Optimal Long-term Executive Contracts

Robert A. Miller* Anh Nguyen†

This version: June 2025
PRELIMINARY AND INCOMPLETE

Abstract

This paper explores the contracting problem when firms and executives learn about their match qualities through realized performances, which are affected by executives' hidden effort choices. Because of the interaction between moral hazard and match value uncertainty, we show that firms use both performance pay and involuntary termination as incentives. Using observed compensation and separation data for US public companies from 1992 to 2022, our structural estimates provide an estimate of involuntary termination that relies only on the optimality of contracts and show that executives' anticipation of the ratchet effect impedes the firms' ability to motivate executives only through incentives. Our counterfactual exercises decompose the welfare loss from moral hazard from that of match value uncertainty and quantify the firm's profit gain from its ability to commit to long-term contracts.

Keywords: Long-term contract, learning, moral hazard, executive compensation, executive turnover.

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

1 Introduction

In the executive labor market, the duration of a relationship between a firm and an executive largely depends on their match quality. This relationship typically ends when beliefs about the match quality become sufficiently pessimistic. Furthermore, these beliefs develop dynamically based on the executive's performance, which is, in turn, influenced by the executive's various activities that are challenging to monitor.

In this paper, we investigate how uncertainty in match quality and moral hazard interact to shape executive compensation contracts, utilizing data from U.S. public companies spanning 1992 to 2022. We develop a structural model of dynamic contracting between a long-lived firm and a long-lived, risk-averse executive with access to the financial market. In this model, the firm aims to maximize its market value, defined as the discounted sum of dividends paid to shareholders plus the firm's value upon termination of the relationship. The firm's per-period output positively depends on (1) the executive's effort choice, which is binary, and (2) the match value between the firm and the executive, which is gradually learned as the executive's performance unfolds. Firms can commit to long-term contracts that specify compensation and the likelihood of termination based on the firm's evolving beliefs about match quality. This commitment enables the firm to internalize the externalities of future compensation rewards on the executive's activity choices in earlier periods.

Our model implies that the optimal compensation schedule depends on past performance only when both moral hazard and match value uncertainty are present. In such cases, the firm also employs involuntary termination as an incentive device, and the extent to which it is used depends on the executive's performance history. We present suggestive evidence consistent with these theoretical implications. Our reduced-form results show that executive compensation is positively correlated with both current and past performance, supporting the hypothesis that firms use performance-based pay to incentivize effort and increase compensation to retain executives with better match quality. Additionally, using previously established classifications of involuntary termination from the literature, we find that involuntary termination is also correlated with both current and past performance.

The key primitive objects of the model are the parameters governing the executive's preferences, including the nonpecuniary costs associated with each effort choice, risk aversion, and the relationship between output, effort, and match value. A central identification challenge arises from the presence of involuntary termination. In the model, a firm–executive relationship can end through two distinct channels. First, the relationship may be terminated involuntarily, with the termination probability chosen by the firm as part of the optimal

contract. Second, separation may occur due to an exogenous outside option faced by the executive, which is independent of the match value. When the cause of separation is not perfectly observed, standard identification arguments in the literature—such as those relying on inversion of the conditional choice probability to identify continuation values—cannot be directly applied in our setting.

We demonstrate that the key primitives of the model can be identified even when the econometrician does not observe whether a separation was voluntary or involuntary, by treating the firing decision as an equilibrium object to be identified. The main intuition behind our identification strategy is that the probability of involuntary termination is optimally chosen by the firm, generating an additional moment condition. One of the identification challenges is to determine when involuntary termination is actually used. To address this, we show that under the optimal contract, the model implies a change in the slope of the relationship between the probability of separation and current performance, conditional on past histories. This change point, which can be nonparametrically identified, represents the performance threshold beyond which the firm ceases to use involuntary termination because the executive's performance is sufficiently strong.

Another identification challenge arises from the presence of future informational rents that an executive acquires by choosing not to exert effort today. These informational rents occur whenever the firm adjusts future compensation schedules based on current performance and consistently rewards better outcomes. Conditional on the same level of performance, an executive who has previously shirked becomes more optimistic about the match value than the firm, and therefore receives a higher continuation value than an executive who has consistently exerted effort. As a result, two factors determine the extent to which the firm rewards high performance (via compensation) and penalizes low performance (via involuntary termination). The first is the likelihood ratio of output under shirking relative to exerting effort. The second, which arises from the interaction between moral hazard and match value uncertainty, is the ratio of the off-path continuation value (after shirking) to the on-path continuation value (under effort). We refer to this second term as the *ratchet ratio*, as it captures the executive's anticipation of an adjusted compensation schedule in response to today's performance—that is, the ratchet effect.

We show that the likelihood ratio and the ratchet effect can be separately identified when there are firm–executive pairs with sufficiently long tenure. Intuitively, after a sufficiently long tenure, the firm's belief about the executive stabilizes, and new output realizations have a limited impact on belief updating. As a result, the sensitivity of compensation to

performance primarily reflects the likelihood ratio, while the ratchet effect converges to one. Moreover, as tenure length increases, involuntary termination becomes unnecessary, and all separations are voluntary. Consequently, differences in the compensation schedule between the early and later years of tenure isolate the contribution of the ratchet effect.

Our estimates show that the per-year cost of moral hazard for firms ranges from \$620 million to \$1.13 billion, substantially exceeding the cost of effort for executives, which averages between \$0.93 million and \$4.2 million. These estimates are larger than those reported in the existing literature, primarily due to the presence of match value uncertainty, which limits the firm's ability to rely solely on compensation to incentivize effort. In particular, increasing performance-based pay also raises the executive's future informational rent from shirking, making it more difficult to induce effort through compensation alone. As a result, relative to a benchmark with moral hazard only, our model predicts that compensation is less sensitive to performance, and firms instead complement pay with involuntary termination to discipline behavior. We estimate that the average probability of firing across tenure years follows an inverse U-shape, with most terminations occurring after the first two years of the relationship. Our estimates also indicate that the magnitude of future informational rents from shirking is substantial.

1.1 Literature Review

Our paper contributes to the literature on job search (Jovanovic, 1979; Miller, 1984; Antonovics and Golan, 2012). Similar to the setting in Jovanovic (1979), our model features ex-ante homogeneous jobs, and both executives and firms learn about match-specific executive quality as performance is realized. Our model departs from the existing literature in two key dimensions. First, in our setting, an executive's private effort choice influences realized performance, making the learning process endogenous to the executive's hidden action. Second, while separation in standard models is often deterministic—occurring when performance falls below a threshold—our model incorporates two distinct channels for separation: voluntary and involuntary. Moreover, involuntary termination may occur probabilistically. This arises because firms may fire executives either (i) when they are clearly worse than a potential replacement (leading to deterministic separation) or (ii) to deter shirking by executives who are otherwise preferable to the outside option.

In our model, involuntary termination arises in equilibrium only when both moral hazard and match value uncertainty are present. To isolate this mechanism, we do not assume the presence of other contractual frictions that might make involuntary termination optimal in

an environment with moral hazard alone (but no match value uncertainty). In particular, we do not impose limited liability, in contrast to prior research showing that involuntary termination is necessary when executives are subject to limited liability (e.g., (Clementi and Hopenhayn, 2006; DeMarzo and Sannikov, 2006)). Similarly, we assume that executives have access to financial markets to smooth consumption, removing the need for the firm to provide insurance. In contrast, (Spear and Wang, 2005) shows that when executives cannot smooth consumption through borrowing or saving, involuntary termination may arise because providing insurance becomes too costly once the executive is sufficiently wealthy.

The interaction between moral hazard and match value uncertainty is a central theme in a large body of principal–agent literature with symmetric uncertainty and imperfect monitoring (Holmström, 1999; Prat and Jovanovic, 2014; DeMarzo and Sannikov, 2016; Cisternas, 2018; Bhaskar and Mailath, 2019; Lyman, 2023), where a key insight is that agents obtain informational rents when they deviate from the equilibrium action due to the ratchet effect. Our modeling approach differs from the existing literature in two main respects. First, we do not impose a parametric structure on the learning or production process. Instead, we assume that match value and output are affiliated, and that the marginal effect of (binary) effort on output may depend on the match value. Second, unlike many existing models that assume risk-neutral agents or allow firms to directly control consumption, our model incorporates risk-averse executives whose consumption decisions are not contractible. This modeling choice reflects both institutional realities and data limitations, as executive consumption is typically unobserved.

In addition, this paper complements the literature on quantifying executive turnover (Parrino (1997); Taylor (2010); Peters and Wagner (2014); Jenter and Kanaan (2015); 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 is not necessarily deterministic, and our identification approach enables us to determine the probability of involuntary termination independently of existing literature estimates. We demonstrate that these estimates are reasonable and positively correlate with prior classifications of involuntary termination found in the literature.

Our paper also contributes to the empirical literature on dynamic moral hazard models (Margiotta and Miller, 2000; Gayle and Miller, 2009, 2015). The key distinction between our work and existing studies is our focus on long-term contracts, as opposed to short-term contracts. In previous models, short-term contracts are considered optimal because the

optimal contract can always be constructed as a sequence of short-term contracts when the effort choice is a sufficient statistic for the distribution of firm output (Fudenberg et al., 1990; Malcomson and Spinnewyn, 1988; Rey and Salanie, 1990). However, this result does not hold in our case due to the interplay of moral hazard and match value uncertainty, which leads to the ratchet effect.

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.

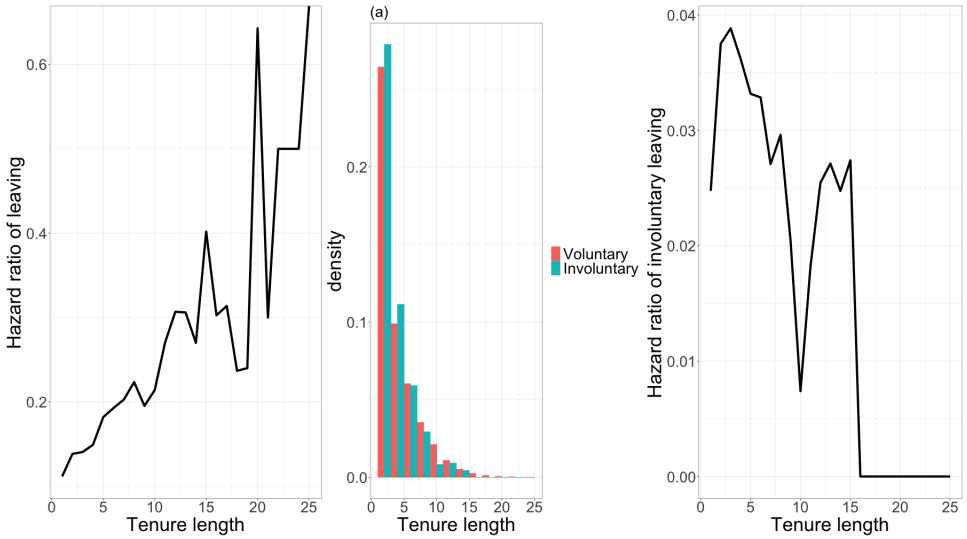
2 Setting and Data

We constructed our dataset using the Compustat Execucomp database, covering the period from 1992 to 2022. This database provides detailed information on executive compensation, including data for S&P 500 companies in 1992 and 1993, and for S&P 1500 companies from 1994 onward, which represent approximately 90% of all publicly traded companies in the US. The dataset also includes information on stock performance and dividend payout data for these firms. For our analysis, we focused exclusively on chief executive officers (CEOs) whose entire employment spells as executives were observed in the data. We exclude firms whose stock performance was in the bottom 2.5% and upper 2.5%. The final dataset comprises 34,719 observations at the executive-firm-year level.

Our analysis distinguishes between voluntary and involuntary termination. Voluntary termination occurs when an executive chooses to leave the firm in favor of better outside opportunities, whereas involuntary termination occurs when the firm unilaterally ends the relationship. Because involuntary termination is often not directly observed, we supplement our data with the executive turnover database provided by Peters and Wagner (2014) and Jenter and Kanaan (2015). This database uses an age-based classification method to determine whether a separation was voluntary or involuntary, identifying 1,398 separations as forced turnovers. We further complement this dataset with additional sources—including news articles, press reports, and SEC filings—and construct a more conservative measure of forced turnovers, which identifies 623 of these cases as involuntary. For each of these measures, we treat the classified instances as involuntary terminations, while acknowledging

that some unclassified separations may also be involuntary (i.e., type II errors). Accordingly, our econometric analysis treats separations not recorded as involuntary as potentially arising from either voluntary or involuntary termination. Figure 1 displays the distribution of tenure lengths at the time of involuntary termination and the hazard ratio between involuntary and voluntary termination using the classification of Peters and Wagner (2014) and Jenter and Kanaan (2015).

Figure 1: Distribution of tenure length at involuntary and voluntary termination



To calculate an executive's compensation, we adopt the approach outlined by Margiotta and Miller (2000) and Gayle et al. (2015), which incorporates the change in wealth from previously held options and stocks into the compensation measure. Consequently, an executive's total compensation consists 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 reflects the value of options and stocks held at the beginning of the period, multiplied by the firm's abnormal return. For our benchmark analysis, we define the abnormal return as the difference between the firm's stock return and the market portfolio's return, with the latter measured by a value-weighted market return.

Table 2 presents summary statistics on executive total compensation and its components in our dataset. Here, we categorize compensation into three components: (1) salary and bonus; (2) other reported payments, which include the value of restricted stock and options granted, long-term incentive payouts, and all other annual payments; and (3) changes in wealth from previous stock and options holdings. The latter two components are the primary

Table 2: Summary statistics

	Large firms		Small firms	
	Mean	SD	Mean	SD
Salary and bonus	842	1392	941	1799
Other reported payments	2149	6914	3078	6559
Change in wealth due to previous holdings	-186	11442	-260	11786
Total compensation	2918	13548	3871	14164
Gross return relative to market return	-0.06	0.21	-0.04	0.2
Executive age	52.87	7.1	52.56	6.95
Tenure duration	3.28	2.75	5.44	3.69
Proportion of internal executives	0.83	0.37	0.84	0.37
Proportion of female executives	0.03	0.16	0.03	0.16
Number of firms	3243		3525	
Number of executives	5880		7125	

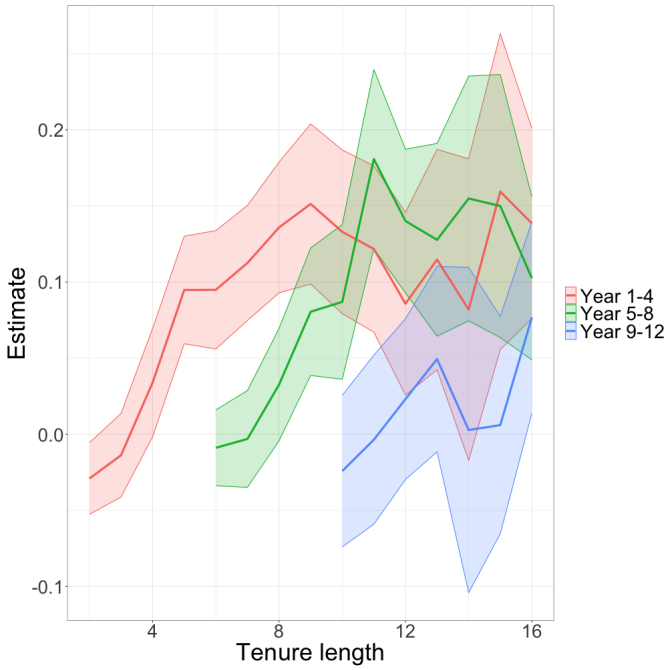
Note: Compensation measures are reported in thousands of dollars. Firm sizes are defined relative to the market size of the median firm in the sample.

sources of variation in total compensation in the data. Figure 11 in the Appendix shows that more than 30% of the total variation in compensation occurs within executives. The data also indicate that half of the executives leave their firms within the first three years of tenure. Executives at larger firms—defined as those with a market value above the median at the time of hiring—tend to have shorter tenures.

In Figure 3, we examine whether an executive's eventual tenure length is positively correlated with the firm's market value during the earlier years of their tenure. The x-axis of the figure represents the coefficient on a tenure length dummy used to predict the firm's market value relative to its initial market value, controlling for industry-year fixed effects. To mitigate survival bias, we estimate the regression on three separate subsamples: firms within the first four years of an executive's tenure (red line), within the second four years (green line), and during years 8 to 12 of tenure (blue line). For example, a firm whose executive eventually stays for five years experiences a market value increase of approximately 10% more than a firm that dismisses its executive within the first year. The regression estimates suggest that an executive's tenure length is positively correlated with firm performance up to a point. However, once tenure becomes sufficiently long, there are no statistically significant differences in firm performance across executives with longer tenures. For instance, among firms whose executives remain with the company for at least nine years, firm performance is no longer correlated with tenure length.

Table 4 shows that compensation is positively correlated with both current and past performance. In this regression, performance is proxied by the firm's gross value added

Figure 3: Differences in firm market value under executives with varying tenure lengths



Note: The y-axis represents the coefficient for each tenure length from the regression:

$$\log(\text{market value}_{j,t}) - \log(\text{market value}_{j,0}) = \alpha_T + \beta_{\text{industry,year}} + \epsilon_{jt},$$

where j denotes the firm index, t represents the tenure year, and T indicates tenure length. Three separate regressions are estimated using subsamples of firm market value: the first four years, the second four years, and years 8 to 12 of executive tenure. market value$_{j,0}$ refers to the firm's market value at the beginning of each corresponding four-year period. The shaded band represents the 95% confidence interval for the estimates, with standard errors adjusted for heteroskedasticity and clustered at the firm level.

in a given period, calculated as the sum of total dividend payments to shareholders and any compensation paid to the executive. This measure is normalized by dividing it by the firm's market value prior to the executive's appointment. The regression results indicate that compensation is higher when executive performance is stronger, and past performance also contributes positively. However, the influence of past performance on compensation depreciates over time.

Table 5 explores the relationship between executive separation, current performance, and past performance, as well as whether this relationship differs depending on whether the separation was likely involuntary. We use the same performance proxy as in Table 4. In columns (1) and (2), the dependent variable indicates whether any separation occurred. Columns (3) and (4) focus on involuntary terminations, using the classification from Peters

Table 4: Correlation between compensation and performance

	Total compensation (in thousands)		
	(1)	(2)	(3)
Performance in t	7.56	5.44	5.46
	(0.99)	(0.82)	(0.82)
Performance in $t-1$	4.99	1.53	1.55
	(0.75)	(0.7)	(0.7)
Internal	-0.06	-0.97	-0.88
	(0.02)	(0.45)	(0.44)
Executive demographics	Yes	Yes	Yes
Industry FEs	No	No	Yes
Year FEs	No	Yes	Yes
Observations	26611	26611	26611
Adjusted R^2	0.04	0.04	0.05

Note: Total compensation is measured in thousands of dollars. Performance is measured as gross value added divided by the firm's market value prior to the executive's appointment. Standard errors, adjusted for heteroskedasticity and clustered at the industry level, are reported in parentheses.

Table 5: Correlation between separation and performance

	All separation	Involuntary separation	Voluntary separation
	(1)	(2)	(3)
Performance in t	-0.5	-0.46	-0.04
	(0.32)	(0.13)	(0.37)
Performance in $t-1$	-0.27	-0.39	0.12
	(0.44)	(0.14)	(0.51)
Internal	-1.07	-0.25	-0.82
	(0.45)	(0.18)	(0.45)
Executive demographics	Yes	Yes	Yes
Industry FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Observations	26611	26611	26611
Adjusted R^2	0.04	0.01	0.05

Note: Separation is measured in percentage points. In this regression, we use the classification of involuntary separation from Peters and Wagner (2014) and Jenter and Kanaan (2015) and classify all other cases as voluntary separation. Standard errors, adjusted for heteroskedasticity and clustered at the industry level, are reported in parentheses.

and Wagner (2014) and Jenter and Kanaan (2015). Columns (5) and (6) apply our more conservative classification of involuntary terminations. The results show that lower current

performance is associated with a higher probability of separation, regardless of whether all separations or only involuntary ones are considered. Similarly, higher past performance is associated with a lower likelihood of involuntary separation, though its effect is smaller than that of current performance. The effects of past performance are more pronounced when the analysis is restricted to involuntary separations.

3 Model

In this section, we present a model of the contracting problem between a firm and its executive, detailing the optimal compensation structure and the schedule for involuntary termination.

A firm seeks to hire an executive from a pool of potential candidates. Once hired, let T denote the duration of employment, and let $t \in \{0, 1, 2, \ldots, T\}$ represent each year of tenure. While the executives are ex-ante identical to the firm, they possess heterogeneous match qualities that are specific to each firm-executive pairing. Let $\theta \in \mathbb{R}_+$ denote the match quality between the firm and an executive. The value of θ is unknown to both parties but is affiliated with the firm's per-period output, which is denoted by X. The firm and the executive have a common prior belief about θ, denoted as F_θ. We assume that for any θ in its support, there exists a sequence of the firm's gross market value realizations up to tenure year t that perfectly reveals θ.[1] Let $\bar{T}$ denote the tenure year at which θ is perfectly revealed. Thus, $\bar{T}$ is a random variable with a positive density over a finite range of values. For simplicity, we assume that the firm and all executives are infinitely lived. Throughout the paper, we use f_Z to denote the probability density function of a random variable Z and F_Z to denote its cumulative distribution function. We use bold symbols to denote vectors.

In each year of employment with the firm, the executive privately chooses an effort level $e \in \{0, 1\}$. We assume that the likelihood ratio of output, conditional on match quality, is separable in match quality and output as follows.

Assumption 1.

$$\frac{f_X(x \mid \theta, e = 0)}{f_X(x \mid \theta, e = 1)} = \gamma(x)\theta \tag{1}$$

where γ is a decreasing positive function.

[1] This assumption can be microfounded as follows. Suppose that θ has a finite support, such as $\theta \in \{L, H\}$. When $\theta = L$, the distribution of the firm's gross market value has a bounded support that overlaps with, but is not identical to, the support of its distribution when $\theta = H$.

An immediate implication of Assumption 1 is that e and x are affiliated. Consequently, an executive who has previously shirked is always more optimistic about their match value—in the sense of likelihood ratio dominance—than an executive who has not, conditional on the same output realizations. Moreover, the firm also holds a more favorable belief (again in likelihood ratio dominance) when the executive has exhibited better performance. Finally, because the likelihood ratio is separable in θ and x, the posterior belief about θ after observing a sequence of outputs $\boldsymbol{x}$ depends only on the number of times the executive chooses to work, not on the specific timing of those choices.

At the beginning of period 0, the executive receives a contract offer from the firm specifying compensations $w_{t+1} : \mathbb{R}^{t-1} \times \mathbb{R} \to \mathbb{R}$ for any given t. For example, $w_{t+1}(\boldsymbol{y}, x)$ denotes the compensation the executive receives at the beginning of period $t+1$ for current performance $X_t = x$ and past performance $\boldsymbol{y}$. Similarly, $\phi_{t+1} : \mathbb{R}^{t-1} \times \mathbb{R} \to \mathbb{R}$ specifies the probability that the executive is terminated involuntarily at the beginning of period $t+1$, conditional on realized performances $(\boldsymbol{y}, x)$. In the subsequent analysis, we omit the arguments of $w_{t+1}(\cdot)$ and $\phi_{t+1}(\cdot)$ whenever this does not cause confusion.

3.1 Executive preferences and consumption choices

The executive derives utility from consumption c_t. Let D_t indicate whether the executive is involuntarily terminated at the beginning of period t. Conditional on the firm not terminating the contract ($D_t = 0$), the executive can choose one of three discrete actions: (1) remain employed without exerting effort—that is, choose $e_t = 0$; (2) remain employed and exert effort ($e_t = 1$); or (3) leave the firm. Let $l_t = 1$ denote the executive's decision to leave the firm.

We normalize the executive's utility from leaving the firm to zero. The executive's lifetime utility is given by:

$$\mathbb{E}\left[\sum_{t=0}^{T-1} -\delta^t \exp(-\rho c_t)\left(((1-e_t)\alpha_0 + e_t\alpha_1)\exp(-\epsilon_t)(1-l_t) + l_t\right) + \sum_{T}^{\infty} -\delta^t \exp(-\rho c_t)\right]$$

where δ is the discount factor, ρ is the coefficient of absolute risk aversion, and T is the realized tenure duration. The term ϵ_t represents an idiosyncratic preference for working at the firm, privately observed by the executive. We assume $\epsilon_0 = 0$, and for $t > 0$, ϵ_t is independently and identically distributed following a logistic distribution. The parameters α_0 and α_1 capture the utility costs associated with choosing $e_t = 0$ and $e_t = 1$, respectively.

We assume that exerting effort is more costly than not exerting effort ($\alpha_1 > \alpha_0$).

The executive has access to financial markets and can smooth consumption over time; however, they cannot hedge against the volatility of w_{t+1} by purchasing a contingent plan in period t. It suffices to consider a finite set of financial securities to characterize the executive's optimal financial portfolio.[2] Specifically, let λ_t denote the price at time 0 of consumption in period t. Let b_t represent the price of a bond that pays a unit of consumption starting from period t, and let 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 t onward. [3] Both the executive and the firm have perfect foresight about a_t and b_t.

Let $v_t = \exp(-\rho w_t / b_t)$. Thus, v_t represents the annuitized compensation from working. Let ξ_t denote the executive's wealth in period t. Let $\mathbb{E}_t$ denote the expectation operator at the beginning of period t conditional on the output realizations and effort choices of past periods, prior to ϵ_t being realized. Therefore, $\mathbb{E}_t$ is related to the executive's belief about θ. Let $p_t := \mathbb{E}_t(\mathbf{1}_{l_t=0})$, which is the probability that the executive stays with the firm voluntarily at the beginning of period t. The executive's value function under the optimal consumption decisions is given by:

$$-\lambda_t b_t \exp\left(-(a_t + \rho\xi_t) b_t^{-1}\right) A_t,$$

where A_t is equal to

$$(1-p_t) + p_t \mathbb{E}\left[\exp(-\epsilon_t b_t^{-1}) \mid p_t\right] \min\left\{ \alpha_0^{b_t^{-1}} \mathbb{E}_t\left(v_{t+1}\left[\phi_{t+1} + (1-\phi_{t+1}) A_{t+1}\right] \mid e_t = 1\right)^{1-b_t^{-1}}, \right.$$
$$\left. \alpha_1^{b_t^{-1}} \mathbb{E}_t\left(v_{t+1}\left[\phi_{t+1} + (1-\phi_{t+1}) A_{t+1}\right] \mid e_t = 0\right)^{1-b_t^{-1}} \right\}.$$

Intuitively, the first component of A_t represents the (normalized) continuation value of not being employed by the firm, while the second component captures the continuation value of being employed by the firm at the optimal effort level. Henceforth, we refer to A_t as the continuation value in period t. This continuation value depends on (1) the executive's belief about their match value θ, which is endogenously determined by their private past effort choices, and (2) the structure of compensation for future performance.

[2] The proof of this result and subsequent derivation of the optimal consumption decisions are similar to that of the existing literature (Margiotta and Miller, 2000; Gayle et al., 2015) and is therefore omitted.

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

3.2 Contracting problem

Incentive constraints We assume that the firm aims to implement a contract that incentivizes the executive to exert effort in all periods prior to contract termination at T. Let $\mathbb{E}_t^*$ denote the expectation operator at the beginning of period t, based on the firm's belief about θ—that is, conditional only on past output realizations under the conjecture that the executive has exerted effort in all previous periods. Therefore, $\mathbb{E}_t^*$ differs from $\mathbb{E}_t$ when the latter reflects a history involving past shirking. Since the effort decision is private and not directly contractible, the firm's contract must satisfy an incentive compatibility (IC) constraint. For any t, the IC constraint is given by:

$$\begin{aligned} &\alpha_1^{(b_t-1)^{-1}} \mathbb{E}_t^* \left(v_{t+1} \left[\phi_{t+1} + (1-\phi_{t+1}) A_{t+1} \right] \mid e_t = 1 \right) \\ \leq\ &\alpha_0^{(b_t-1)^{-1}} \mathbb{E}_t^* \left(v_{t+1} \left[\phi_{t+1} + (1-\phi_{t+1}) A_{t+1} \right] \mid e_t = 0 \right) \end{aligned} \tag{IC}$$

Participation constraints In period 0, since $\epsilon_0 = 0$, the firm sets the contract such that the executive is indifferent between accepting and not accepting the contract (and thus always chooses to accept). Therefore, the participation constraint for period 0 is specified as follows:

$$\alpha_1^{(b_0-1)^{-1}} \mathbb{E}_0^* \left(v_1 \left[\phi_1 + (1-\phi_1) A_1 \right] \mid e_0 = 1 \right) \leq 1 \tag{IR_0}$$

where the right-hand side represents the normalized value of the executive's outside option, which is equal to 1.

In subsequent periods $t > 0$, because ϵ_t is unobserved by the firm and unbounded, the participation constraint identifies the expected utility an executive receives at the beginning of period t. Equivalently, this can be formulated as follows: the firm chooses to retain the executive (voluntarily) with probability $p_t : \mathbb{R}^{t-2} \times \mathbb{R} \to [0,1]$. For example, conditional on performance in period $t-1$, $X_{t-1} = x$, and past output realizations $\boldsymbol{y}$, the probability that the executive chooses to stay with the firm prior to the realization of ϵ_t is given by $p_t(\boldsymbol{y}, x)$.

Let ϵ_t^* denote the threshold at which the executive is indifferent between staying with the firm and quitting at the beginning of period t:

$$\exp(-\epsilon_t^* b_t^{-1}) \alpha_1^{b_t^{-1}} \mathbb{E}_t^* \left(v_{t+1} \left[\phi_{t+1} + (1-\phi_{t+1}) A_{t+1} \right] \mid e_t = 1 \right)^{1-b_t^{-1}} = 1.$$

Therefore, $p_t = \Pr(\epsilon_t \geq \epsilon_t^*)$. Define $\underline{u}_t(p) := \exp\left(F_\epsilon^{-1}(1-p)(b_t-1)^{-1} \right)$. Therefore, $\underline{u}_t(p)$ denotes the (normalized) utility level from compensation at which an executive accepts the

contract with probability p, and $\underline{u}_t$ is decreasing in p. The participation constraint for $t \geq 1$ is then given by:

$$\alpha_1^{(b_t-1)^{-1}} \mathbb{E}_t \left(v_{t+1} \left[\phi_{t+1} + (1-\phi_{t+1}) A_{t+1} \right] \mid e_t = 1 \right) \leq \underline{u}_t(p_t). \tag{IR}$$

Firm's problem Let $\Pi_{0,t}$ denote the market value of the firm if the executive is replaced at the beginning of period t. $\Pi_{0,t}$ is independent of the firm's belief about the replaced executive's quality. Because there is a 1-1 mapping between w_t and v_t, we formulate the firm optimization problem as follows:

$$\Pi_0 = \max_{(p_t(\cdot), v_{t+1}(\cdot), \phi_{t+1}(\cdot))} \mathbb{E} \left[\lambda_0 (x_0 - w_1) + \lambda_1 \Pi_1 \right] \tag{$\mathcal{P}$}$$

where

$$\Pi_t = \lambda_t \mathbb{E}_t \left[p_t(1-\phi_t) \left[x_t - w_{t+1} + \lambda_{t+1}/\lambda_t \Pi_{t+1} \right] + (1 - p_t(1-\phi_t)) \Pi_{0,t} \right]$$

subject to (IR_0), (IR), and (IC). That is, the firm is maximizing the shareholders' expected payment, defined as the discounted sum of per-period dividends weighted by the price of consumption in period t, along with the firm's market value in the event of an executive replacement.

3.3 Necessary conditions

Let $\boldsymbol{y} \in \mathbb{R}^{t-1}$ denote a sequence of realized outputs. For $\tau \leq t$, define:

$$g_t^\tau(\boldsymbol{y}, x) := \frac{\mathbb{E}_t(f_X(x \mid e_{t',t' \leq \tau} = 1, e_\tau = 0, \boldsymbol{y})}{\mathbb{E}_t^*(f_X(x \mid e_t = 1, \boldsymbol{y}))} \tag{2}$$

Therefore, g_t^τ represents the likelihood ratio between the distribution of output in period t, conditional on past outputs $\boldsymbol{y}$, when the executive shirks for the first time in period τ, and the distribution when the executive has never shirked. When the executive shirks for the first time in period τ, they subsequently choose the optimal effort level in each following period, based on their private belief. This optimal effort choice for $\tau' > \tau$ may involve further shirking. In particular, g_t^t corresponds to the likelihood ratio commonly considered in the static setting, where the executive faces an immediate trade-off between lower effort costs from shirking and a worse distribution of performance in the same period. However, for $\tau < t$, g_t^τ reflects how the distribution of output differs when the executive's belief about

θ deviates from that of the firm due to prior private shirking. Subsequently, we use $\mathbb{E}_t^\tau$ to denote the expectation operator corresponding to the executive's belief at the beginning of period t, after shirking for the first time in period τ. Because of the affiliation property, $g_t^\tau(\boldsymbol{y}, x)$ is decreasing in x.

Let A_t^* denote the continuation value of the executive on the equilibrium path, i.e., when the executive's belief about θ is similar to that of the firm. Let A_t^τ denote the continuation value associated with $\mathbb{E}_t^\tau$—i.e., after an executive shirks for the first time in τ. Let μ_t^{IC} and μ_t^{IR} denote the Lagrangian multipliers associated with the IC and IR constraints in period t. Similarly, let $\mu_t^{q,0}$ and $\mu_t^{q,1}$ represent the Lagrangian multipliers associated with the constraints $\phi_t \geq 0$ and $\phi_t \leq 1$.

In Sections 3.3.1 and 3.3.2, we illustrate the key determinants of $(v_{t+1}, \phi_{t+1}, p_{t+1})$ for the initial two periods ($t = 0$ and $t = 1$). Similar sets of necessary conditions applicable to periods $t \geq 2$ are omitted for brevity.

3.3.1 Necessary conditions for period-1 compensation and termination

The executive receives a payment of w_1 at the beginning of period 1 for his performance in period 0 and faces a termination probability of ϕ_1.

The FOC with respect to $v_1(\cdot)$ at $X_0 = x$ is expressed as follows:[4]

$$\begin{aligned} 0 = \lambda_0 v_1^{-1} - (\mu_0^{IC} + \mu_0^{IR})\alpha_1^{(b_0-1)^{-1}} \left[(A_1^*(1-\phi_1) + \phi_1\right] \\ + \mu_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_0^0 \left[(A_1^0(1-\phi_1) + \phi_1\right] \end{aligned} \tag{3}$$

The first term in Equation (3) represents the effect of increasing v_1—that is, lowering w_1—on the firm's payoff, while the remaining terms capture its impact on the executive's IC and IR constraints in period 0. Furthermore, Equation (3) indicates that three factors determine how v_1 is set. First, the likelihood ratio g_0^0, which is equal to $\gamma(x_0)\mathbb{E}(\theta \mid x_0)$. A higher likelihood ratio is associated with a higher v_1, and equivalently, a lower w_1. Second, the future continuation value under the firm's belief, denoted by A_1^*, which is dictated by the structure of future compensation plans. Third, the continuation value A_1^0 when the executive chooses to shirk, which depends on both future compensation plans and the extent to which the executive's belief deviates from the firm's belief. When compensation is always positively

[4]For simplicity, we have assumed that point-wise optimization is always possible, and there exists a strictly positive value of v_1, which implies a finite value of w_1 that satisfies the FOC with respect to ϕ_1. When this is not the case, the optimization problem requires the firm to choose a lower bound on wages, $\underline{w} > -\infty$, in addition to the current contractual tools.

correlated with θ, A_1^0 is strictly smaller than A_1^*—that is, the executive anticipates a positive informational rent from shirking today due to adjustments in future compensation based on revealed information about their match quality. Conversely, if future contracts reward poor performance, then $A_1^0 > A_1^*$. When future compensation is constant, $A_1^0 = A_1^*$. The relative difference between A_1^0 and A_1^* is a form of ratchet effect, and we henceforth refer to the ratio $\frac{A_1^0}{A_1^*}$ as the *ratchet ratio*. A smaller ratchet ratio implies that the executive can earn greater informational rents in future discounted value by shirking today.

Next, the FOC w.r.t $p_1(\cdot)$ at $X_0 = x$ is given by:

$$0 = \lambda_1(1-\phi_1)\mathbb{E}_1\left[(x_1 - w_2 + \lambda_2/\lambda_1\Pi_2) - \Pi_{0,1}\right] + \mu_1^{IR}\underline{u}'(p_1) \tag{4}$$

The first term of Equation (4) represents the change in the firm's value from replacing the executive. Because ϵ is unbounded and the firm is not constrained by limited liability, $p_1\lambda_1(1-\phi_1)\mathbb{E}_1\left[(x_1 - w_2 + \lambda_2/\lambda_1\Pi_2) - \Pi_{0,1}\right]$ approaches 0 from the right as p_1 approaches 0. That is, there always exists a value of p_1 such that the change in the firm's value is positive. The second term reflects the impact of adjusting p_1 on the IR constraint in period 1. Specifically, an increase in p_1 makes it more costly for the firm to satisfy the IR constraint in that period. Because $\underline{u}'(p_1)$ is negative, the firm must ensure—through the design of future compensation—that the expected value from retaining the executive with probability p_1 exceeds the value from hiring a replacement.

Finally, the FOC w.r.t $\phi_1(\cdot)$ at $X_0 = x$ is given by:

$$\begin{aligned} 0 = \lambda_1 p_1 \mathbb{E}_1\left[\Pi_{1,0} - (x_1 - w_2 + \lambda_2/\lambda_1\Pi_2)\right] - (\mu_0^{IC} + \mu_0^{IR})\alpha_1^{(b_0-1)^{-1}}(1-A_1^*)v_1 \\ +\mu_0^{IC} g_0^0 \alpha_0^{(b_0-1)^{-1}}(1-A_1^0)v_1 - \mu_1^{q,1} + \mu_1^{q,0} \end{aligned} \tag{5}$$

The first term of Equation (5) represents the change in the firm's expected market value in period 1 if the executive is replaced. As discussed above, this term is always weakly negative. The remaining terms capture the impact of increasing the probability of involuntary termination on the executive's IR and IC constraints. When future compensation rewards better performance, implying $A_1^0 < A_1$, an increase in ϕ_1 reduces the informational rent from shirking and thus strengthens the executive's incentives to exert effort. However, when there is no ratchet effect ($A_1^0 = A_1$), this benefit is no longer present. In that case, Equation (3) implies that $\mu_1^{q,0} > 0$, and $\phi_1 = 0$ is always an equilibrium.

3.3.2 Necessary conditions for period-2 compensation and termination

The FOC w.r.t $v_2(\cdot)$ at $X_0 = y, X_1 = x$ is expressed as follows:

$$\begin{aligned} 0 = \lambda_1 v_2^{-1} - \left(\mu_1^{IR} + \lambda_1^{IC}\right) \alpha_1^{(b_1-1)^{-1}} \left[(A_2^*(1-\phi_2) + \phi_2\right] \\ + \mu_1^{IC} \alpha_0^{(b_1-1)^{-1}} g_1^1 \left[(A_2^1(1-\phi_2) + \phi_2\right] \\ + (1-\phi_1)\left[- \left(\mu_0^{IR} + \mu_0^{IC}\right) \alpha_1^{(b_0-1)^{-1}} v_1 \frac{\partial A_1^*}{\partial v_2} + \mu_0^{IC} \alpha_0^{(b_0-1)^{-1}} v_1 g_1^0 \frac{\partial A_1^0}{\partial v_2}\right] \end{aligned} \tag{6}$$

where, at a given belief

$$\frac{\partial}{\partial v_2} A_1 = \frac{A_1 - (1-p_1)}{\mathbb{E}\left[v_2((1-\phi_2)A_2 + \phi_2)\right]} (1 - b_1^{-1}) \left[(1-\phi_2)A_2 + \phi_2\right] \tag{7}$$

Relative to the FOC with respect to v_1 in Equation (3), the trade-off involved in increasing v_2—that is, decreasing w_2—includes additional effects on the incentive and participation constraints in period 1. This difference reflects the firm's ability to internalize the externalities that its period-1 compensation schedule imposes on the constraints in period 0, by committing to a long-term contract. These externalities are governed by the likelihood ratio g_1^0, the continuation value from shirking in period 0, A_2^0, and the continuation value from shirking in period 1, A_2^1. These off-equilibrium-path continuation values are, in turn, shaped by the divergence between the firm's belief and the executive's belief, the executive's subsequent effort choices following past shirking, and the structure of compensation from period 3 onward.

Similarly, the FOC w.r.t ϕ_2 is given by

$$\begin{aligned} 0 = \lambda_2 p_2 \mathbb{E}_2^* \left[\Pi_{0,2} - (x_2 - w_3 + \lambda_3/\lambda_2 \Pi_3)\right] - (\mu_1^{IC} + \mu_1^{IR})(1 - A_2^*) v_2 \\ + \mu_1^{IC} g_1^1 (1 - A_2^1) v_2 - \mu_2^{q,1} + \mu_2^{q,0} \\ + (1-\phi_1)\left[-(\mu_0^{IC} + \mu_0^{IR}) v_1 \frac{\partial A_1^*}{\partial \phi_2} + \mu_0^{IC} v_1 g_1^0 \frac{\partial A_1^0}{\partial \phi_2}\right] \end{aligned} \tag{8}$$

where

$$\frac{\partial}{\partial \phi_2} A_1 = \frac{A_1 - (1-p_1)}{\mathbb{E}\left[v_2((1-\phi_2)A_2 + \phi_2)\right]} (1 - b_1^{-1}) \left[1 - A_2\right] v_2 \tag{9}$$

3.4 Characterization of the optimal contract

We first characterize the optimal compensation and firing decisions when θ has been perfectly learned.

Proposition 1 (**post-$\bar{T}$ contracts**). *For $t > \bar{T}$:*

- ***(P-1)*** *There is no firing ($\phi_t = 0$) and w_{t+1} is increasing in x_t.*
- ***(P-2)*** *p_t and Π_t are increasing in θ.*
- ***(P-3)*** *Conditional on θ, $\frac{\partial p_t}{\partial x_\tau} = \frac{\partial w_{t+1}}{\partial x_\tau} = 0$ for $\bar{T} \leq \tau < t$.*

The proof of Proposition 1 is in the Appendix. Because θ is perfectly learned for $t \geq \bar{T}$, we show that it is without loss for the firm to ignore the the externality of compensations in period t on previous period's IR and IC constraints. Therefore, compensations in each period mirror those of one-period contracts, where the sensitivity of compensation to performance is determined only by the likelihood ratio g_t^t. Because the belief on θ degenerates, $g_t^t(x_t)$ is proportional to $\gamma(x_t)$ and is decreasing in x_t (Assumption 1). Therefore, compensations are strictly increasing in performance.

The result that firing never occurs when θ is perfectly learned relies on the observation that, in this case, there is no ratchet effect. As a result, setting $\phi > 0$ does not contribute to incentivizing the executive to exert effort. Moreover, since firing imposes an additional cost on the executive's ex-ante participation constraint compared to simply offering low compensation, the firm is strictly better off allowing the executive to exit voluntarily by offering a low expected wage, rather than firing them with positive probability.

The proof of Proposition 1 also illustrates the relationship between the executive's expected wages and their marginal productivity. Because the match value is firm-specific, the executive is never paid their full marginal product. However, we show that an executive with a higher θ earns higher expected utility. Moreover, the firm's market value from hiring a higher-θ executive increases with θ.

Next, we consider the contracts set prior to $\bar{T}$. We define a contract as *monotonic* if compensation is weakly increasing in all past outputs, and the probability of termination is weakly decreasing in all past outputs. The following proposition shows that there exists a monotonic contract that satisfies the first-order conditions.

Proposition 2. *There exists a monotonic contract that satisfies the necessary conditions of ($\mathcal{P}$).*

In the proof, we show that there exist weakly positive values of $(\mu_t^{IR}, \mu_t^{IC}, \mu_t^{q,0}, \mu_t^{q,1})$ such that the FOCs are satisfied when the contract is monotonic. Intuitively, this monotonicity result follows from the affiliation property between θ and output x. Affiliation implies that the ratchet ratio under a monotonic contract is decreasing in output—that is, the informational rent from past deviation is positively correlated with output. In contrast, a non-monotonic contract that allocates higher informational rents from past shirking to low output levels makes it more difficult to incentivize the executive not to shirk. We further show that, under a monotonic contract, the marginal benefit of using termination to deter shirking also decreases with output, implying that a monotonic firing policy is locally optimal.

In addition, under a monotonic contract, we can show that the optimal deviation involves shirking only once. This is summarized in the following result.

Corollary 1. *Under a monotonic incentive-compatible contract, an executive who has previously shirked will choose to work in all subsequent periods.*

The following proposition establishes a property of the firing policy ϕ.

Proposition 3 (**Pre-$\bar{T}$ Contracts**)**.** *Suppose that the optimal contract is monotonic. Let $\boldsymbol{y}$ denote the history of outputs up to period $t-1$. If*

$$\min_{x \in supp(X_t)} \left\{ \mathbb{E}^*_{t+1} \left(x_{t+1} + \frac{\lambda_{t+2}}{\lambda_{t+1}} \Pi_{t+2} \right) - \Pi_{0,t+1} \right\}$$

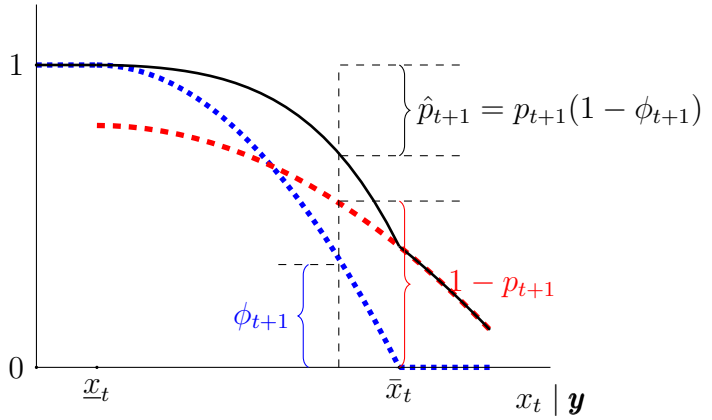
is sufficiently small, then there exist functions $\bar{x}_t : \mathbb{R}^{t-1} \to \mathbb{R}$ and $\underline{x}_t : \mathbb{R}^{t-1} \to \mathbb{R}$ such that $\bar{x}_t(\boldsymbol{y}) > \underline{x}_t(\boldsymbol{y})$ for all $\boldsymbol{y}$, and:

$$\phi_{t+1}(\boldsymbol{y}, x) = \begin{cases} 0 & \text{if } x \geq \bar{x}_t(\boldsymbol{y}), \\ 1 & \text{if } x \leq \underline{x}_t(\boldsymbol{y}), \\ \in (0,1) & \text{if } \underline{x}_t(\boldsymbol{y}) < x < \bar{x}_t(\boldsymbol{y}), \text{ with } \lim_{x \uparrow \bar{x}_t(\boldsymbol{y})} \phi_{t+1}(\boldsymbol{y}, x) > 0. \end{cases}$$

We prove Proposition 3 by analyzing the economic trade-off the firm faces when using firing to incentivize the executive. Increasing ϕ_t makes it easier to induce effort in period $t-1$, but comes at the cost of potentially losing the executive to an outside replacement—who may be of lower match value—in the following period. When the firm retains the executive with some probability, the presence of the ratchet effect implies that the firm needs to reward more for performance in the preceding period, which comes at a cost as the executive is risk averse. We show that when the firm's belief about an executive is sufficiently pessimistic, the

firm's maximum expected profit from retention is strictly lower than the cost of having to use higher-power rewards. Therfore, when X_t is sufficiently low, $\phi_{t+1} = 1$. On the other hand, when an executive achieves high performance, the likelihood ratio is sufficiently small, and implementing a positive firing probability undermines the compensation scheme designed to reward performance. Therefore, there is no benefit to using a positive firing probability, and $\phi_{t+1} = 0$. Finally, we show that $v(\cdot)$ and $\phi(\cdot)$ are continuous in their arguments, which implies that $\underline{x}_t < \bar{x}_t$. . Figure 6 illustrates the shape of the aggregate retention probability, accounting for the firing policy q.

Figure 6: Illustration of the separation probability in equilibrium



Note: The dashed red curve represents the quitting probability, $1 - p_{t+1}$. This probability is undefined when the executive is fired with certainty ($X_t \leq \underline{x}_t(\boldsymbol{y})$). The dashed blue curve represents the firing probability, ϕ_{t+1}, which equals 1 for low outcomes ($X_t \leq \underline{x}_t(\boldsymbol{y})$) and 0 when the executive's performance is sufficiently good ($X_t \geq \bar{x}_t(\boldsymbol{y})$). The solid black line represents the separation probability, given by $1 - p_{t+1}(1 - \phi_{t+1})$.

Discussion In our model, the firm can use both rewards and punishments to incentivize the executive. Rewards are governed by p, which measures the utility level promised to the executive on the equilibrium path. Punishments are captured by the firing decisions ϕ. In the absence of match value uncertainty (but with moral hazard), the ratchet ratio is always equal to 1, and the marginal cost of increasing ϕ always outweighs the marginal cost of lowering p. This is because when $\phi > 0$, the executive is forced to leave the firm regardless of their private preference for staying, which may be positive. Therefore, setting $\phi > 0$ is optimal only when there is some match value uncertainty interacting with moral hazard. This trade-off is also reflected in the differences between contracts before and after θ is perfectly realized.

A similar intuition arises when there is match value uncertainty but no moral hazard. In this case, the ratchet ratio is again equal to 1, and using $\phi > 0$ is strictly suboptimal.

Finally, we note that because ϕ is used to discipline the executive against moral hazard, the firm must possess *some* level of commitment. For example, if ϕ can be renegotiated after output is realized, the firm is always better off ex post by offering very low expected compensation in the future to induce the executive to leave voluntarily, rather than firing them. However, as long as the firm can commit to setting ϕ *before* the executive chooses their effort level, the main intuition of the model continues to hold.

4 Identification

In this section, we establish the identification argument under the assumption that firms and executives are ex-ante homogeneous and the contract is monotonic. In Section **??**, we explain how heterogeneity in the observed characteristics of firms and executives is incorporated into the estimation procedure. We also allow $f_X(\cdot \mid e_t = 1)$ to vary with t. Accordingly, we use the notations $f_{X_t}(\cdot \mid e_t = 1)$ in the analysis that follows.

We focus on the following objects of interest. First, the set of parameters that characterize the executive's preferences: $\{\rho, \alpha_0, \alpha_1\}$. Second, the prior belief about θ and the way match quality and effort jointly affect output, summarized by f_θ, $\gamma(x)$, and $f_{X_t}(\theta \mid e_t = 1)$. Third, we are also interested in the probability of forced termination, ϕ_t, and the ratchet ratio $\frac{A_t^\tau}{A_t^*}$ for any $\tau < t$. Although ϕ_t and $\frac{A_t^\tau}{A_t^*}$ are not primitives as they depend on the contract set by the firm, they measure the welfare loss from asymmetric information at the optimal contract.

From the data, we observe realizations of compensation (w_{t+1}), output x_t (measured as gross value added), and the firm's market value after x_t is realized but before ϵ_{t+1} is realized (conditional on $D_t \neq 1$). We also observe the prices of financial securities (a_t, b_t, λ_t).

Using the theoretical results from Section 3, we can further reduce the number of objects that need to be identified. In particular, Corollary 1 implies that the continuation value when an executive has previously shirked is independent of *when* the shirking occurred and depends only on the history of observed outputs.

Corollary 2. *Let $\tau < t$. Under a monotonic contract, conditional on the output history $\boldsymbol{y} \in \mathbb{R}^{t-1}$, we have*

$$A_t^\tau(\boldsymbol{y}) = A_t^{t-1}(\boldsymbol{y}),$$

where $A_t^{t-1}(\cdot)$ is a decreasing function. Moreover,

$$\frac{g_t^t(x, \boldsymbol{y})}{g_t^\tau(x, \boldsymbol{y})} = \frac{\gamma(x)}{\gamma(y_\tau)},$$

where y_τ is the τ-th element of $\boldsymbol{y}$.

Recall that Proposition 3 identifies $\bar{x}_t(\boldsymbol{y})$ as the kink in the graph of the separation probability as a function of x_t, conditional on $\boldsymbol{y}$. In addition, $\underline{x}_t(\boldsymbol{y}) = \arg\min_x \Pr\left(l_{t+1}(1 - D_{t+1}) + D_{t+1} = 1 \mid x, \boldsymbol{y}\right) = 1$. Therefore, both $\bar{x}_t(\boldsymbol{y})$ and $\underline{x}_t(\boldsymbol{y})$ are identified.

Identification of $\bar{T}$ We now show that $\bar{T}$ is an identified function of the history of output. From Proposition 1, once θ has been realized, compensation at time t depends on $(x_0, x_1, \ldots, x_{\bar{T}-1})$ but is independent of x_{t-1} for all $t \geq \bar{T} + 1$. Therefore, $\bar{T}$ is identified as follows

$$\bar{T} = \min\{\arg\min_t \sum_w (f_{W_t}(w \mid x_0, x_1, ..., x_{\bar{T}, ..., x_{t-1}}) - f_{W_t}(w \mid x_0, x_1, ..., x_{\bar{T}}))^2\}$$

Identification of α_1 and ρ Because $\bar{T}$ is identified, we use Proposition 1 to identify the risk aversion coefficient ρ and the nonpecuniary cost of working α_1. Consider two output histories (up to $\bar{T}$), $\boldsymbol{y}$ and $\boldsymbol{y}'$, that yield different distributions of w_{t+1}. Proposition 1 states that $w_{t+1}(\cdot, \boldsymbol{y})$ and $w_{t+1}(\cdot, \boldsymbol{y}')$ can always be ordered via likelihood ratio dominance, where $w_{t+1}(\cdot, \boldsymbol{y}')$ dominates $w_{t+1}(\cdot, \boldsymbol{y})$ if and only if $\mathbb{E}_t(\theta \mid \boldsymbol{y}') > \mathbb{E}_t(\theta \mid \boldsymbol{y})$. Because compensation in period $t+1$ does not depend on output in period t, we have $p_{t+1}(\cdot, \boldsymbol{y}) = p_{t+1}(\boldsymbol{y})$. Moreover, since $q_{t+1}(\cdot, \boldsymbol{y}) = 0$ and $q_{t+1}(\cdot, \boldsymbol{y}') = 0$, both $p_{t+1}(\boldsymbol{y})$ and $p_{t+1}(\boldsymbol{y}')$ are identified using only data on separation.

Therefore, ρ and α_1 must rationalize:

$$\underline{u}(p_t(\boldsymbol{y})) = \alpha_1^{(b_t - 1)^{-1}} \mathbb{E}^*(\exp(-\rho w_{t+1}/b_{t+1} \mid \boldsymbol{y}) A_{t+1}^*(\boldsymbol{y})) \tag{10}$$

$$\underline{u}(p_t(\boldsymbol{y}')) = \alpha_1^{(b_t - 1)^{-1}} \mathbb{E}^*(\exp(-\rho w_{t+1}/b_{t+1} \mid \boldsymbol{y}') A_{t+1}^*(\boldsymbol{y}')) \tag{11}$$

where $A_{t+1}^*(\cdot)$ is an identified mapping of $p_{t+1}(\cdot)$ and is therefore identified. WLOG, consider $(\theta \mid \boldsymbol{y}) < (\theta \mid \boldsymbol{y}')$. Equations (10) and (11) each provide an identified, upward-sloping relationship between ρ and α_1. We show that these two functions intersect at most once, using the facts that the distribution of $w_{t+1}(\boldsymbol{y})$ is dominated in the likelihood ratio sense by the distribution of $w_{t+1}(\boldsymbol{y}')$, and that $A_{t+1}^*(\boldsymbol{y}) > A_{t+1}^*(\boldsymbol{y}')$. Therefore, α_1 and ρ are identified.

We prove this result in Lemma 6 in the Appendix.

Identification of $\gamma(x)$ As shown in Proposition 1, the post-$\bar{T}$ contracts resemble static contracts. Therefore, conditional on the same output history $\boldsymbol{y}$ (up to $\bar{T}$), the annuitized compensation v_{t+1} is a linear function of $\gamma(x_t)$:

$$v_{t+1}(\boldsymbol{y}, x_t) = \beta_0(\boldsymbol{y}) - \beta_1(\boldsymbol{y})\gamma(x_t).$$

for some unknown functions $\beta_0(\cdot)$ and $\beta_1(\cdot)$. This implies

$$\frac{\frac{\partial^2}{\partial x_t \partial \boldsymbol{y}} v_{t+1}(\boldsymbol{y}, x_t)}{\frac{\partial}{\partial \boldsymbol{y}} v_{t+1}(\boldsymbol{y}, x_t)} = \frac{\partial}{\partial x_t}\gamma(x_t)\left(\gamma(x_t)\right)^{-1}$$

Because x_t and $\boldsymbol{y}$ are both observed, $\gamma(\cdot)$ is identified up to a multiplicative constant. We further pin down this constant using the property of the likelihood ratio that $\mathbb{E}(\gamma(x_t) \mid \boldsymbol{y}) = 1$.

Identification of α_0 Conditional on the past output history (up to $\bar{T}$) being $\boldsymbol{y}$, the IC constraint in period $t+1$ holds with equality:

$$\begin{aligned} &\alpha_1^{(b_t-1)^{-1}}\, \mathbb{E}_t^* \left[\exp\left(-\frac{\rho w_{t+1}}{b_{t+1}}\right) A_{t+1}^*(\boldsymbol{y}) \,\middle|\, \boldsymbol{y} \right] \\ = \alpha_0^{(b_t-1)^{-1}}\, &\mathbb{E}_t^* \left[\exp\left(-\frac{\rho w_{t+1}}{b_{t+1}}\right) A_{t+1}^*(\boldsymbol{y})\gamma(x_t)\theta \,\middle|\, \boldsymbol{y} \right]. \end{aligned} \tag{12}$$

Since the posterior on $\theta \mid \boldsymbol{y}$ degenerates, $\mathbb{E}_t^*\left[\gamma(x)\theta \mid \boldsymbol{y}\right] = 1$. Therefore

$$\begin{aligned} &\alpha_1^{(b_t-1)^{-1}}\, \mathbb{E}_t^* \left[\exp\left(-\frac{\rho w_{t+1}}{b_{t+1}}\right) A_{t+1}^*(\boldsymbol{y}) \,\middle|\, \boldsymbol{y} \right] \\ = \alpha_0^{(b_t-1)^{-1}}\, &\mathbb{E}_t^* \left[\exp\left(-\frac{\rho w_{t+1}}{b_{t+1}}\right) A_{t+1}^*(\boldsymbol{y}) \frac{\gamma(x_t)}{\mathbb{E}_t^*(\gamma(x) \mid \boldsymbol{y})} \,\middle|\, \boldsymbol{y} \right]. \end{aligned} \tag{13}$$

Therefore, as $\rho, \gamma(\cdot)$, and α_1 have all been identified, α_0 is identified from Equation (13).

Identification of the ratchet ratio and the firing probability Below, we outline the identification argument for the period-1 ratchet ratio and firing probability under the assumption that the optimal contract is monotonic.

Let $p_1^0(x)$ denote the probability that an executive with belief $\mathbb{E}_1^0$—i.e., one who previously shirked in period 0 after observing $X_0 = x$—chooses to stay with the firm prior to realizing

ϵ_1. Under a monotonic contract, we have $p_1^0(\cdot) > p_1(\cdot)$. We consider three separate cases. In the first case, suppose $X_0 = x \geq \bar{x}_0$, i.e., $\phi_1 = 0$. In this case, both $p_1(\cdot)$ and $A_1^*(\cdot)$ are identified. For notational simplicity, we omit the argument of output when it is clear from context. The first-order conditions with respect to v_1, along with the individual rationality constraint, yield:

$$\mu_0^{IR} = \lambda_0 \tag{14}$$

$$A_1^0 g_0^0 = \frac{(\mu_0^{IC} + \lambda_0)\alpha_1^{(b_0-1)^{-1}} A_1^* - \lambda_0 v_1^{-1}}{\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}}} \tag{15}$$

Because there is no private information in period 0, the individual rationality (IR) constraint implies that the expected utility provided to a newly hired executive must equal 1—the value of the outside option. This condition pins down μ_0^{IR}. Equation (15) therefore identifies $A_1^0 g_0^0$ up to a constant, namely μ_0^{IC}.

Because v_2 is also set optimally, we obtain:[5]

$$\begin{aligned}(A_1^0 - (1-p_1^0))g_0^0 = (&-\lambda_1 - \lambda_1 \mathbb{E}_1^* \left[(X_1 - w_2 + \lambda_2/\lambda_1 \Pi_2) - \Pi_{0,1}\right] p_1 (1-p_1) \\ &+(\mu_0^{IC} + \lambda_0)\alpha_1^{(b_0-1)^{-1}} v_1 (A_1^* - (1-p_1))\Big) \left(\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}} v_1\right)^{-1}\end{aligned} \tag{16}$$

Dividing Equation (15) by Equation (16), we obtain an expression that identifies $\frac{A_1^0}{A_1^0-(1-p_1^0)}$ up to μ_0^{IC} and is increasing in μ_0^{IC}. The denominator of this expression represents the continuation value from choosing to stay with the firm, weighted by the probability that the executive chooses to stay. In the Appendix (Lemma 2), we show that the denominator is a known increasing function of p_1^0, while A_1^0 is a known decreasing function of p_1^0. Therefore, the ratio $\frac{A_1^0}{A_1^0-(1-p_1^0)}$ is a known decreasing function of p_1^0. As a result, both p_1^0 and A_1^0 are identified up to μ_0^{IC} from Equations (15) and (16). Finally, from Equation (15), g_0^0 is identified up to a constant, μ_0^{IC}.

In the second case, $X_0 = x \leq \underline{x}$. Therefore, $\phi_1 = 1$. The FOC w.r.t v_1 implies

$$g_0^0 = \frac{(\mu_0^{IC} + \lambda_0)\alpha_1^{(b_0-1)^{-1}} - \lambda_0 v_1^{-1}}{\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}}} \tag{17}$$

Eq. (17) thus identifies g_0^0 up to μ_0^{IC} for $x \leq \underline{x}$.

[5]This equation combines the FOCs w.r.t p_1, and the expectation over x_1 of the FOC w.r.t v_2 multiplied by v_2

In the third case, $X_0 = x \in (\underline{x}, \bar{x})$. Similar to the first case, we obtain

$$(A_1^0(1-\phi_1)+\phi_1)g_0^0 = \frac{(\lambda_0+\mu_0^{IC})\alpha_1^{(b_0-1)^{-1}}(A_1^*(1-\phi_1)+\phi_1)-\lambda_0 v_1^{-1}}{\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}}} \tag{18}$$

and

$$\begin{aligned}
&(A_1^0-(1-p_1^0))g_0^0 \\
&= (-\lambda_1-\lambda_1(1-\phi_1)p_1\mathbb{E}_1^*\left[(X_1-w_2+\lambda_2/\lambda_1\Pi_2)-\Pi_{0,1}\right](1-p_1) \\
&\quad +(\mu_0^{IC}+\lambda_0)\alpha_1^{(b_0-1)^{-1}}v_1(A_1^*-(1-p_1))\Big)\left(\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}}v_1\right)^{-1}
\end{aligned} \tag{19}$$

In addition, because $\phi_1 \in (0,1)$, the FOC w.r.t ϕ_1 provides an additional identifying equation:

$$\begin{aligned}
(1-A_1^0)g_0^0 &= (\lambda_1 p_1^*\mathbb{E}_1^*\left[(X_1-w_2+\lambda_2/\lambda_1\Pi_2 \mid x)-\Pi_{0,1}\right] \\
&\quad +(\lambda_0+\mu_0^{IC})\alpha_1^{(b_0-1)^{-1}}v_1(1-A_1^*)\Big)\left(\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}}\right)^{-1}
\end{aligned} \tag{20}$$

Multiplying Eq. (20) by $(1-\phi_1)$ and adding to Eq. (18), we obtain

$$g_0^0 = \frac{\lambda_1 p_1^*(1-\phi_1)\mathbb{E}_1\left[(X_1-w_2+\lambda_2/\lambda_1\Pi_2 \mid x)-\Pi_{0,1}\right]+(\lambda_0+\mu_0^{IC})\alpha_1^{(b_0-1)^{-1}}}{\mu_0^{IC}\alpha_0^{(b_0-1)^{-1}}} \tag{21}$$

Because $(1-\phi_1)p_1$ is equal to the probability of retention—which is identified from the data—Equation (21) identifies g_0^0 up to the constant μ_0^{IC} for $x \in (\underline{x}, \bar{x})$. Moreover, g_0^0 is also identified up to μ_0^{IC} for $x \notin (\underline{x}, \bar{x})$. Using the fact that $\mathbb{E}_0^*(g_0^0) = 1$, we can identify μ_0^{IC}. Therefore, g_0^0 is fully identified.

Next, we show that Equations (20) and (19), taken together, identify p_1^0 as a linearly increasing function of p_1. On the other hand, Equation (19) alone identifies p_1^0 as a decreasing function of p_1. Therefore, for any x, both p_1^0 and p_1 are identified. The firing probability ϕ_1 is then identified as the ratio between the retention probability—identified from the data—and p_1. The ratchet ratio is also identified.

The following proposition extends the identification argument presented above to an arbitrary period t:

Proposition 4. *If the optimal contract is monotonic, for any period t and any sequence $\boldsymbol{y}$ of outputs up to period $t-1$, both $\phi_{t+1}(\cdot, \boldsymbol{y})$ and $\tilde{A}_{t+1}(\cdot, \boldsymbol{y})$ are identified. Moreover, for any $\tau < t$, $g_t^\tau(\cdot, \boldsymbol{y})$ is also identified. Subsequently, $f_{X_t}(\cdot \mid \theta)$ and the prior belief f_θ are identified.*

We briefly summarize the proof of Proposition 4. Because $\gamma(\cdot)$ is identified, g_t^τ is a known linear function of g_t^t. Therefore, the number of objects to be identified is the same as in the case of $t = 0$, namely, $\tilde{A}_{t+1}$, q_{t+1}, and g_t^t. These objects are identified by forward induction—i.e., assuming that the relevant parameters for earlier periods have already been identified—using the same argument applied at $t = 0$.

To see the identification argument for $f_{X_t}(\cdot \mid \theta)$ and $f_\theta(\cdot)$, recall that the belief about θ degenerates after period $\bar{T}$. Let $\boldsymbol{y}^{\bar{T}}$ denote an output history up to $\bar{T}$, and define $\hat{\theta} = \mathbb{E}(w_t \mid \boldsymbol{y}^{\bar{T}})$. Because the contract beyond $\bar{T}$ depends only on θ and is ordered via first-order stochastic dominance (Proposition 1), there exists a strictly increasing function $\sigma(\cdot)$ such that $\theta = \sigma(\hat{\theta})$. It therefore suffices to identify $\sigma(\cdot)$. On the other hand, under the parametric assumption in Assumption 1, for any t-output x and past history $\boldsymbol{y}$, the conditional expectation $\mathbb{E}_t^*(\sigma(\hat{\theta}) \mid x, \boldsymbol{y})$ is identified and is equal to $g_t^t(x, \boldsymbol{y})/\gamma(x)$. Therefore, $\sigma(\cdot)$ is identified. Hence, $f_{X_t}(\cdot \mid \theta, e_t = 1)$ is also identified.

Finally, we note that the assumption that $\gamma(\cdot)$ is independent of t can be relaxed, allowing the cost of moral hazard to depend on the executive's tenure length. To relax this assumption, we require the existence of different history paths that lead the firm to hold the same belief about θ. More concretely, for each history $\boldsymbol{y}$, there exists a (known) alternative history $\boldsymbol{y}'$ and outputs x, x' such that $\theta \mid (\boldsymbol{y}, x) \overset{d}{=} \theta \mid (\boldsymbol{y}', x')$. Under this assumption, $\gamma_t(\cdot)$ is identified for any t by exploiting the variation in compensation and firing probabilities across $(\boldsymbol{y}, x)$ and $(\boldsymbol{y}', x')$.

5 Estimation

The estimation procedure follows the identification argument outlined in the previous section. In the estimation procedure, due to the small sample size, we assume that $\bar{T} = 8$; that is, learning ceases with certainty after 8 years (3882 observations at the firm-executive-year level). This value is chosen based on empirical patterns observed in the data: firm market value does not appear to increase with tenure beyond 8 years (Figure 3). Moreover, most instances of involuntary termination occur before $t = 8$ (Figure 1), which aligns with our theoretical prediction that once learning concludes, involuntary termination no longer occurs.

To account for variation in observed characteristics across firms and executives, we homogenize the firm's market value by regressing the logarithm of market value on industry fixed effects, the unemployment rate, the consumer price index, executive gender, and executive age. We then normalize Π_t, X_t, and w_{t+1} by dividing each by the predicted firm

market value from this regression. We also allow $\log(\alpha_1)$ and $\log(\rho)$ to be linear functions of industry dummies, the executive's age at the start of their tenure as CEO, and gender.

We further parameterize $\gamma(\cdot)$ to be a spline function of X_t. We first estimate α_1 and ρ using GMM, with the moments constructed from the participation constraint of executives. We then estimate $\gamma(\cdot)$ from the variation in annuitized compensation across $(x, \boldsymbol{y})$ with nonlinear least squares. Finally, α_0 and θ are nonparametrically estimated given α_1, ρ, γ using the binding IC constraints and the property of the likelihood ratio — i.e., $\mathbb{E}_t^*(\theta\gamma(x) \mid \theta) = 1$.

Estimation of q_t and $\frac{\tilde{A}_t}{A_t}$ for $t < \bar{T}$ In this step, instead of conditioning compensation and separation decisions directly on the full history of past performance, we reduce the dimensionality of the conditioning variables by leveraging the observation that the effect of past periods on current compensation can be summarized by two sufficient statistics, defined as follows:

$$\Delta_t := \sum_{t'=1}^{t-1} (\mu_{t'}^{IR} + \mu_{t'}^{IC}) v_{t'+1} \prod_{\tau=t'}^{t-1} \frac{A_\tau - (1 - p_\tau)}{\mathbb{E}_\tau \left[A_{\tau+1}(1 - q_{\tau+1} + q_{\tau+1}) v_{\tau+1} \right]},$$

and

$$\tilde{\Delta}_t := \sum_{t'=1}^{t-1} \mu_{t'}^{IC} v_{t'+1} \prod_{\tau=t'}^{t-1} \frac{A_\tau^{\tau-1} - (1 - p_\tau^{\tau-1})}{\mathbb{E}_\tau \left[A_{\tau+1}^{\tau-1}(1 - q_{\tau+1} + q_{\tau+1}) v_{\tau+1} \right]} \gamma(x_\tau).$$

Thus, the relevant conditioning variables summarizing the past reduce to Δ_t, $\tilde{\Delta}_t$, and the firm's belief about θ. In estimation, we assume that the firm's belief about θ at each t can be parameterized by the prior mean at the beginning of period t.

We further parameterize $\underline{x}$, $\bar{x}$, and q_{t+1} as spline functions of Δ_t, $\tilde{\Delta}_t$, and $\mathbb{E}_t^*(\theta)$, with the restriction that q_{t+1} equals 0 at $\underline{x}$, 1 at $\bar{x}$, and is strictly decreasing between these two points. We also assume that $\frac{\tilde{p}_{t+1} - p_{t+1}}{1 - p_{t+1}}$ is a positive spline function of $\mathbb{E}_t^*(\theta)$, Δ_t, and $\tilde{\Delta}_t$.

Given these parameters, we compute $\mathbb{E}_t^*(\theta)$ from the FOC with respect to v_{t+1}, and obtain p_{t+1} from kernel estimates of the retention probability, divided by $(1 - q_{t+1})$. We then compute A_{t+1} based on p_{t+1}. Given $\underline{x}$, we calculate Π_0 as the expected market value of executives with $x < \underline{x}$. Using the FOC with respect to p_{t+1}, we compute μ_{t+1}^{IR}. The parameters are then chosen to minimize (1) the probability of forced turnovers for $x > \bar{x}$, (2) the absolute value of the FOC with respect to q_{t+1}, (3) the squared values of the IC and IR constraints, and (4) the squared value of the difference between A_t and the same continuation value constructed from $(A_{t+1}, q_{t+1}, w_{t+1})$. Standard errors are bootstrapped based on 100 iterations.

Discussion Because of the small sample size, our estimation procedure does not directly exploit the kink property in Proposition 3 to identify $\bar{x}$. Instead, we rely on observed data on forced turnovers, under the assumption that the probability of a forced turnover is zero for $x > \bar{x}$. As a result, our estimates are robust to type-II errors in the classification of forced turnovers.

5.1 Estimation Results

Table 7: Estimated risk premium in millions of dollars

Industry	Average	SD
Finance and Insurance	0.05	0.12
Information	0.13	0.48
Aerospace	0.02	0.08
Chemical	0.17	0.64
Other	0.09	0.35

Note: SEs are not reported. SD statistics refer to the standard deviation across estimates for different executives and firms in the data.

Table 7 reports the estimated risk premiums, defined as the difference between average annual compensation and the certainty equivalent, measured in millions of dollars, across different industries (based on NAICS classification). The average risk premium across industries ranges from \$20,000 to \$170,000, with substantial variation across executives and firms within the same industry.

Table 8: Estimated cost of effort in millions of dollars

Industry	Average	SD
Finance and Insurance	4.2	0.91
Information	2.98	0.77
Aerospace	3.74	0.86
Chemical	0.93	0.2
Other	2.56	1.39

Note: SEs are not reported. SD statistics refer to the standard deviation across estimates for different executives and firms in the data.

Table 8 reports the estimated cost of effort (in dollars) for executives across different industries. Relative to other industries, firms in the chemical industry exhibit the lowest cost of effort, whereas those in the finance and insurance industry have the highest. For

example, the cost of exerting effort for an executive in the chemical industry is 0.93 millions, compared to 4.2 millions for those in finance and insurance.

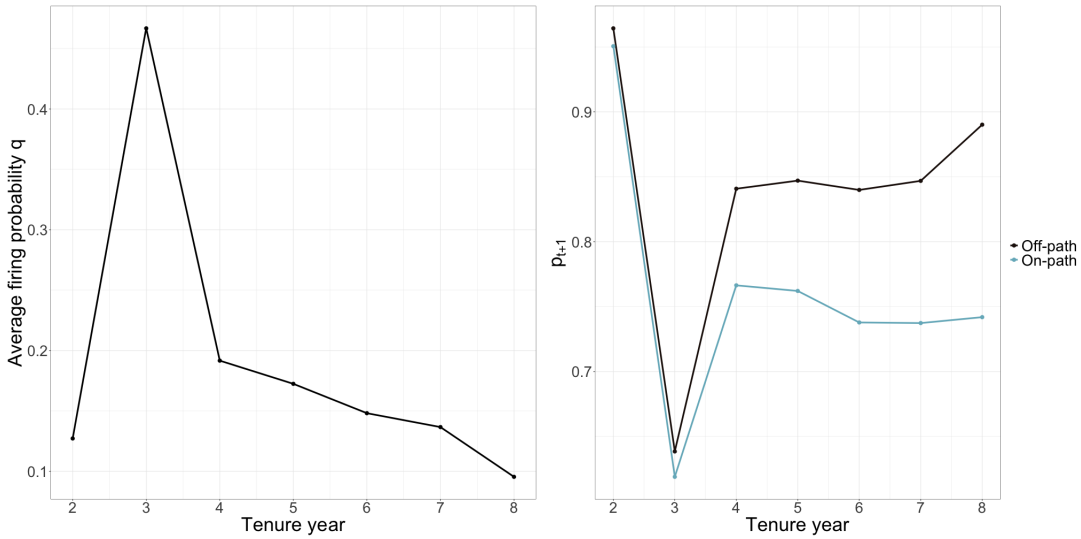
Table 9: Estimated per-year cost of moral hazard in millions of dollars

Industry	Average	SD
Finance and Insurance	1050	877
Information	1131	778
Aerospace	626	490
Chemical	884	715
Other	836	853

Note: SEs are not reported. SD statistics refer to the standard deviation across estimates for different executives and firms in the data.

Table 9 reports the annual per-period output difference between firms whose executives shirk and those whose executives exert effort, while Figure 13 illustrates the estimated likelihood ratio $\gamma(\cdot)$. These output losses are substantially greater for firms than the cost of effort borne by executives, consistent with our modeling assumption that firms have a strong incentive to induce effort. For instance, in the information industry, a firm loses \$1.13 billion in output per period when an executive shirks, compared to a cost of effort of only \$2.98 million for the executive. The lower output under shirking arises from the negative relationship between output and the likelihood ratio of output under effort versus shirking.

Figure 10: Estimated firing probability and informational rent from shirking



Finally, Figure 10 shows the estimated firing probability across tenure years (left panel) and the informational rent executives earn from shirking (right panel). Our estimates in-

dicate that some firing occurs after the first year, with an average firing probability of approximately 42% at the end of year two. Firing probabilities then decline significantly, averaging between 10% and 20% in later years. The estimates also suggest that executives accrue substantial informational rents from shirking. For instance, when shirking in period 3, an executive anticipates a higher continuation value, reflected in a 0.83 probability of staying with the firm in the next period, compared to an on-path voluntary retention probability of only 0.77.

6 Counterfactuals (to be added)

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.

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.

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 F. Kanaan (2015): "CEO turnover and relative performance evaluation," *the Journal of Finance*, 70, 2155–2184.

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.

Peters, F. S. and A. F. Wagner (2014): "The executive turnover risk premium," *The Journal of Finance*, 69, 1529–1563.

Prat, J. and B. Jovanovic (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.

A Additional Tables and Figures

B Proof of Section 3

B.1 Proof of Proposition 1

To prove this Proposition, we first show two lemmas.

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

Proof of Lemma 1. Let ϵ_t^* denote the threshold of ϵ_t at which the executive is indifferent between staying and (voluntarily) leaving. Therefore, $\epsilon_t^* = F_\epsilon^{-1}(1 - p_t)$. Let $\varepsilon = \exp(\epsilon)$, and $\varepsilon^* = \exp(\epsilon^*)$. Therefore, the density function of ε is $\frac{1}{(1+\varepsilon)^2}$

$$A - (1 - p) = \int_{\varepsilon^*} \frac{1}{\varepsilon} f_\varepsilon \varepsilon d\varepsilon \underline{u}(p)^{1-b^{-1}}$$

$$\implies \frac{\partial}{\partial p} A - (1 - p) = \frac{1}{\varepsilon^*} \underline{u}(p)^{1-b^{-1}} - \int_{\varepsilon^*} \frac{1}{\varepsilon} f_\varepsilon \varepsilon d\varepsilon (1 - b^{-1}) \underline{u}(p)^{-b^{-1}} \underline{u}(p) \frac{1}{f_\varepsilon \varepsilon^*} (b - 1)^{-1}$$

$$\geq 0$$

Figure 11: Decomposition of variance in compensations within and across executives

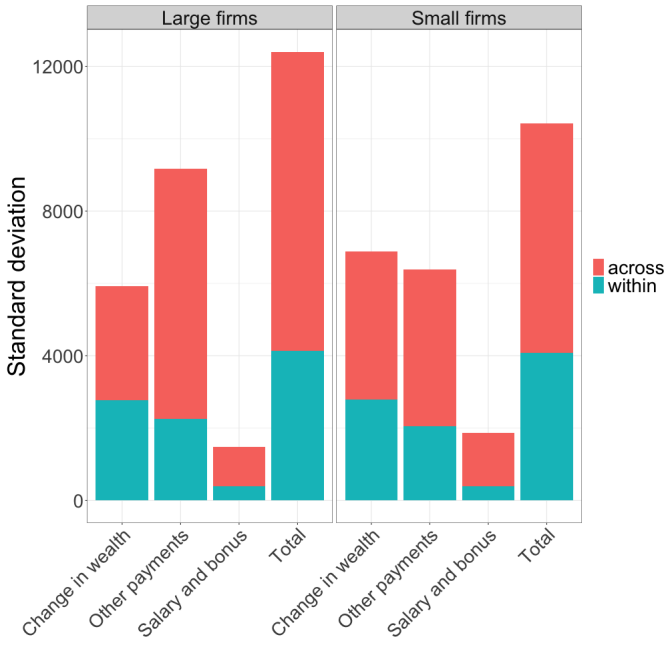
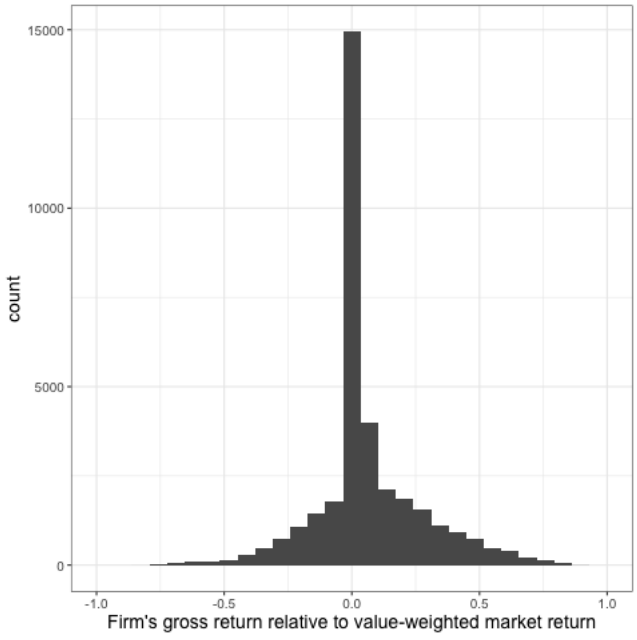
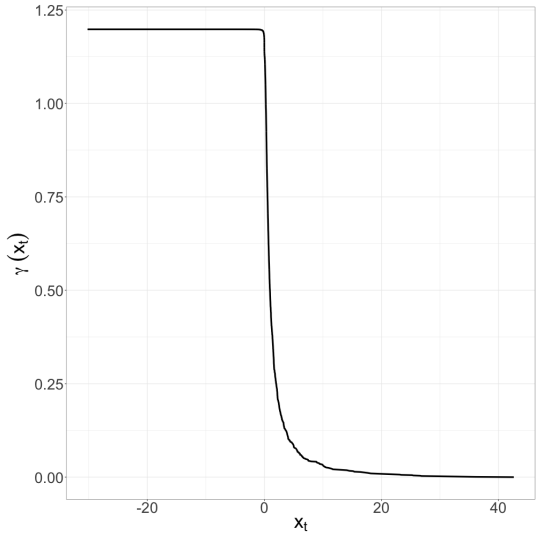


Figure 12: Distribution of constructed gross return



where the final inequality is because $b_t > 1$, and f_ε is decreasing. □

Figure 13: Estimated likelihood ratio $\gamma(\cdot)$



Lemma 2. *Conditional on $b_t = b$:*

$$\underline{u}(p) = \frac{1-p}{p}(b-1)^{-1}$$
$$(\log(\underline{u}(p)))' = -\left(p(1-p)(b-1)\right)^{-1}$$

The FOC w.r.t v_{t+1} is given by:

$$\begin{aligned} 0 = \lambda_t v_{t+1}^{-1} - (\mu_t^{IC} + \mu_t^{IR})\alpha_1^{(b_{t+1}-1^{-1})}\left[A_{t+1}^*(1-\phi_{t+1}) + \phi_{t+1}\right] \\ +\mu_t^{IC}\alpha_0^{(b_{t+1}-1^{-1})} g_t^t\left[A_{t+1}(1-\phi_{t+1}) + \phi_{t+1} \mid e_t = 0\right] \\ -\sum_{\tau=0}^{t-1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1}\frac{\partial A_{\tau+1}^*}{\partial v_{t+1}} \\ +\sum_{\tau=0}^{t-1}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(g_t^\tau v_{\tau+1}\frac{\partial A_{\tau+1}^\tau}{\partial v_{t+1}}) \end{aligned} \tag{22}$$

Taking expectation of (22) multiplied by v_{t+1} over x_{t+1} yields:

$$\begin{aligned} 0 = \lambda_t - \mu_t^{IR}\underline{u}(p_t) - \mathbb{E}_t\Bigg[\sum_{\tau=0}^{t-1} v_{t+1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1}\frac{\partial A_{\tau+1}^*}{\partial v_{t+1}} \\ -\sum_{\tau=0}^{t-1}\mu_\tau^{IC} v_{t+1}\alpha_0^{(b_{\tau+1}-1)^{-1}}(v_{\tau+1}\frac{\partial A_{\tau+1}^\tau}{\partial v_{t+1}})\Bigg] \end{aligned} \tag{23}$$

The FOC w.r.t p_t is given by

$$0 = \lambda_t(1-\phi_t)\mathbb{E}_t\left[(x_t - w_{t+1} + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right] + \mu_t^{IR}\underline{u}'(p_t) \tag{24}$$

The FOC w.r.t ϕ_{t+1} is given by

$$\begin{aligned}
0 = -\lambda_{t+1}p_{t+1}\mathbb{E}_{t+1}\left[(x_{t+1} - w_{t+2} + \lambda_{t+2}/\lambda_{t+1}\Pi_{t+2}) - \Pi_{0,t+1}\right] \\
-(\mu_t^{IC} + \mu_t^{IR})\alpha_1^{(b_{t+1}-1^{-1})}\left[1 - A_{t+1}^*\right]v_{t+1} \\
+\mu_t^{IC}\alpha_0^{(b_{t+1}-1^{-1})}g_t^t\left[1 - A_{t+1} \mid e_t = 0\right]v_{t+1} \\
-\sum_{\tau=0}^{t-1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}}v_{\tau+1}\frac{\partial A_{\tau+1}^*}{\partial \phi_{t+1}} \\
+\sum_{\tau=0}^{t-1}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(g_t^\tau v_{\tau+1}\frac{\partial A_{\tau+1}^\tau}{\partial \phi_{t+1}}) - \mu_{t+1}^{q,1} + \mu_{t+1}^{q,0}
\end{aligned} \tag{25}$$

Consider $t \geq \bar{T}$. That is θ has been learned. Therefore, $g_t^\tau = 1$, and $(A_t \mid e_\tau = 0) = A_t^*$. Eq. (25) for ϕ_{t+1} is simplified to

$$\begin{aligned}
0 = -\lambda_t p_{t+1}\mathbb{E}_{t+1}\left[(x_{t+1} - w_{t+2} + \lambda_{t+1}/\lambda_t\Pi_{t+2}) - \Pi_{0,t+1}\right] \\
-(\mu_t^{IC} + \mu_t^{IR})\alpha_1^{(b_{t+1}-1^{-1})}\left[1 - A_{t+1}^*\right] + \mu_t^{IC}\alpha_0^{(b_{t+1}-1^{-1})}g_t^t\left[1 - A_{t+1}^*\right] \\
-\sum_{\tau=0}^{t-1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}}v_{\tau+1}\frac{\partial A_{\tau+1}^*}{\partial \phi_{t+1}} + \sum_{\tau=0}^{t-1}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(v_{\tau+1}\frac{\partial A_{\tau+1}^\tau}{\partial \phi_{t+1}}) + \mu_{t+1}^{q,0} - \mu_{t+1}^{q,1}
\end{aligned} \tag{26}$$

Similarly, Eq. (22) becomes

$$\begin{aligned}
0 = \lambda_t v_{t+1}^{-1} - (\mu_t^{IC} + \mu_t^{IR})\alpha_1^{(b_{t+1}-1^{-1})}\left[A_{t+1}^*(1-\phi_{t+1}) + \phi_{t+1}\right] \\
+\mu_t^{IC}\alpha_0^{(b_{t+1}-1^{-1})}g_t^t\left[A_{t+1}^*(1-\phi_{t+1}) + \phi_{t+1}\right] \\
-\sum_{\tau=0}^{t-1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}}v_{\tau+1}\frac{\partial A_{\tau+1}^*}{\partial v_{t+1}} \\
+\sum_{\tau=0}^{t-1}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(v_{\tau+1}\frac{\partial A_{\tau+1}^\tau}{\partial v_{t+1}})
\end{aligned} \tag{27}$$

We first show that $\phi_{t+1} = 1$ is not optimal. Suppose, for a contradiction, that $\phi_{t+1} = 1$

under the optimal contract. (26) becomes:

$$0 = -\lambda_t p_{t+1} \mathbb{E}_{t+1}\left[(x_{t+1} - w_{t+2} + \lambda_{t+1}/\lambda_t \Pi_{t+2}) - \Pi_{0,t+1}\right] \\ -\lambda_t(1 - A^*_{t+1}) - \mu^{q,1}_{t+1} \tag{28}$$

Therefore, it must be that $\mathbb{E}_{t+1}\left[(x_{t+1} - w_{t+2} + \lambda_{t+1}/\lambda_t \Pi_{t+2}) - \Pi_{0,t+1}\right]$ is negative for any positive value of p_{t+1}. However, because ϵ is unbounded, there exists $\rho > 0$ sufficiently small such that $\mathbb{E}_{t+1}(w_{t+2}|p_{t+1} = \rho, \phi_{t+2} = 0) > \mathbb{E}_{t+1}\left[(x_{t+1} + \lambda_{t+1}/\lambda_t \Pi_{t+2}) - \Pi_{0,t+1}\right]$ where w_{t+2} satisfies the IC constraint in period $t+1$. Therefore, this is a contradiction.

Now, we show that $\phi_{t+1} = 0$ is optimal. Because $A^*_{t+1} < 1$ by definition, Eq. (27) implies

$$0 > \lambda_t v^{-1}_{t+1} - (\mu^{IC}_t + \mu^{IR}_t)\alpha_1^{(b_{t+1}-1^{-1})} \\ +\mu^{IC}_t \alpha_0^{(b_{t+1}-1^{-1})} g^t_t \\ -\sum_{\tau=0}^{t-1}(\mu^{IC}_\tau + \mu^{IR}_\tau)\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1} \frac{\partial A^*_{\tau+1}}{\partial A^*_t} \frac{A^*_t - (1-p_t)}{\mathbb{E}\left[v_{t+1}(A^*_{t+1}(1-\phi_{t+1}) + \phi_{t+1})\right]} \\ +\sum_{\tau=0}^{t-1} \mu^{IC}_\tau \alpha_0^{(b_{\tau+1}-1)^{-1}} (v_{\tau+1} \frac{\partial A_{\tau+1}}{\partial A_t} \frac{A_t - (1-p_t)}{\mathbb{E}\left[v_{t+1}(A_{t+1}(1-\phi_{t+1}) + \phi_{t+1})\right]} \mid e_\tau = 0) \tag{29}$$

With some algebra, we can show that $\frac{\partial A_{\tau+1}}{\partial v_{t+1}} v_t + \frac{\partial A_{\tau+1}}{\partial \phi_{t+1}}(1-\phi_{t+1}) = \frac{\partial A_{\tau+1}}{\partial A_t} \frac{A_t - (1-p_t)}{\mathbb{E}[v_{t+1}(A_{t+1}(1-\phi_{t+1})+\phi_{t+1})]}$. Multiplying (27) by v_{t+1} and adding (26) multiplied by $1 - \phi_{t+1}$ yields

$$0 = -\lambda_{t+1} p_{t+1}(1-\phi_{t+1}) \mathbb{E}_{t+1}\left[(x_{t+1} - w_{t+2} + \lambda_{t+2}/\lambda_{t+1} \Pi_{t+2}) - \Pi_{0,t+1}\right] \\ -(\mu^{IC}_t + \mu^{IR}_t)\alpha_1^{(b_{t+1}-1^{-1})} v_{t+1} + \mu^{IC}_t \alpha_0^{(b_{t+1}-1^{-1})} g^t_t v_{t+1} \\ -\sum_{\tau=0}^{t-1}(\mu^{IC}_\tau + \mu^{IR}_\tau)\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1} \frac{\partial A_{\tau+1}}{\partial A_t} \frac{A_t - (1-p_t)}{\mathbb{E}\left[v_{t+1}(A_{t+1}(1-\phi_{t+1}) + \phi_{t+1})\right]} v_{t+1} \\ +\sum_{\tau=0}^{t-1} \mu^{IC}_\tau \alpha_0^{(b_{\tau+1}-1)^{-1}} v_{\tau+1} \left(\frac{\partial A_{\tau+1}}{\partial A_t} \frac{A_t - (1-p_t)}{\mathbb{E}\left[v_{t+1}(A_{t+1}(1-\phi_{t+1}) + \phi_{t+1})\right]} \mid e_\tau = 0 \right) v_{t+1} + \mu^{q,0}_{t+1} \tag{30}$$

Combine Eq. (31) with Eq. (29)

$$0 \leq -\lambda_{t+1} p_{t+1}(1-\phi_{t+1}) \mathbb{E}_{t+1}\left[(x_{t+1} - w_{t+2} + \lambda_{t+2}/\lambda_{t+1} \Pi_{t+2}) - \Pi_{0,t+1}\right] \\ -\lambda_t + \mu^{q,0}_{t+1} \tag{31}$$

Therefore, $\mu^{q,0}_{t+1} > 0$, which implies $\phi_{t+1} = 0$.

$$0 = \lambda_t - \mu_t^{IR}\underline{u}(p_t) - (A_t - (1-p_t))\left[\sum_{\tau=0}^{t-1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1}\frac{\partial A^*_{\tau+1}}{\partial A_t}\right.$$
$$\left. - \sum_{\tau=0}^{t-1}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(v_{\tau+1}\frac{\partial A^\tau_{\tau+1}}{\partial A_t})\right] \quad (32)$$

$$\implies 0 = \lambda_t + \lambda_t \mathbb{E}_t\left[(x_t - w_{t+1} + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right]\frac{\underline{u}(p_t)}{\underline{u}(p_t)'} \quad (33)$$

$$-(A_t - (1-p_t))\left[\sum_{\tau=0}^{t-1}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1}\frac{\partial A^*_{\tau+1}}{\partial A_t}\right.$$
$$\left. - \sum_{\tau=0}^{t-1}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(v_{\tau+1}\frac{\partial A^\tau_{\tau+1}}{\partial A_t})\right] \quad (34)$$

$$\implies 0 = \lambda_t + \lambda_t \mathbb{E}_t\left[(x_t - w_{t+1} + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right]\frac{\underline{u}(p_t)}{\underline{u}(p_t)'} \quad (35)$$

$$-\frac{A_t - (1-p_t)}{A_t}\left[\sum_{\tau=0}^{t-2}(\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1}\frac{\partial A^*_{\tau+1}}{\partial v_t}v_t\right.$$
$$\left. - \sum_{\tau=0}^{t-2}\mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}}(v_{\tau+1}\frac{\partial A^\tau_{\tau+1}}{\partial v_t}v_t)\right]$$
$$-\frac{A_t - (1-p_t)}{A_t}\left[(\mu_{t-1}^{IC} + \mu_{t-1}^{IR})\alpha_1^{(b_t-1)^{-1}} v_t A_t - \mu_{t-1}^{IC}\alpha_0^{(b_t-1)^{-1}} v_t A_t\right] \quad (36)$$

$$\implies 0 = \lambda_t\frac{1-p_t}{A_t} + \lambda_t \mathbb{E}_t\left[(x_t - w_{t+1} + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right]\frac{\underline{u}(p_t)}{\underline{u}(p_t)'} \quad (37)$$

$$\implies 0 = \lambda_t A_t^{-1} - \lambda_t \mathbb{E}_t\left[(x_t - w_{t+1} + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right]p_t(1-b_t)^{-1} \quad (38)$$

$$\implies \lambda_t\left[A_t^{-1}p_t^{-1} + \mathbb{E}_t(w_{t+1})(1-b_t)^{-1}\right] = \lambda_t \mathbb{E}_t\left[(x_t + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right](1-b_t)^{-1} \quad (39)$$

where Eq. (34) is derived from $\frac{\partial A_{\tau+1}}{\partial A_t} = \frac{\partial A_{\tau+1}}{\partial v_t}A_t^{-1}v_t$, and Eq. (36) is derived from the FOC w.r.t v_t. Therefore, Eq. (39) implies that p_t is independent of all past histories, conditional on θ (and thus conditional on $(x_0, ..., x_{T-1})$.

Next, we show that for $t > \bar{T}$, p_t is increasing in θ. Note that at the optimal choice of p_t, the derivative w.r.t p_t of the LHS of (39) must be positive. Therefore, at $t = \max T$, an executive is certain to retire in the next period. Therefore, $\Pi_{t+1} = \Pi_{0,t+1}$ and is independent of θ. Therefore, p_t and π_t are increasing in θ for $t = \max T$. Using backward induction, we show that the RHS of (39) is always increasing in θ. Therefore, p_t is increasing in θ for any $t > \bar{T}$.

Because p_t only dependent on θ for $t > \bar{T}$, (22) implies that w_{t+1} is also independent of past histories, conditional on θ. The contract at t is now identical to that of a short-term contract, where the slope of compensation is an inverse function of the likelihood ratio, which is $\gamma(x_t)\theta$. Therefore, w_{t+1} is monotonic in x_t.■

B.2 Proof of Proposition 2

We organize the proof of Proposition 3 into several lemmas.

To simplify notation, we assume in this proof that $\bar{T}$ is a constant. The argument extends straightforwardly to the case where $\bar{T}$ is a deterministic function of the output history, though at the cost of more cumbersome notation.

Lemma 3. *Under a monotonic contract, the most profitable deviation for the executive involves shirking in exactly one period.*

Proof. Consider an executive who has previously shirked once. Let $l^*(x)$ denote the executive's subsequent decision whether to leave the firm on the equilibrium path. Let $\bar{\theta} \in supp(\theta)$.

$$\frac{\mathbb{E}^*(vA\theta\gamma(x) \mid \theta \leq \bar{\theta}, l^*(x))F_\theta(\bar{\theta})}{\mathbb{E}^*(vA\theta\gamma(x) \mid l^*(x))} \tag{40}$$

The above is a well-defined cumulative distribution of $\bar{\theta}$. Because the contract is monotonic, vA is decreasing in x and in θ. Therefore, from the affiliation property of θ and x:

$$\mathbb{E}^*(vA\theta\gamma(x) \mid \theta \leq \bar{\theta}, l^*(x)) \leq \mathbb{E}^*(vA \mid \theta \leq \bar{\theta}, l^*(x))\mathbb{E}(\gamma(x)\theta \mid \theta \leq \bar{\theta}, l^*(x)) = \mathbb{E}^*(vA \mid \theta \leq \bar{\theta}, l^*(x))$$

Therefore, using the IC constraint in period t:

$$\frac{\alpha_0^{(b-1)^{-1}}\mathbb{E}^*(vA\theta\gamma(x) \mid \theta \leq \bar{\theta}, l^*(x))}{\alpha_0^{(b-1)^{-1}}\mathbb{E}^*(vA\theta\gamma(x) \mid l^*(x))} \succeq_{FOSD} \frac{\alpha_1^{(b-1)^{-1}}\mathbb{E}^*(vA \mid \theta \leq \bar{\theta}, l^*(x))}{\alpha_1^{(b-1)^{-1}}\mathbb{E}^*(vA \mid l^*(x))} \tag{41}$$

which implies

$$\frac{\mathbb{E}^*(vA\gamma(x)\theta^2 \mid l^*(x))}{\mathbb{E}^*(vA\gamma(x)\theta \mid l^*(x))} \geq \frac{\mathbb{E}^*(vA\theta \mid l^*(x))}{\mathbb{E}^*(vA \mid l^*(x))} \tag{42}$$

Therefore,

$$\frac{\mathbb{E}(\alpha_0^{(b-1)^{-1}}vA\gamma(x)\theta \mid l^*(x))}{\mathbb{E}(\alpha_1^{(b-1)^{-1}}vA \mid l^*(x))} \geq \frac{\mathbb{E}^*(\alpha_0^{(b-1)^{-1}}vA\gamma(x)\theta \mid l^*(x))}{\mathbb{E}^*(\alpha_1^{(b-1)^{-1}}vA \mid l^*(x))} \tag{43}$$

where $\mathbb{E}$ now denotes the belief of an executive who has shirked once. Therefore

$$\frac{\mathbb{E}(\alpha_0^{(b-1)^{-1}} vA\gamma(x)\theta \mid l^*(x))}{\mathbb{E}(\alpha_1^{(b-1)^{-1}} vA \mid l^*(x))} \geq \frac{\min_{l(x)} \mathbb{E}^*(\alpha_0^{(b-1)^{-1}} vA\gamma(x)\theta \mid l(x))}{\mathbb{E}^*(\alpha_1^{(b-1)^{-1}} vA \mid l^*(x))} \geq 1 \tag{44}$$

Therefore, for any $l(x)$

$$\frac{\mathbb{E}(\alpha_0^{(b-1)^{-1}} vA\gamma(x)\theta \mid l(x))}{\mathbb{E}(\alpha_1^{(b-1)^{-1}} vA \mid l(x))} \geq 1 \tag{45}$$

which implies

$$\frac{\min_l \mathbb{E}(\alpha_0^{(b-1)^{-1}} vA\gamma(x)\theta \mid l(x))}{\min_l \mathbb{E}(\alpha_1^{(b-1)^{-1}} vA \mid l(x))} \geq 1 \tag{46}$$

Thus, an executive who has shirked once will also choose to work in the next period. The proof extends straightforwardly to any past history of shirking. □

Lemma 4. *Consider a monotonic contract. The ratchet ratio $\frac{A}{A^*}g$ is decreasing in past outputs.*

Proof. Let l denote the leaving decision of an executive at any future output realization. We first show $\frac{A|l}{A^*|l}$ is decreasing in past outputs y. Consider $y' > y$ and $\bar{\theta}$ in the support of θ.

First, note that $\mathbb{E}(A \mid \theta, l, y)$ is a decreasing function of θ. For $a < a'$

$$\frac{\int^{a'} A df_A(a|y)}{\int^{a} A df_A(a \mid y, l)} \geq \frac{\int^{a'} A df_A(a|y,l)}{\int A df_A(a|y,l)} \tag{47}$$

Therefore:

$$\frac{\mathbb{E}(A\theta|\theta \geq \bar{\theta}, y, l)}{\mathbb{E}(A|\theta \geq \bar{\theta}, y, l)} \geq \frac{\mathbb{E}(A\theta \mid y, l)}{\mathbb{E}(A \mid y, l)} \tag{48}$$

which implies

$$\frac{\mathbb{E}(A\theta|\theta \leq \bar{\theta}, y, l)}{\mathbb{E}(A\theta \mid y, l)} \preceq_{FOSD} \frac{\mathbb{E}(A|\theta \leq \bar{\theta}, y, l)}{\mathbb{E}(A \mid y, l)} \tag{49}$$

Because of the affiliation property between θ and y, $\frac{f_\theta(\theta|y')}{f_\theta(\theta|y)}$ is increasing in θ. Therefore

$$\frac{\mathbb{E}(A\theta|y', l)}{\mathbb{E}(A\theta \mid y, l)} \leq \frac{\mathbb{E}(A|y', l)}{\mathbb{E}(A \mid y, l)} \implies \frac{\mathbb{E}(A\theta|y', l)}{\mathbb{E}(A \mid y', l)} \leq \frac{\mathbb{E}(A\theta|y, l)}{\mathbb{E}(A \mid y, l)} \tag{50}$$

Next, we show that $\min_l \mathbb{E}(A \mid y, l)$ is an increasing function $m(\cdot)$ of $\mathbb{E}(A|y, l = 0)$, and this function has decreasing elasticity. This follows directly from the definition of A, and the parametric distribution of ϵ.

Therefore, for $a' > a$:

$$\frac{\partial}{\partial y}\left[\ln m(a) - \ln m(a')\right] \tag{51}$$

$$= \frac{1}{m(a)} m'(a)\frac{\partial a}{\partial y} - \frac{1}{m(a')} m(a')\frac{\partial a}{\partial y'} \tag{52}$$

$$= \frac{1}{m(a)} m'(a) a \frac{\partial a}{\partial y} a^{-1} - \frac{1}{m(a')} m(a') a' \frac{\partial a'}{\partial y} a'^{-1} \tag{53}$$

$$\leq -\frac{\partial a'}{\partial y} a'^{-1} \left(\frac{m'(a)a}{m(a)} - \frac{m(a')a'}{m(a')} \right) \leq 0 \tag{54}$$

□

Lemma 5. *Consider a monotonic contract. Let p^* denote the probability that an executive with belief $\mathbb{E}^*$ accepts the contract prior to realizing ϵ and receives continuation value A^*. Let $\tilde{p}$ denote the probability that the executive accepts the contract under belief $\tilde{\mathbb{E}} \neq \mathbb{E}^*$, which results from having shirked once in the past, and earns a continuation value $\tilde{A}$. Then, the following expressions are decreasing in past output:*

$$\frac{1 - \tilde{A}}{1 - A^*} \times g \quad \textit{and} \quad \frac{\tilde{A} - (1 - \tilde{p})}{A^* - (1 - p^*)} \times g.$$

Proof. The proof is similar to that of Lemma 4 and is therefore omitted. □

Proof of Proposition 2. Consider $t = \bar{T} - 1$. The FOC w.r.t v_{t+1} is given by (27), where $\phi_{t+1} = 0$ from Proposition 1. Using the result of Proposition 1, we abuse notation and use $A^*_{t+1}(\theta)$ to indicate the continuation value of an executive with a (revealed) type θ. Eq. (27) is thus equivalent to

$$\begin{aligned} 0 = \lambda_t v_{t+1}^{-1} - (\mu_t^{IC} + \mu_t^{IR})\alpha_1^{(b_{t+1}-1^{-1})} A^*_{t+1}(\theta) + \mu_t^{IC}\alpha_0^{(b_{t+1}-1^{-1})} g_t^t A^*_{t+1}(\theta) \\ - \sum_{\tau=0}^{t-1} (\mu_\tau^{IC} + \mu_\tau^{IR})\alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1} \frac{\partial A^*_{\tau+1}}{\partial v_{t+1}} + \sum_{\tau=0}^{t-1} \mu_\tau^{IC}\alpha_0^{(b_{\tau+1}-1)^{-1}} \gamma(x_\tau)(v_{\tau+1}\frac{\partial A^\tau_{\tau+1}}{\partial v_{t+1}}) \end{aligned} \tag{55}$$

where $g_t^t(x_t) = g_t^t(x_t|\theta) = \gamma(x_t)$ using Assumption 1. Because $\gamma(x_t)$ is decreasing in x_t, $v_{t+1}A^*_{t+1}$ is increasing in x_t. Therefore, a monotonic contract satisfies this neccessary condition.

Next, we consider whether there exists a monotonic contract that satisfies the FOC w.r.t p_t at $t = \bar{T} - 1$. Consider the expectation of the FOC w.r.t v_{t+1} multiplied by v_{t+1} (Eq. (23)), combined with the FOC w.r.t p_t, and the FOC w.r.t v_t.

$$0 = \lambda_t - \mu_t^{IR}\underline{u}(p_t) - \sum_{\tau=0}^{t-1}\left(\mu_\tau^{IR} + \mu_\tau^{IC}\right)\alpha_1^{(b_\tau-1)^{-1}} v_{\tau+1}\mathbb{E}_t\left[\frac{\partial A^*_{\tau+1}}{\partial v_{t+1}} v_{t+1}\right] + \mu_\tau^{IC}\alpha_0^{(b_\tau-1)^{-1}} v_{\tau+1}\mathbb{E}_t\left[\frac{\partial A_{\tau+1}}{\partial v_{t+1}} v_{t+1} \mid e_\tau = 0\right]\gamma(x_\tau) \tag{56}$$

$$\implies 0 = \lambda_t + \lambda_t\mathbb{E}_t\left[(x_t - w_{t+1} + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right](1-\phi_t)\frac{\underline{u}(p_t)}{\underline{u}(p_t)'} - \sum_{\tau=0}^{t-1}\left(\mu_\tau^{IR} + \mu_\tau^{IC}\right)\alpha_1^{(b_\tau-1)^{-1}} v_{\tau+1}\mathbb{E}_t\left[\frac{\partial A^*_{\tau+1}}{\partial v_{t+1}} v_{t+1}\right] + \mu_\tau^{IC}\alpha_0^{(b_\tau-1)^{-1}} v_{\tau+1}\mathbb{E}_t\left[\frac{\partial A_{\tau+1}}{\partial v_{t+1}} v_{t+1} \mid e_\tau = 0\right]\gamma(x_\tau) \tag{57}$$

Henceforth, we refer to $\lambda_t\mathbb{E}_t\left[(x_t + \lambda_{t+1}/\lambda_t\Pi_{t+1}) - \Pi_{0,t}\right]$ as the marginal benefit of increasing p_t, and the marginal cost is given by

$$-\left((1-\phi_t)\frac{\underline{u}(p_t)}{\underline{u}(p_t)'}\right)^{-1}\left(\lambda_t - \lambda_t\mathbb{E}_t w_{t+1}(1-\phi_{t+1})\frac{\underline{u}(p_t)}{\underline{u}(p_t)'} - \sum_{\tau=0}^{t-1}\left(\mu_\tau^{IR} + \mu_\tau^{IC}\right)\alpha_1^{(b_\tau-1)^{-1}} v_{\tau+1}\mathbb{E}_t\left[\frac{\partial A^*_{\tau+1}}{\partial v_{t+1}} v_{t+1}\right] + \mu_\tau^{IC}\alpha_0^{(b_\tau-1)^{-1}} v_{\tau+1}\mathbb{E}_t\left[\frac{\partial A_{\tau+1}}{\partial v_{t+1}} v_{t+1} \mid e_\tau = 0\right]\gamma(x_\tau)\right) \tag{58}$$

We prove by contradiction. Suppose that there exists $y < y'$ such that $p_t(y) > p_t(y')$. This implies that, at a given p, the difference in the marginal cost of p_t at y' and y outweighs $y' - y$. Specifically, the change in marginal cost is given by

$$\lambda_t\left[\mathbb{E}_t\left(w_{t+1} \mid y'\right) - \mathbb{E}_t\left(w_{t+1} \mid y\right)\right] + \mu_{t-1}^{IC}\left[(A_t - (1-p_t) \mid y')\gamma(y') - (A_t - (1-p_t) \mid y)\gamma(y)\right]$$

Using similar logic as in Lemma 4. $\frac{A_t - (1-p_t)|y}{A_t^* - (1-p_t^*)|y}$ is decreasing in y. Because the changes in wages cannot exceed the expected change in output, $p_t \mid y' \geq p_t \mid y$.

Similarly, the FOC w.r.t ϕ_{t+1} indicates that the marginal benefit of raising ϕ_{t+1} is smaller for y' than for y. Therefore, it is optimal for the firm to set $\phi_{t+1} \mid y' \leq \phi_{t+1} \mid y$.

Suppose that there exists x_0, x_0' and $x_0 < x_0'$ such that $A_1(x_0) < A_1(x_0')$. Consider an

alternative contract $\tilde{v}_1$ such that

$$\tilde{v}_1(y,x) = \begin{cases} \tilde{v}_1(x_0,x) \text{ if } y \in \{x_0, x_0'\} \\ \tilde{v}_1(y,x) \text{ if } y \notin \{x_0, x_0'\} \end{cases}$$

Because $\tilde{v}_1$ is incentive compatible at x_0, it is also incentive compatible at x_0'. □

Proof of Proposition 3

To reduce notations, we prove Proposition 3 for $t = 0$. The proof is straightforwardly extended to any $t > 0$.

First, we show that there exists $\underline{x}_0$. Suppose, for a contradiction, that at $X_0 = x_0$, $p_1(x_0) > 0$. Consider a sufficiently low value of x_0 such that $\alpha_0 g_0 - \alpha_1 \geq 0$ and $\alpha_0 g_0(A_1^0(1-\phi_1)+\phi_1) - \alpha_1(A_1(1-\phi_1)+\phi_1) \geq 0$. Consider an alternative contract $\tilde{v}_1$ such that:

$$\tilde{v}_1(x) = \begin{cases} v_1(x)/\exp(\delta) \text{ if } x \neq x_0 \\ \frac{\alpha_0 g_0(x_0)(A_1^0(x_0)(1-\phi_1(x_0))+\phi_1(x_0)) - \alpha_1(A_1(x_0)(1-\phi_1(x_0))+\phi_1(x_0))}{\alpha_0 g_0(x_0) - \alpha_1} v_1(x_0)/\exp(\delta) \text{ if } x = x_0 \end{cases}$$

Note that at $\delta = 0$, $\tilde{v}_1(\cdot)$ satisfy the IC constraint, and the executive is better off. Therefore, there exists a value $\delta > 0$ such that the executive is indifferent between $v_1(\cdot)$ and $\tilde{v}_1(\cdot)$. Specifically:

$$\delta = \log(v_1(x_0)(A_1(x_0)(1-\phi_1(x_0))+\phi_1(x_0))) - \log(\tilde{v}_1(x_0))$$

Note that $\delta > \log(\tilde{v}_1(x_0)) - \log(v_1(x_0))$. Therefore, the contract costs less than the original contract and satisfies both IC and IR constraints. When $\mathbb{E}\left[(x_1 - w_2 + \lambda_2/\lambda_1\Pi_2) - \Pi_{0,1}\right]$ is sufficiently close to 0, the cost-saving from $\tilde{v}_1$ exceeds the profit from $\mathbb{E}\left[(x_1 - w_2 + \lambda_2/\lambda_1\Pi_2) - \Pi_{0,1}\right]$. Therefore, for sufficiently low values of $\mathbb{E}\left[(x_1 + \lambda_2/\lambda_1\Pi_2) - \Pi_{0,1}\right]$, $\underline{x}_0$ exists.

Next, we show that there exists $\bar{x}_0$. Suppose, for a contradiction, that at $X_0 = x_0$, $\phi_1(x_0) > 0$. Consider a sufficiently large value of x_0 such that $\alpha_1 v_1(A_1(1-\phi_1)+\phi_1) - \alpha_0 v_1(A_1^0(1-\phi_1)+\phi_1)g_0^0 > 0$ and $\alpha_1 A_1 - \alpha_0 A_1^0 g_0^0 > 0$. Consider an alternative contract $\tilde{v}_1$ such that:

$$\tilde{v}_1(x) = \begin{cases} v_1(x)/\exp(\delta) \text{ if } x \neq x_0 \\ \frac{\alpha_1 v_1(A_1(1-\phi_1)+\phi_1) - \alpha_0 v_1(A_1^0(1-\phi_1)+\phi_1)g_0^0}{\alpha_1 A_1 - \alpha_0 A_1^0 g_0^0}/\exp(\delta) \text{ if } x = x_0 \end{cases}$$

We choose δ and $\tilde{v}_1$ such that the new contract satisfy IC and IR. Note that, by construction,

when $\delta = 0$, the new contract satisfies the IC constraint and yields strictly higher payoff to the executive. Therefore,

$$\delta = \log(v_1(x_0)(A_1(x_0)(1-\phi_1(x_0))+\phi_1(x_0))) - \log(\tilde{v}_1(x_0)A_1(x_0)) + \log(f_{X_0}(x_0)) > 0$$

Note that $\delta > \log(\tilde{v}_1(x)) - \log(v_1(x))$. Therefore, the contract costs less than the original contract and satisfies both IC and IR constraints. In addition, because $\mathbb{E}\left[\Pi_{0,1} - (x_1 - w_2 + \lambda_2/\lambda_1\Pi_2)\right] \leq 0$, the firm earns higher expected profit in $t = 2$ by setting $\phi_1 = 0$. Therefore, this is a profitable deviation. Therefore, $\bar{x}_1$ exists.

Finally, we show that $\underline{x}_0(\cdot) \neq \bar{x}_0(\cdot)$. We prove a more general result, which is that $v(\cdot), \phi(\cdot)$, and $p(\cdot)$ are continuous. Suppose, for a contradiction, that there exists x_0 and x_0' such that $x_0' - x_0 \leq \delta$ but $|v_1(x_0)(A_1(x_0)(1-\phi_1(x_0)+\phi_1(x_0)) - v_1(x_0')(A_1(x_0')(1-\phi_1(x_0')+\phi_1(x_0'))| > \Delta > 0$ for any δ sufficiently small. Consider an alternative contract $(\tilde{v}, \phi_1)$ such that

$$\tilde{v}(\boldsymbol{x}) = \begin{cases} \mathbb{E}(v(\boldsymbol{x})|X_0 \in \{x_0, x_0'\}) \\ v(\boldsymbol{x}) \text{ if } X_0 \notin \{x_0, x_0'\} \end{cases}$$

and

$$\tilde{\phi}(\boldsymbol{x}) = \begin{cases} \mathbb{E}(\phi(\boldsymbol{x})|X_0 \in \{x_0, x_0'\}) \\ \phi(\boldsymbol{x}) \text{ if } X_0 \notin \{x_0, x_0'\} \end{cases}$$

Let $\tilde{A}$ denote the continuation value from $(\tilde{v}, \tilde{\phi})$. By construction, $\mathbb{E}(A(\boldsymbol{x}) \mid X_0 \in \{x_0, x_0'\}) = \mathbb{E}(\tilde{A}(\boldsymbol{x}) \mid X_0 \in \{x_0, x_0'\})$. Note that because $A(p)$ is concave in p, $\tilde{p}_1(x_0) = \tilde{p}_1(x_0') > \mathbb{E}(p_1(x) \mid X_0 \in \{x_0, x_0'\})$. In addition, because v is convex in w $\mathbb{E}(\tilde{w}_2 \mid X_0 \in \{x_0, x_0'\}) < \mathbb{E}(\tilde{w}_2 \mid X_0 \in \{x_0, x_0'\})$. Therefore, the new contract satisfies both IC and IR constraints and generate strictly higher profit for the firm. Therefore, v and ϕ must be continuous.

C Proof of Section 4

Lemma 6. *Eq.* (10) *and* (11) *uniquely identify* α_1 *and* ρ.

Proof. Suppose that there exist (α_1, ρ) and $(\tilde{\alpha}_1, \tilde{\rho})$ that can rationalize the data, where $\alpha_1 \neq \tilde{\alpha}_1$ or $\rho \neq \tilde{\rho}$ or both. Without loss of generality, we assume that $\rho > \tilde{\rho}$. Thus, $\alpha_1 < \tilde{\alpha}_1$. This implies

$$\mathbb{E}\left[\alpha_1^{(b_t-1)^{-1}} \exp\left(-\rho w_{t+1}/b_{t+1}\right) - \tilde{\alpha}_1^{(b_t-1)^{-1}} \exp\left(-\tilde{\rho} w_{t+1}/b_{t+1}\right) \mid \boldsymbol{y}\right] = 0 \tag{59}$$

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.
Consider (59) at $(\boldsymbol{y}')$. Let $\bar{w}$ and $\underline{w}$ denote the (unconditional) support of W_{t+1}:

$$0 = \int_{\underline{w}}^{\bar{w}} \Delta(w) f_{W_t}(w \mid \boldsymbol{y}) \frac{f_W(w \mid \boldsymbol{y}')}{f_W(w \mid \boldsymbol{y})} dw \tag{60}$$

$$\implies 0 = -\int_{\underline{w}}^{\bar{w}} \int_{\underline{w}}^{w} \Delta(w') f_W(w' \mid \boldsymbol{y}) dw' \frac{\partial}{\partial w} \frac{f_w(w \mid \boldsymbol{y}')}{f_W(w \mid \boldsymbol{y})} dw \tag{61}$$

where Eq. (61) follows from integration by parts. Because $\Delta(w)$ crosses 0 only once from below, $\int_{\underline{w}}^{w} \Delta(w')\ f_w(w' \mid (\boldsymbol{y}) dw' < 0$ for any $w' < \bar{w}$. The likelihood ratio is increasing; therefore, the LHS of (61) is strictly negative. This is a contradiction. □

Proof of Proposition 4. Our identification argument proceeds as follows. Suppose that $q_\tau(\cdot)$, $\tilde{A}_\tau(\cdot)$, $g_\tau^{\tau'}(\cdot)$ are all identified for $\tau \leq t$, then $q_{t+1}(x, \boldsymbol{y})$, $\tilde{A}_{t+1}(x, \boldsymbol{y})$, and g_t^τ are identified.

Using Corollary 2, note that $\frac{\partial A_{\tau+1}^\tau}{\partial v_{t+1}}$ is an identified linear function of A_{t+1}^t for any τ. Because of the parametric assumption in Assumption 1, $g_t^t/g_t^\tau = \gamma(x_t)/\gamma(x_\tau)$. Therefore,

$$\begin{aligned} 0 = \lambda_t v_{t+1}^{-1} &- (\mu_t^{IC} + \mu_t^{IR}) \alpha_1^{(b_{t+1}-1^{-1})} \left[A_{t+1}^*(1-\phi_{t+1}) + \phi_{t+1}\right] \\ &+ \mu_t^{IC} \alpha_0^{(b_{t+1}-1^{-1})} \gamma_t \mathbb{E}_t^*(\theta) \left[A_{t+1}^t(1-\phi_{t+1}) + \phi_{t+1}\right] \\ &- \sum_{\tau=0}^{t-1} (\mu_\tau^{IC} + \mu_\tau^{IR}) \alpha_1^{(b_{\tau+1}-1)^{-1}} v_{\tau+1} \frac{\partial A_{\tau+1}^*}{\partial v_{t+1}} \\ &+ \sum_{\tau=0}^{t-1} \mu_\tau^{IC} \alpha_0^{(b_{\tau+1}-1)^{-1}} (\gamma_\tau \mathbb{E}_t^*(\theta) v_{\tau+1} \frac{\partial A_{\tau+1}^\tau}{\partial v_{t+1}}) \end{aligned} \tag{62}$$

□