there exists t values of $\mathcal{S}_t$ that yields the same $(g_t^\tau)_{\tau \leq t}$ and $\{A_{t+1}^\tau\}_{\tau \leq t}$. Rewrite (33) as follows:

$$\sum_{\tau=0}^t m_{\mathcal{S}_t, \tau}^1 g_t^\tau = l_{\mathcal{S}_t}^1 - \lambda^{q_{t+1},0}(\mathcal{S}_t) \quad (34)$$

where $m_{\mathcal{S}_t, \tau}^1 > 0$, and $m_{\mathcal{S}_t, \tau}^1$ and $l_{\mathcal{S}_t}^1$ are already identified.

Similarly, we rewrite the FOC w.r.t v_{t+1} as follows:

$$-A_{t+1}(1 - q_{t+1}) + \sum_{\tau=0}^t m_{\mathcal{S}_t, \tau}^2 (1 - q_{t+1}) A_{t+1}^\tau g_t^\tau = l_{\mathcal{S}_t}^2 \quad (35)$$

where $m_{\mathcal{S}_t, \tau}^2, l_{\mathcal{S}_t}^2 \geq 0$ and $m_{\mathcal{S}_t, \tau}$ and $l_{\mathcal{S}_t, \tau}$ are already identified.

Consider the FOC w.r.t v_{t+2} and taking expectation over v_{t+2} . This equation can be rewritten as:

$$-\lambda_{IR}^{t+1} \underline{u}(p_{t+1}) + \sum_{\tau=0}^t m_{\mathcal{S}_t, \tau}^3 (A_{t+1}^\tau - (1 - p_t^\tau)) g_t^\tau = l^3 \quad (36)$$

where $m_{\mathcal{S}_t, \tau}^3 > 0$ and l_3 are all identified.

Conditional on A_{t+1} being identified, q_{t+1} is identified. Therefore, the system of equations in (35) identifies $A_{t+1}^\tau g_t^\tau$ for any $\tau \leq t$. From (34), g_t^τ is a linear identified mapping of

$\{\lambda^{q_{t+1},0}(\mathcal{S}_t)\}$. Combined with (33), $\{\lambda^{q_{t+1},0}(\mathcal{S}_t)\}$ is identified. Therefore, conditional on A_{t+1} being identified, $\{g_t^\tau\}_{\tau \leq t}$ and $\{A_{t+1}^\tau\}_{\tau \leq t}$ identified.

Now, we show that A_{t+1} is identified. If there exists $\mathcal{S}_t$ such that $q_{t+1} = 0$, A_{t+1} is identified. We now consider the case for which $q_{t+1} > 0$ for all $\mathcal{S}_t$. In this case, (34) identifies g_t^τ because $\lambda^{q_{t+1},0} = 0$. From (35), A_{t+1}^τ is an identified *increasing* function of A_{t+1} . From Lemma B.3, $A_{t+1}^\tau - (1 - p_t^\tau)$ is increasing in p_t^τ , and is therefore decreasing in A_{t+1}^τ and A_{t+1} . Therefore, the LHS of (36) is an increasing identified function of A_{t+1} . Thus, A_{t+1} is identified. ■

References

- ANTONOVICS, K. AND L. GOLAN (2012): “Experimentation and job choice,” *Journal of Labor Economics*, 30, 333–366.
- ATAL, J. P., H. FANG, M. KARLSSON, AND N. R. ZIEBARTH (2020): “Long-term health insurance: Theory meets evidence,” Tech. rep., National Bureau of Economic Research.
- BHASKAR, V. AND G. J. MAILATH (2019): “The curse of long horizons,” *Journal of Mathematical Economics*, 82, 74–89.
- CISTERNAS, G. (2018): “Two-sided learning and the ratchet principle,” *The Review of Economic Studies*, 85, 307–351.
- CLEMENTI, G. L. AND H. A. HOPENHAYN (2006): “A theory of financing constraints and firm dynamics,” *The Quarterly Journal of Economics*, 121, 229–265.
- DEMARZO, P. M. AND Y. SANNIKOV (2006): “Optimal security design and dynamic capital structure in a continuous-time agency model,” *The journal of Finance*, 61, 2681–2724.
- (2016): “Learning, termination, and payout policy in dynamic incentive contracts,” *The Review of Economic Studies*, 84, 182–236.
- DEMSETZ, H. AND K. LEHN (1985): “The structure of corporate ownership: Causes and consequences,” *Journal of political economy*, 93, 1155–1177.
- FUDENBERG, D., B. HOLMSTROM, AND P. MILGROM (1990): “Short-term contracts and long-term agency relationships,” *Journal of economic theory*, 51, 1–31.
- GAYLE, G.-L., L. GOLAN, AND R. A. MILLER (2015): “Promotion, turnover, and compensation in the executive labor market,” *Econometrica*, 83, 2293–2369.
- GAYLE, G.-L. AND R. A. MILLER (2009): “Has moral hazard become a more important factor in managerial compensation?” *American Economic Review*, 99, 1740–69.
- (2015): “Identifying and testing models of managerial compensation,” *The Review of Economic Studies*, 82, 1074–1118.
- GHILI, S., B. HANDEL, I. HENDEL, AND M. D. WHINSTON (2021): “Optimal long-term health insurance contracts: characterization, computation, and welfare effects,” .

- HENDEL, I. AND A. LIZZERI (2003): “The role of commitment in dynamic contracts: Evidence from life insurance,” *The Quarterly journal of economics*, 118, 299–328.
- HERMALIN, B. E. AND M. S. WEISBACH (1998): “Endogenously chosen boards of directors and their monitoring of the CEO,” *American Economic Review*, 96–118.
- HOFMANN, A. AND M. BROWNE (2013): “One-sided commitment in dynamic insurance contracts: Evidence from private health insurance in Germany,” *Journal of Risk and Uncertainty*, 46, 81–112.
- HOLMSTRÖM, B. (1999): “Managerial incentive problems: A dynamic perspective,” *The review of Economic studies*, 66, 169–182.
- JENTER, D. AND K. LEWELLEN (2021): “Performance-induced CEO turnover,” *The Review of Financial Studies*, 34, 569–617.
- JOVANOVIC, B. (1979): “Job matching and the theory of turnover,” *Journal of political economy*, 87, 972–990.
- LYMAN, N. (2023): “Reputation, Entrenchment, and Dynamic Managerial Incentives,” *Working Paper*.
- MALCOMSON, J. M. AND F. SPINNEWYN (1988): “The multiperiod principal-agent problem,” *The Review of Economic Studies*, 55, 391–407.
- MARGIOTTA, M. M. AND R. A. MILLER (2000): “Managerial compensation and the cost of moral hazard,” *International Economic Review*, 41, 669–719.
- MILLER, R. A. (1984): “Job matching and occupational choice,” *Journal of Political Economy*, 92, 1086–1120.
- PARRINO, R. (1997): “CEO turnover and outside succession a cross-sectional analysis,” *Journal of financial Economics*, 46, 165–197.
- PASTORINO, E. (Forthcoming): “Careers in firms: the role of learning and human capital,” *Journal of Political Economy*.
- PETERS, F. S. AND A. F. WAGNER (2014): “The executive turnover risk premium,” *The Journal of Finance*, 69, 1529–1563.

- PRAT, J. AND B. JOVANOVIĆ (2014): “Dynamic contracts when the agent’s quality is unknown,” *Theoretical Economics*, 9, 865–914.
- REY, P. AND B. SALANIE (1990): “Long-term, short-term and renegotiation: On the value of commitment in contracting,” *Econometrica: Journal of the Econometric Society*, 597–619.
- SPEAR, S. E. AND C. WANG (2005): “When to fire a CEO: Optimal termination in dynamic contracts,” *Journal of Economic Theory*, 120, 239–256.
- TAYLOR, L. A. (2010): “Why are CEOs rarely fired? Evidence from structural estimation,” *The Journal of Finance*, 65, 2051–2087.