Optimal Long-term Executive Contracts

Robert A. Miller* Anh Nguyen†

This version: January 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.

*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. In our model, both firms and executives initially share symmetric uncertainty about their match quality, which gradually resolves 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.

We first present suggestive evidence consistent with the presence of both moral hazard and match quality uncertainty. Specifically, we use the firm's gross return as a measure of performance and nonparametrically construct a predicted market value based solely on the executive's past performance, with this latter variable capturing the executive's historical performance. Our reduced-form results show that executive compensation is positively correlated with both current and past performance, supporting our hypothesis that firms use performance-based pay to incentivize executives and increase compensation to retain those with better match quality. Additionally, we find that current performance is negatively correlated with the probability of separation, but executives who underperform relative to shareholders' expectations face a higher risk of involuntary termination. The results further suggest that voluntary separation is an endogenous decision made by executives, who weigh the discounted future utility of remaining with the firm against the benefits of leaving.

We then construct and estimate a model that does not a priori impose a parametric functional form on the relationship between match quality and the firm's market value. In this model, the firm's objective is to maximize its market value, which is equal to the discounted sum of dividends paid to shareholders. A better matched executive is associated with a higher expected market value. To retain such executives, the firm has to balance higher executive compensation against lower dividends. The executive's action is binary—either to work or to shirk—where working incurs a cost to the executive but is the preferred

action for the firm. We assume the firm designs contracts to incentivize the executive to work until the relationship is terminated. The executive's optimal action is determined by the distribution of present compensation and the expected continuation value associated with each action. Our model accommodates both voluntary and involuntary separations. Voluntary separations occur when an executive opts for a better outside opportunity, which is independent of the match value of the relationship. Involuntary separations, on the other hand, arise as part of the firm's optimal contract design.

Our first contribution is to demonstrate that the latent equilibrium objects of the model, as well as its primitives, can be identified using observed compensation realizations and retention data. In the model, the key primitive objects are the nonpecuniary cost of exerting effort and the likelihood ratio of performance, which captures the likelihood of performance being produced under shirking versus working. Together, these two primitives determine the welfare cost of moral hazard. In addition, we treat the firing decision as an equilibrium object to be identified, enabling us to independently estimate the probability of involuntary termination without relying on existing estimates in the literature. Another object of interest is the ratchet ratio, which reflects the executive's anticipation that future compensation schedules will be influenced by today's performance, creating a ratchet effect. We define the ratchet ratio as the inverse of the future payoff under shirking relative to that under putting in effort.

Our identification argument relies on the optimality of the compensation contract and the firing frequency. Intuitively, the strength of performance-based compensation is determined by both the likelihood ratio and the ratchet ratio. Specifically, compensation is higher when the performance realization signals that the executive is more likely to achieve the outcome by putting in effort (i.e., when the likelihood ratio under shirking versus working is low) and when the executive earns a lower future informational rent by deviating today. These two ratios also influence the necessity of using involuntary termination, and an executive is more likely to be fired when his ratchet ratio is lower. Another identifying moment in the model relates to the level of compensation, which reflects the firm's desire to retain the executive as well as to minimize the ratchet effect resulting from past actions.

Our estimates suggest that firms frequently use involuntary termination to incentivize executives, and the firing probability is strictly positive for 85% of the sample. When employed as an incentive, the probability of involuntary termination can reach up to 30%. We also show that it is important to account for the ratchet effect. The average estimated ratchet ratio is 0.94, indicating that executives can earn significant informational rent by choosing

to shirk. This informational rent translates into an equivalent additional annual compensation of approximately $11 million. Furthermore, the ratchet ratio is negatively correlated with the likelihood ratio, making it more difficult for firms to incentivize executives to work through compensations alone.

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 executives and firms learn about match-specific executive quality as executive performances are realized. Our model differs from the existing models in this literature in two aspects. First, in our model, an executive's private effort choice affects the realized performance. This implies that the learning process is endogenous to the hidden action of the executive. Second, whereas in most models, the separation between a firm and a worker is always a voluntary action made by the worker, in our model, separation can be involuntary. In addition, our model identifies two distinct channels for involuntary separation: (1) deterministic involuntary separation, occurring when the executive performs worse than an outside option, and (2) probabilistic involuntary separation, used to motivate the executive.

More broadly, involuntary termination in the optimal contract between an agent and a principal can arise for several reasons: the agent has limited liability (Clementi and Hopenhayn, 2006; DeMarzo and Sannikov, 2006), the agent is risk-averse and providing insurance through the contract is too costly (Spear and Wang, 2005), or the executive's realized quality is lower than that of a potential replacement (Hermalin and Weisbach, 1998). Our paper differs from these papers in that it considers a setting in which agents are risk averse and have access to a sufficiently complete market to hedge against risks not associated with their performance pay. The incorporation of risk aversion is empirically motivated by the extensive literature that studies the role of risk in compensation contracts, such as Demsetz and Lehn (1985) and Peters and Wagner (2014).

The interaction between moral hazard and match value uncertainty is a central theme in a large body of literature on principal-agent models featuring symmetric uncertainty and imperfect monitoring (Holmström, 1999; Prat and Jovanovic, 2014; DeMarzo and Sannikov, 2016; Cisternas, 2018; Bhaskar and Mailath, 2019; Lyman, 2023). Our paper differs from the existing literature in two ways. First, we do not assume a parametric learning process. Second, unlike existing models, which often require either risk-neutral executives or firms

that can directly control consumption decisions, our model does not allow firms to manage the consumption choices of risk-averse executives. Our modeling choice is motivated by both institutional settings and the fact that executives' consumption decisions are typically unobservable in the data.

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.

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 employs an age-based classification method to determine whether a separation was voluntary or involuntary, identifying 1,398 separations as forced turnovers. For our analysis, we also adopt a more conservative approach, classifying only 623 of these cases as forced turnovers based on additional sources, including news articles, press reports, and SEC filings. For the remaining cases in our data, we remain agnostic regarding the nature of the separation.

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 2.1 presents summary statistics on executive total compensation and its components in our dataset. 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. Additionally, there is significant variation in compensation both across executives and within individual executive contracts.

Throughout the paper, we assume that a firm's market value reflects shareholders' beliefs about the executive's match quality with the firm. Accordingly, if the executive's match quality is learned over time, we would expect compensation to depend on past market values. However, without specifying a full learning model, the precise contribution of past information embedded in these market values to compensation remains unclear. To address this,

Table 2.1: Summary statistics

	Large firms		Small firms	
	Mean	SD	Mean	SD
Salary and bonus	617	518	1212	2111
Other reported payments	1083	2058	4353	8927
Change in wealth due to previous holdings	-117	1808	-360	16298
Total compensation	1677	2800	5404	19104
Gross return relative to market return	-0.06	0.21	-0.04	0.2
Executive age	52.74	7.6	52.73	6.42
Tenure duration	3.67	2.92	4.79	3.58
Proportion of female executives	0.03	0.17	0.02	0.15
Number of firms	1971		2703	
Number of executives	3669		4779	

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.

we adopt a nonparametric approach to proxy for the information contained in past market values. Specifically, we compute the first and second moments of the conditional distribution of the firm's future market value, based on its past value realizations under the same executive, using Kernel estimation. In this construction, we normalize the firm's market value by regressing its log value on industry fixed effects, unemployment rate, inflation, and the executive's age and gender, using only the residuals of this regression. Figure B.2 shows the distribution of the constructed moments. Additionally, in the presence of moral hazard, executive compensation should also be positively correlated with a per-period performance measure. For this, we use the firm's gross return as the proxy for per-period performance. The distribution of firm gross return in the data is presented in Figure B.1 in the Appendix.

Table 2.2 demonstrates that compensation is positively correlated with both per-period performance and past performance. In this regression, the proxy for past performance is the first and second moments of the predicted market value (PMV), as previously explained. A higher first moment reflects a higher expected market value based on past performance, while a higher second moment is negatively correlated with the executive's tenure. The regression results indicate that executives are more likely to receive higher compensation when the firm's gross return is higher or when the predicted market value based on their past performance is higher. These findings are consistent with our hypothesis that firms use performance-based pay to motivate executives and offer higher compensation to retain better-matched executives.

Table 2.3 explores the correlation between separation, per-period performance, and past performance, as well as whether this correlation varies depending on whether the separation

Table 2.2: Correlation between compensation and performance

	Total compensation			
	(1)	(2)	(3)	(4)
Gross return	6.2518	6.4323	8.2702	8.2716
	(1.0706)	(1.0924)	(1.2464)	(1.2534)
First moment of PMV	0.1132	0.1113	0.1098	0.1092
	(0.0148)	(0.0152)	(0.0147)	(0.0146)
Second moment of PMV	−0.0004	−0.0004	−0.0004	−0.0004
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Age of executive				−0.0442
				(0.0124)
Female executive				0.3252
				(0.1449)
Constant	1.6591	1.4271	0.6088	3.1476
	(0.1783)	(0.0741)	(0.3364)	(0.7890)
Industry FEs	No	Yes	Yes	Yes
Year FEs	No	No	Yes	Yes
Observations	34,719	34,719	34,719	34,719
Adjusted R^2	0.0350	0.0402	0.0521	0.0531

Note: Total compensation is measured in thousands of dollars. The second and third coefficients represent the correlation between total compensation and the first and second moments of the predicted market value change, respectively. Standard errors, adjusted for heteroskedasticity and clustered at the industry level, are reported in parentheses.

was likely involuntary. In the second column of Table 2.3, we use the classification method from Peters and Wagner (2014) and Jenter and Kanaan (2015). In the third column, we apply our more conservative approach to classify involuntary terminations. The results show that lower per-period performance, as measured by gross return, is associated with a higher probability of separation. When considering all separations (column (1)), past performance does not exhibit a significant correlation with separation. However, higher past performance is linked to a greater likelihood of involuntary separation (columns (2) and (3)), implying that executives who underperform relative to shareholders' expectations are more likely to be terminated involuntarily. Interestingly, the regression results indicate that the correlation between separation and performance is similar for both voluntary and involuntary separations. This finding suggests that voluntary separation is endogenous, with executives choosing to leave the firm when they anticipate lower future compensation due

Table 2.3: Correlation between separation and performance

	Separation		
	(1)	(2)	(3)
Gross return	−3.022	−3.739	−2.663
	(1.230)	(1.365)	(1.150)
Gross return x Involuntary		5.552	−5.104
		(3.046)	(3.175)
First moment of PMV	0.016	−0.031	−0.018
	(0.020)	(0.020)	(0.018)
First moment of PMV x Involuntary		0.140	0.128
		(0.055)	(0.052)
Second moment of PMV	0.002	0.002	0.003
	(0.001)	(0.001)	(0.001)
Second moment of PMV x Involuntary		0.001	−0.001
		(0.002)	(0.001)
Age of executive	0.545	0.550	0.546
	(0.062)	(0.062)	(0.062)
Female executive	−0.555	−0.816	−0.578
	(1.328)	(1.341)	(1.314)
Constant	−14.072	−14.783	−13.919
	(18.296)	(18.438)	(18.253)
Industry FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Observations	34,719	34,719	34,719
Adjusted R^2	0.039	0.039	0.039

Note: Separation is measured in percentage points. In Column (2), we use the classification of involuntary separation from Peters and Wagner (2014) and Jenter and Kanaan (2015). In Column (3), we apply our more conservative classification. Standard errors, adjusted for heteroskedasticity and clustered at the industry level, are reported in parentheses.

to poor performance today.

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. While the executives

are ex-ante identical to the firm, they possess heterogeneous match qualities that are specific to each firm-executive pairing. The match quality between the firm and an executive is unknown to both parties but positively influences the firm's gross market value. We assume that the firm and all executives are infinitely lived. Once hired, let T denote the duration of employment, and let $t \in \{0, 1, 2, \ldots, T\}$ represent each year of tenure. The firm can employ an executive for at most $\bar{T}$ years, meaning $T \leq \bar{T}$, after which the executive retires. We define $\mathcal{S}_t$ as the firm's prior belief about its market value at the beginning of period t, and $\mathcal{S}_{t+1}$ as its posterior belief after observing new information. Let $\tilde{V}_t$ denote the change in the market value of the firm in period t. We sometimes abuse notation and refer to $\{\tilde{V}_{t+1}, \mathcal{S}_t\}$ as the posterior belief $\mathcal{S}_{t+1}$.

The executive receives wage w_{t+1} at the beginning of period $t+1$ for work performed in period t and faces involuntarily termination with probability q_{t+1}. At the beginning of period $t = 0$, the executive receives a contract offer from the firm specifying wages $\{w_{t+1}|\mathcal{S}_t, \tilde{V}_{t+1}\}_{t\leq\bar{T}}$ and (involuntary) termination probabilities $\{q_{t+1}|\mathcal{S}_t, \tilde{V}_{t+1}\}_{t\leq\bar{T}}$.

3.1 Executive preferences and consumption choices

The executive derives utility from consumption c_t. In any period t, conditional on the firm not terminating the contract ($D_t = 0$), the executive can make one of three choices: (1) remain employed without exerting effort ($d_{0,t} = 1$), (2) remain employed and exert effort ($d_{1,t} = 1$), or (3) leave the firm ($d_{2,t} = 1$). Note that $d_t = (d_{0,t}, d_{1,t}, d_{2,t})$ are mutually exclusive choices. The firm does not observe the executive's past or current effort decisions. If the executive is not employed by the firm, their per-period utility is normalized to 1.

The executive's lifetime utility is given by:

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

where δ is the discount factor, ρ is the coefficient of absolute risk aversion, and T is the realized tenure duration. ϵ_t represents an idiosyncratic preference for working at the firm. We assume $\epsilon_0 = 0$, and $\{\epsilon_t\}_{t\geq 1}$ are independently distributed over time following a logistic distribution. The parameters α_0 and α_1 represent the utility costs associated with choosing $d_{0,t} = 1$ and $d_{1,t} = 1$, respectively. We assume that exerting effort is more costly than not exerting effort ($\alpha_1 > \alpha_0$), but effort has a positive impact on the firm's payoff and is the preferred choice for the firm.

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. Following the framework of Margiotta and Miller (2000), it suffices to consider a finite set of financial securities to characterize the executive's optimal financial portfolio. 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. [1]

We assume that the executive and the firm have perfect foresight about $\{a_t, b_t\}_{t \leq \bar{T}}$. Let ξ_t denote the executive's wealth in period t. Let p_t denote the probability that the executive will stay with the firm prior to ϵ_t being realized. The executive's value function, prior to ϵ_t being realized and under optimal consumption decisions, is given by:[2]

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

where A_t is equal to

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

Here, $v_t = \exp(-\rho w_t / b_t)$ represents the compensation from working, annuitized to facilitate consumption smoothing. Importantly, both p_t and A_t depend on the executive's information set at the beginning of period t, which includes the firm's belief about its market value $\mathcal{S}_t$ as well as the executive's private history of effort choices.

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.

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 the termination of the contract.

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

[2] The derivation of the optimal consumption decisions are similar to that of Margiotta and Miller (2000) and is therefore omitted.

Since the effort decision is private and not directly contractible, the firm's contract must satisfy an incentive compatibility (IC) constraint. At any time t, the IC constraint is given by:

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

where α_1 and α_0 are the utility costs of exerting and not exerting effort, respectively, as defined earlier. In constructing the IC constraint, we utilize the observation that the firm's belief about its future market value aligns with that of an executive who has consistently chosen the equilibrium action—i.e., exerting effort—in all past periods. Therefore, the executive's information set for the IC constraint includes only $\mathcal{S}_t$ and his effort choice in period t.

Participation constraints In addition, the firm must satisfy the executive's participation constraint. In period 0, since $\epsilon_0 = 0$, there is no private preference shock, and we assume that when indifferent between accepting and not accepting the offer, the executive always chooses to accept. Therefore, the firm sets the contract such that the executive is indifferent, and the participation constraint for period 0 is specified as follows:

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

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

In subsequent periods, since ϵ_t is unobserved by the firm and unbounded, the participation constraint is structured such that the executive remains with the firm with a probability p_t. Let ϵ_t^* denote the threshold at which the executive, following the equilibrium strategies $(d_{1,\tau})_{\tau<t} = 1$, is indifferent between staying with the firm and quitting.

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

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

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

where $\underline{u}(p_t|d_{1,t}=1,\mathcal{S}_t)$ represents the executive's minimum required continuation value to remain with the firm, given the probability of staying p_t, their effort choice $d_{1,t}=1$, and the firm's belief $\mathcal{S}_t$.

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

$$V_0 = \max_{(p_t,v_{t+1},q_{t+1}|\mathcal{S}_t)} \mathbb{E}_0\left[\lambda_0(\tilde{V}_1 - w_{t+1}) + \lambda_1 V_1\right] \tag{$\mathcal{P}$}$$

where

$$V_t = \lambda_t \mathbb{E}_t\left[p_t(1-q_t)\left[\tilde{V}_{t+1} - w_{t+1} + \lambda_{t+1}/\lambda_t V_{t+1}\right] + (1-p_t(1-q_t))V_{0,t}\right]$$

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

For $\tau \leq t$, let

$$g^{\tau}_{t+1} := \frac{f(\tilde{V}_{t+1}|\mathcal{S}_t, d_{0,\tau}=1)}{f(\tilde{V}_{t+1}|\mathcal{S}_t, d_{1,\tau}=1, d_{1,-\tau}=1)}$$

That is, g^{τ}_t denotes the likelihood ratio of performance in period t when the executive shirks in period τ (and choose his optimal actions in subsequent periods) relative to when the executive never shirks. We assume that $g^t_{t+1}(\cdot|\mathcal{S}_t)$ is decreasing when $\tilde{V}_1$ is sufficiently large, $\lim_{\tilde{V}_{t+1}\to\infty} g^t_{t+1} = 0$, and $\lim_{\tilde{V}_{t+1}\to\infty} g^{\tau}_{t+1} < \infty$ for $\tau < t$.

Conditional on $\mathcal{S}_t$, let λ^{IC}_t and λ^{IR}_t denote the Lagrangian multipliers associated with the IC and IR constraints in period t, respectively. Similarly, let $\lambda^{q,0}_t$ and $\lambda^{q,1}_t$ represent the Lagrangian multipliers associated with the constraints $q_t \geq 0$ and $q_t \leq 1$. Since the idiosyncratic preference shock ϵ_t is unbounded, p_t is always strictly less than 1 (i.e., $p_t \in [0,1)$). For notational convenience, we normalize λ^{IC}_t and λ^{IR}_t by dividing them by the probability of reaching $\mathcal{S}_t$, given by $\prod_{\tau=0}^{t-1} p_\tau(1-q_\tau)(1-q_t)$.

In Sections 3.3.1 and 3.3.2, we illustrate the key determinants of $(v_{t+1}, q_{t+1}, p_{t+1})$ 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 he faces a termination probability of q_1. To simplify notation, we omit $\tilde{V}_1$ and $\mathcal{S}_0$ from the conditioning variables of A_1, p_1, and V_1 whenever this does not cause confusion.

Conditional on the realization of $\mathcal{S}_1$, the FOC with respect to v_1 is expressed as follows:

$$\begin{aligned} 0 = \lambda_0 v_1^{-1} - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{(b_0-1)^{-1}} \left[(A_1|d_{1,0}=1, \tilde{V}_1, \mathcal{S}_0)(1-q_1) + q_1\right] \\ + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1^0 \left[(A_1|d_{0,0}=1, \tilde{V}_1, \mathcal{S}_0)(1-q_1) + q_1\right] \end{aligned} \tag{1}$$

The first term in Eq. (1) represents the increase in the firm's payoff when increasing v_1—i.e., lowering w_1, and the remaining terms represent its effect on the executive's IC and IR constraints in period 0. Furthermore, Eq. (1) indicates that the variation in v_1 is linked to the variation in the likelihood ratio g_1 and the ratio $\frac{A_1|d_{0,0}=1,\tilde{V}_1,\mathcal{S}_0)}{A_1(d_{1,0}=1,\tilde{V}_1,\mathcal{S}_0)}$. The latter ratio, termed the *ratchet ratio*, represents the executive anticipating changes in his future compensation schedule in response to revealed information about his match quality—i.e., a ratchet effect. A smaller ratchet ratio signifies that the executive can earn more informational rent in future discounted value by shirking today. The ratchet ratio reaches its maximum value of 1 only when (1) future compensations are independent of performance (constant wages) or (2) $\tilde{V}_1$ perfectly reveals the match value, and firms and executives do not further update their beliefs in the future.

The FOC w.r.t q_1 is given by:

$$\begin{aligned} 0 = \lambda_1 p_1 \mathbb{E}\left[V_{0,1} - (\tilde{V}_2 - w_2 + \lambda_2/\lambda_1 V_2)\right] - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{(b_0-1)^{-1}}(1 - (A_1|d_{1,0}=1))v_1 \\ + \lambda_0^{IC} g_1 \alpha_0^{(b_0-1)^{-1}}(1 - (A_1|d_{0,0}=1))v_1 - \lambda_1^{q,1} + \lambda_1^{q,0} \end{aligned} \tag{2}$$

The first term of Eq. (2) represents the change in the firm's expected market value in period 1 if the executive is replaced. The remaining terms represent the impact of increasing the probability of involuntary termination on the executive's IR and IC constraints.
Multiplying Equation (2) by $1 - q_1$ and combining it with Equation (1), which is multiplied

by v_1, results in the following:

$$\begin{aligned} 0 = \lambda_1 p_1 (1 - q_1) \mathbb{E}\left[V_{0,1} - (\tilde{V}_2 - w_2 + \lambda_2 / \lambda_1 V_2) \right] \\ + \lambda_0 - (\lambda_0^{IC} + \lambda_0^{IR}) \alpha_1^{(b_0 - 1)^{-1}} v_1 + \lambda_0^{IC} \alpha_0^{(b_0 - 1)^{-1}} g_1 v_1 + \lambda_1^{q,0} \end{aligned} \tag{3}$$

The following proposition establishes that an executive is terminated with certainty when his performance is sufficiently poor, and is never terminated involuntarily when his performance is adequately high.

Proposition 1. *There are thresholds $\bar{V}_1$ and $\underline{V}_1$, where $\underline{V}_1$ such that:*

- $\tilde{V}_1 \geq \bar{V}_1$ *if and only if* $q_1 = 0$ *(no termination).*
- $\tilde{V}_1 \leq \underline{V}_1$ *if and only if* $q_1 = 1$ *(certain termination).*
- $\underline{V}_1 < \tilde{V}_1 < \bar{V}_1$ *if and only if* $q_1 \in (0, 1)$.

The proof is provided in the Appendix. In the proof, we demonstrate that the threshold for definite termination, $\underline{V}_1$, is such that the difference in the firm's future valuation from replacing the executive equals or exceeds 0. To confirm the existence of $\bar{V}_1$, observe that as $\tilde{V}_1$ approaches infinity, g_1 tends toward zero. From Equation (3), it follows that $\lim_{\tilde{V}_1 \to \infty} \lambda_1^{q,0} > 0$. Consequently, $\lim_{\tilde{V}_1 \to \infty} q_1 = 0$. In the last step, we show that $\bar{V}_1$ is unique under the assumption that g_1^0 is monotone in $\tilde{V}_1$ for sufficiently large values of $\tilde{V}_1$. We also show that $q_1(\tilde{V}_1)$ approaches 0 from the left with a strictly negative slope, implying that $q_1(\tilde{V}_1)$ exhibits a kink point at $\bar{V}_1$. This result is later used for our identification argument and is discussed in more detail in Section 4.

Proposition 1 also highlights two distinct channels for involuntary termination. When $\tilde{V}_1 \leq \underline{V}_1$, an executive is replaced because his match value is low, and the firm benefits from replacing him with a new executive. However, when $\tilde{V}_1 > \underline{V}_1$, although the executive is on average better than his replacement, he is still replaced probabilistically. Therefore, in cases where $\tilde{V}_1 > \underline{V}_1$, involuntary termination serves as an incentive.

3.3.2 Necessary conditions for period-2 compensation and termination

To simplify notation, we omit $\tilde{V}_2$ and $\mathcal{S}_1$ from the conditioning variables of A_2, p_2, q_2, and V_2 whenever this does not cause confusion.

Conditional on the realizations of $\mathcal{S}_2$ and $\mathcal{S}_1$, the FOC w.r.t v_2 is expressed as follows:

$$\begin{aligned} 0 = \lambda_1 v_2^{-1} - \left(\lambda_1^{IR} + \lambda_1^{IC}\right) \alpha_1^{(b_1-1)^{-1}} \left[(A_2|d_{1,1}=1, \mathcal{S}_1, \tilde{V}_2, d_{1,-1}=1)(1-q_2)+q_2\right] \\ + \lambda_1^{IC} \alpha_0^{(b_1-1)^{-1}} g_2^1 \left[(A_2|d_{0,1}=1, \mathcal{S}_1, \tilde{V}_2, d_{1,-1}=1)(1-q_2)+q_2\right] \\ + (1-q_1)\left[-\left(\lambda_0^{IR}+\lambda_0^{IC}\right)\alpha_1^{(b_0-1)^{-1}} v_1 \frac{\partial (A_1|d_{1,0}=1)}{\partial v_2} + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} v_1 g_2^0 \frac{\partial (A_1|d_{0,0}=1)}{\partial v_2}\right] \end{aligned} \tag{4}$$

where

$$\frac{\partial}{\partial v_2}(A_1|\cdot, \tilde{V}_1) = \frac{(A_1 - (1-p_1)|\cdot, \tilde{V}_1)}{\mathbb{E}\left[v_2((1-q_2)A_2+q_2)|\cdot, \tilde{V}_1\right]}(1-b_1^{-1})\left[(1-q_2)(A_2|\cdot, \tilde{V}_1)+q_2\right] \tag{5}$$

Relative to the FOC w.r.t v_1 in Eq. (1), the trade-off when increasing v_2—effectively a decrease in w_2—includes additional impacts on the incentive and participation constraints of period 1. This difference represents the firm's ability to internalize the externality caused by its compensation schedule in period 1 on period-0 constraints through its commitment to a long-term contract. These externalities are determined by the likelihood ratio g_2^0 (when the executive shirks in period 0), and the ratchet ratio $\frac{A_2|d_{0,0}=1,\tilde{V}_2,\mathcal{S}_1}{A_2|d_{1,0}=1,\tilde{V}_2,\mathcal{S}_1,d_{1,-0}=1}$. Consequently, the variation in present value accounting for the possibility of being involuntarily terminated at the beginning of period 2 across $\tilde{V}_2$ arises from the variation in the likelihood ratios g_2^1 and g_2^0, and the ratchet ratios $\frac{A_2|d_{0,0}=1,\tilde{V}_2,\mathcal{S}_1}{A_2|d_{1,0}=1,\tilde{V}_2,\mathcal{S}_1,d_{1,-0}=1}$ and $\frac{A_2|d_{0,1}=1,\tilde{V}_2,\mathcal{S}_1}{A_2|d_{1,1}=1,\tilde{V}_2,\mathcal{S}_1,d_{1,-1}=1}$.

The FOC w.r.t q_2 is given by

$$\begin{aligned} 0 = \lambda_2 p_2 \mathbb{E}\left[V_{0,2} - (\tilde{V}_3 - w_3 + \lambda_3/\lambda_2 V_3)\right] - (\lambda_1^{IC}+\lambda_1^{IR})(1-(A_2|d_{1,1}=1, \tilde{V}_2, \mathcal{S}_1, d_{1,-1}=1))v_2 \\ + \lambda_1^{IC} g_2 (1-(A_2|d_{0,1}=1, \tilde{V}_2, \mathcal{S}_1, d_{1,-1}=1))v_2 - \lambda_2^{q,1} + \lambda_2^{q,0} \\ + (1-q_1)\left[-(\lambda_0^{IC}+\lambda_0^{IR}) v_1 \frac{\partial (A_1|d_{1,0}=1)}{\partial q_2} + \lambda_0^{IC} v_1 g_2^0 \frac{\partial (A_1|d_{0,0}=1)}{\partial q_2}\right] \end{aligned} \tag{6}$$

where

$$\frac{\partial}{\partial q_2}(A_1|\cdot, \tilde{V}_1) = \frac{(A_1 - (1-p_1)|\cdot, \tilde{V}_1)}{\mathbb{E}\left[v_2((1-q_2)A_2+q_2)|\cdot, \tilde{V}_1\right]}(1-b_1^{-1})\left[1-(A_2|\cdot, \tilde{V}_1)\right] v_2 \tag{7}$$

Using similar reasoning as outlined in Proposition 1, we can also establish the existence of thresholds $\underline{V}_2$ and $\bar{V}_2$ such that the executive is definitively terminated if $\tilde{V}_2 \leq \underline{V}_2$ and is

never terminated if $\tilde{V}_2 \geq \bar{V}_2$. The same argument can be extended to any period t.

Finally, the firm also sets p_1 optimally. The FOC w.r.t p_1 is given by

$$0 = \lambda_1(1-q_1)\mathbb{E}\left[(\tilde{V}_2 - w_2 + \lambda_2/\lambda_1 V_2) - V_{0,1}\right] + \lambda_1^{IR}\underline{u}'(p_1) \tag{8}$$

Discussion For period $t \geq 3$, we obtain a similar set of FOCs for v_{t+1}, p_t, and q_{t+1}. In particular, the FOC w.r.t v_{t+1} shows that the variation in v_{t+1} is determined by the variation in the likelihood ratios $\{g_{t+1}^{\tau}\}_{\tau \leq t}$ and the ratchet ratios $\{\frac{A_{t+1}|d_{0,\tau}=1}{A_{t+1}|d_{1,\tau}=1}\}_{\tau \leq t}$. We also obtain the following result, which generalizes the result in Proposition 1:

Proposition 2. *Conditional on the firm's belief $\mathcal{S}_t$, there exist thresholds $\underline{V}_{t+1}$ and $\bar{V}_{t+1}$ such that the contract is terminated with certainty if $\tilde{V}_{t+1} \leq \underline{V}_{t+1}$, and it is terminated with some probability $q_t \in (0,1)$ if and only if $\tilde{V}_{t+1} \in (\underline{V}_{t+1}, \bar{V}_{T+1})$. If $\tilde{V}_{t+1} > \bar{V}_{t+1}$, the executive is never fired—i.e., $q_t = 0$.*

4 Identification and Estimation

In this section, we establish the identification argument under the assumption that firms and executives are ex-ante homogeneous. In Section 4.1, we explain how the heterogeneity in the observed characteristics of firms and executives is incorporated into the estimation procedure.

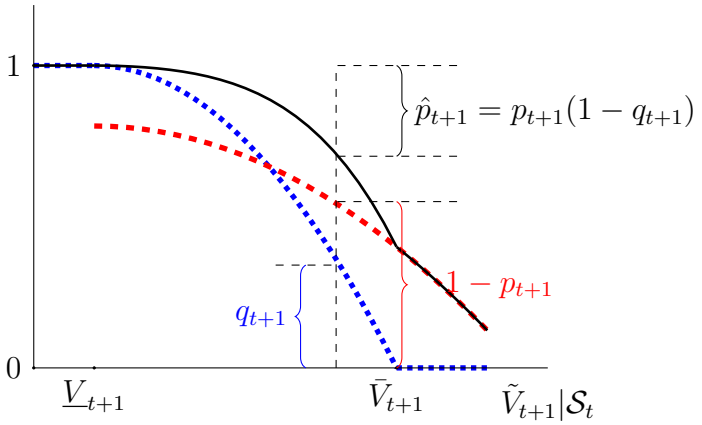
We focus on four key objects of interest. First, the set of parameters that characterize the executive's preferences: $\{\rho, \alpha_0, \alpha_1\}$. Second, the likelihood ratios, $\{g_{t+1}^{\tau}\}_{\tau \leq t}$, which represent the marginal benefit of executive effort on shareholders' payoff. Third, the ratchet ratios, $\left\{\frac{A_{t+1}|d_{0,\tau}=1}{A_{t+1}|d_{1,\tau}=1}\right\}_{\tau \leq t}$. Lastly, we aim to measure the latent probability of forced termination, q_t.

From the data, we observe the realizations of compensation (w_{t+1}), market value (V_t), gross added value $(\tilde{V}_t)$, whether an executive leaves the firm $(d_{2,t}(1-D_t))$, and the prices of financial securities (a_t, b_t, λ_t). Thus, $\mathcal{S}_t$ and the value from hiring a new executive $V_{0,t}$ are both identified, and the retention probability $\hat{p}_t = p_t(1-q_t)$ is also identified. We assume that the maximum tenure T is observed by the econometrician. Throughout this section, we assume that an executive's firing status is not directly observed. We also assume that the distribution of $\tilde{V}_{t+1}|\mathcal{S}_{t+1}, \mathcal{S}_{t-1}$ exhibits a single peak, a condition we have verified as satisfied in our data.

Our identification argument proceeds as follows. First, we identify whether q_t is interior using the following result, which is derived from Propositions 1 and 2.

Corollary 1. *Conditional on $\mathcal{S}_t$, the graph of $\hat{p}_{t+1}$ as a function of $\tilde{V}_{t+1}$ exhibits a unique kink at $\bar{V}_{t+1}$.*

Figure 4.1: Illustration of the optimal separation probability



Note: The dashed red curve represents the quitting probability, $1 - p_{t+1}$. This probability is undefined when the executive is fired with certainty ($\tilde{V}_{t+1} \leq \underline{V}_{t+1}$). The dashed blue curve represents the firing probability, q_{t+1}, which equals 1 for low outcomes ($\tilde{V}_{t+1} \leq \underline{V}_{t+1}$) and 0 when the executive's performance is sufficiently good ($\tilde{V}_{t+1} \geq \bar{V}_{t+1}$). The solid black line represents the separation probability identified from the data, given by $1 - p_{t+1}(1 - q_{t+1})$.

Figure 4.1 illustrates the result of Corollary 1. As the derivative of q_{t+1} with respect to $\tilde{V}_{t+1}$ is strictly negative as q_{t+1} approaches zero from the left, the graph of q_{t+1} exhibits a change in slope. Consequently, the separation probability $\tilde{p}_{t+1}$ also displays a kink. Since $\hat{p}_{t+1}$ is identified, the location of the kink is also identified. Therefore, whether q_t is interior is now identified.

Next, we introduce the following conjecture:

Conjecture 1. *Under the optimal contract, an executive shirks at most once during their tenure.*

Conjecture 1 relaxes the IC constraint by restricting the executive's effort choices in subsequent periods if they choose to shirk in period t. In the Appendix, we demonstrate that this result holds under a sufficient set of conditions regarding how the latent match value affects the distribution of the firm's market value. Specifically, these conditions are satisfied if: (1) an executive who has previously shirked always holds a belief about the match value—and consequently the firm's market value—that is superior in a first-order

stochastic sense to the firm's belief, and (2) motivating an executive to exert effort becomes easier when their match value is higher.

Conjecture 1 establishes a relationship between the compensation schedule at similar posteriors $\mathcal{S}_{t+1}$ but different priors $\mathcal{S}_t$. Consider a simple setting where $\tilde{V}_t$ is binary (H or L). At the start of $t = 2$, the firm's posterior belief following the history $(\tilde{V}_1 = H, \tilde{V}_2 = L)$ is identical to that after $(\tilde{V}_1 = L, \tilde{V}_2 = H)$. Additionally, the likelihood ratio g_2^1 in the first scenario matches the likelihood ratio g_2^0 in the second scenario (and vice versa). Furthermore, v_2 conditioned on $(\tilde{V}_1 = H, \tilde{V}_2 = L)$ is characterized by the same likelihood ratios and ratchet ratios (albeit with different weights) as v_2 conditioned on $(\tilde{V}_1 = L, \tilde{V}_2 = H)$.

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

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

Assumption 1 states that the variation in compensation profiles across different priors must be sufficiently rich. This assumption is satisfied if, for example, the distribution of $w_{t+1}|s$ first-order stochastic dominates $w_{t+1}|s'$, Assumption 1 is satisfied. In addition, because w_{t+1} is observed and $\mathcal{S}_t$ is identified, this assumption can be tested.

Proposition 3. *Under Assumption 1, the parameters $\{\alpha_0, \alpha_1, g_{t+1}^{\tau}, A_{t+1}(d_{0,\tau} = 1, d_{1,-\tau} = 1), A_{t+1}(d_{1,\cdot} = 1)\}$ are identified for any $t \geq 0$ and $\tau \leq t$.*

The proof of Proposition 3 is in the Appendix. First, note that at $t = T-1$, the executive retires in the next period. Thus, the continuation values $A_T(d_{0,T-1} = 1)$ and $A_T(d_{1,\cdot} = 1)$ are both equal to 1. Consequently, we show that $q_{T-1} = 0$. The retention probability in period $T - 1$ is therefore solely determined by the probability that the executive voluntarily stays with the firm and is identified. Since p_{T-1} is identified, the continuation value A_{T-1} is also identified. Conditional on $\mathcal{S}_{T-1}$, this implies that $\mathbb{E}\left[\alpha_1^{(b_T-1)^{-1}} \exp(-\rho w_T / b_T)\right]$ is identified.

Therefore, α_1 and ρ are identified from the variation in $\mathbb{E}\left[\alpha_1^{(b_T-1)^{-1}}\exp(-\rho w_T/b_T)\right]$ across different distributions of w_T due to variations in $\mathcal{S}_{T-1}$.

In the next step, we demonstrate that $\{\alpha_0, A_1, A_1^0, g_1^0\}$ can be identified. Recall that the FOC with respect to v_1 in Eq. (1) shows that the variation in v_1 is driven by the variation in g_1^0 and the variation in $\frac{A_1(d_{0,0}=0)}{A_1(d_{1,0}=1)}$. Similarly, by taking the expectation of the FOC with respect to v_2 in Eq. (4) with respect to $\tilde{V}_2$, p_1 is also determined by the same likelihood ratio and ratchet ratio.

We then consider two cases. In the first case, $\tilde{V}_1 \geq \bar{V}_1$, where $q_1 = 0$. Here, p_1 is identified and equals the retention probability $\hat{p}_1$. In the proof, we show that the two equations mentioned above are sufficient to identify both the ratchet ratio and the likelihood ratio. In the second case, $\tilde{V}_1 < \bar{V}_1$, we obtain an additional moment from the FOC with respect to q_1 (Eq. (2)). This equation alone is sufficient to identify g_1^0. Subsequently, the FOC with respect to v_1 is used to identify the ratchet ratio. Finally, the effort cost of shirking, α_0, is determined by the incentive compatibility constraint in period 0, which binds strictly.

Next, we show that $A_2, A_2^0, A_2^1, g_2^0, g_2^1$ are identified. In this step, conditional on $\mathcal{S}_2$, we identify pairs of $(\tilde{V}_1', \tilde{V}_1)$ such that it is equally likely to observe them conditional on $\mathcal{S}_0$ and $\mathcal{S}_2$. That is, $f(\tilde{V}_1|\mathcal{S}_2, \mathcal{S}_0) = f(\tilde{V}_1'|\mathcal{S}_2, \mathcal{S}_0)$. Because the distribution of $\tilde{V}_1|\mathcal{S}_0, \mathcal{S}_2$ exhibits a single peak, $\tilde{V}_1$ is uniquely pinned down for a given pair of values of $\tilde{V}_1$ and $\mathcal{S}_2$, and $\tilde{V}_1$ and V_1 yield similar likelihood ratios at $\mathcal{S}_2$. The FOCs with respect to v_2 and q_2 at $(\tilde{V}_1, \mathcal{S}_2)$ and $(\tilde{V}_1', \mathcal{S}_2)$ jointly identify two likelihood ratios $\{g_2^0, g_2^1|\mathcal{S}_1, \mathcal{S}_2\}$ and the corresponding ratchet ratios. In the final step, we generalize the argument for any period t.

4.1 Estimation

As discussed previously, when setting termination and compensation for periods $t+\tau$, the firm takes into account their impact on the executive's participation and incentive constraints in period t. However, as τ increases, these effects diminish. Therefore, to reduce the computational burden in the estimation procedure, we assume that the firm only considers the effects of termination and compensation in periods $t+1$ and $t+2$ (but not beyond $t+3$) on the executive's participation and incentive constraints in period t.

In the data, the observed characteristics for firms include industry category and firm size, which we define as the market value of the firm prior to hiring a new executive. For executives, the observed characteristics include age, whether they were hired internally or externally, and gender. We assume that these characteristics contribute to the firm's market value through a multiplicative index. To calculate this index, we perform a regression of the

log of the firm's market value against industry fixed effects, firm size, executive characteristics, and the unemployment rate to adjust for economic conditions. We then homogenize both added gross value and compensation by dividing them by this multiplicative index. This ensures that the adjusted compensation remains optimal within the firm's optimization problem.

We estimate $\mathcal{S}_t$ using a similar procedure as for the estimation of predicted market value in Section 2. We impose an assumption that these distributions can be summarized by the first two moments (the mean and the variance). In period $t = 0$, $\mathcal{S}_t$ is the distribution of the homogenized added gross value of newly hired executives, which we compute using kernel estimation. In subsequent periods, $\mathcal{S}_{t+1}$ is the conditional distribution of the homogenized $\tilde{V}_{t+1}$ conditional on the moments of $\mathcal{S}_t$, which we also estimate using kernel estimation. We choose the optimal uniform bandwidths for the moments of $\mathcal{S}_t$ and $\tilde{V}_t$ using cross validation (Racine and Li, 2004).

Next, at each $\mathcal{S}_{t+1}$ and $\mathcal{S}_{t-1}$, we estimate pairs of $(\tilde{V}_t, \tilde{V}_t')$ by solving for the following equality:

$$f(\tilde{V}_t|\mathcal{S}_{t+1}, \mathcal{S}_{t-1}) = f(\tilde{V}_t'|\mathcal{S}_{t+1}, \mathcal{S}_{t-1})$$

and empirically verify that it has a unique root.

Due to the small sample size, we opt not to estimate the likelihood ratio and ratchet ratio nonparametrically. Instead, we parameterize these ratios as spline functions of the moments of the prior beliefs ($\mathcal{S}_t$), the realized added gross value ($\tilde{V}_{t+1}$), and bond price b_{t+1}. We assume similar spline parameterizations for the location of the kink point ($\bar{V}_{t+1}$) and the firing probability q_t when it is positive.

The parameters to be estimated include those of the spline functions for g_{t+1}^t, $\frac{A_{t+1}(d_{0,t}=1)}{A_{t+1}(d_{1,t}=1)}$, $\bar{V}_t$, and q_t as well as the parameters that characterize executive preferences $(\alpha_0, \alpha_1, \rho)$. We estimate the parameters related to $\bar{V}_t$ separately from the rest of the parameters using the fact that $\hat{p}_t$ exhibits a kink point at $\bar{V}_t$; therefore, the parameters characterizing $\bar{V}_t$ maximizes the difference between the left-derivative and the right-derivative at $\bar{V}_t$. We estimate the remaining parameters using GMM, where the moments are the FOCs with respect to q_{t+1}, v_{t+1}, and p_t, the predicted retention probability maximizing the likelihood of the observed retention pattern in the data, the complementarity slackness conditions of the IC and IR constraints, and the statistical property that $\mathbb{E}(g_t^{t+1}|\mathcal{S}_t) = 1$.

4.2 Estimation results

Table 4.2 presents the parameter estimates for executive preferences and involuntary turnover. The estimated risk aversion (ρ) indicates that the average certainty equivalent compensation is \$4.37 million, which is \$80,000 less than the estimated average compensation of \$4.45 million. Figure B.3 in the Appendix illustrates the distribution of estimated certainty equivalent compensation in the data. The difference in the magnitudes of α_0 and α_1 suggests that, to maintain the same level of utility as when working, the average certainty equivalent compensation while shirking is only \$4.18 million.

Table 4.2: Summary of estimates

Risk aversion	0.9
Non-pecuniary cost of shirking (α_0)	0.79
Non-pecuniary cost of working (α_1)	0.95
Average probability of involuntary termination q (when positive)	0.0031
Average frequency of positive involuntary termination	0.85

Table 4.2 also suggests that firms frequently use involuntary termination to incentivize executives against moral hazard. The estimated firing probability q is positive for 85% of the sample, and the predicted firing probability can be as high as 32.9%. Table 4.3 also suggests that our measure of (probabilistic) involuntary termination is positively correlated with other measures of involuntary termination (as discussed in Section 2) and is also positively correlated with realized separation.

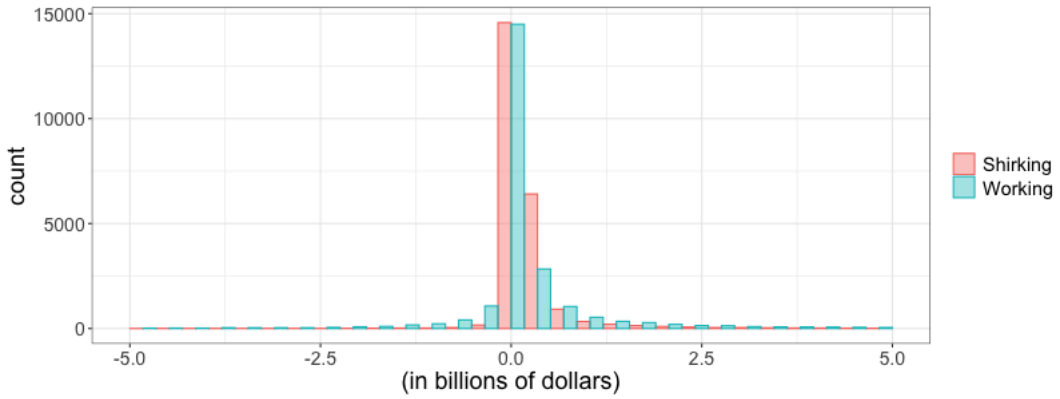
Table 4.3: Correlation between q and other classifications of involuntary turnover

	Estimated q (in percentage point)	
Involuntary termination (*)	0.032	
	(0.014)	
Involuntary termination (**)		0.088
		(0.019)
Separation	0.060	0.060
	(0.014)	(0.014)
Constant	0.254	0.252
	(0.005)	(0.005)

Note: The first measure of involuntary termination follows the classification by Peters and Wagner (2014) and Jenter and Kanaan (2015), while the second measure uses our more conservative classification.

To assess the magnitude of the likelihood ratio g, we computed the implied distribution of firms' yearly gross value changes for scenarios where an executive works versus when he shirks, as illustrated in Figure 4.4. The average yearly gross value change observed in the data is \$312.63 million, which, according to our model, corresponds to the distribution of annual performance when the executive exerts effort. In contrast, our estimates show that shirking leads to a lower average gross value change of \$283.12 million. Additionally, the distribution of annual performance under shirking is more concentrated around zero, suggesting that shirking exposes executives to less risk.

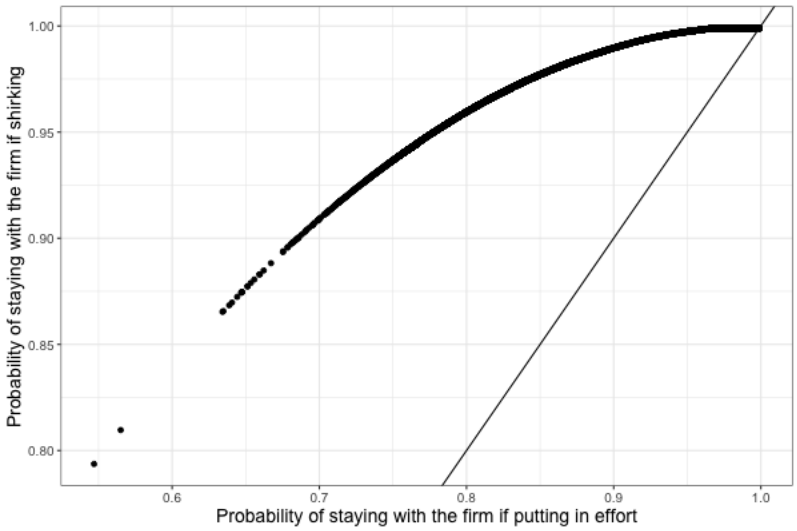
Figure 4.4: Gross value change comparison between shirking and working



Finally, the average ratchet ratio is estimated at 0.94, with a range of 0.81 to 0.98. This translates into an equivalent additional annual compensation of approximately \$11 million. Figure 4.5 illustrates the magnitude of the implied ratchet effect. The x-axis represents the probability that an executive voluntarily stays with the firm in the next period if their performance was achieved by working, while the y-axis represents this probability under shirking, conditional on the same per-period performance and shareholders' prior beliefs. These probabilities are positive functions of the continuation value A associated with each action. As future compensation increases, executives are more likely to remain with the firm, and the difference between the continuation values of working and shirking diminishes, causing the ratchet ratio to approach 1. Conversely, when low future compensation reduces the likelihood of staying, executives earn a higher future informational rent, resulting in a ratchet ratio below 1. For example, if an executive is likely to stay with the firm voluntarily with a probability of 0.7 under working, this probability increases to 0.91 under shirking. Figure B.4 in the Appendix shows that the ratchet ratio is negatively correlated with the likelihood ratio, highlighting the difficulty of motivating executives to work through

compensation alone.

Figure 4.5: Estimated ratchet ratio



Note: The x-axis represents the probability that an executive will choose to stay with the firm in period $t+1$ if he had not shirked previously, while the y-axis represents the probability that an executive will choose to stay in period $t+1$ if he had shirked in period t. Each point on the graph represents the computed ratchet effect at each observation in the data. The black line indicates the 45-degree line.

5 Counterfactuals

(To be added)

6 Conclusion

In this paper, we develop a structural model of long-term contracting to examine the interaction between moral hazard and uncertainty in match value between firms and executives. Our model shows that, in the optimal contract, firms rely on both performance-based compensation and involuntary termination to incentivize executives to take firm-optimal actions. The estimates suggest that performance-based compensation alone is less effective as executives anticipate future ratcheting of compensation, which incentivizes them to take non-optimal actions today in exchange for future informational rent. Our findings indicate that this ratchet effect is substantial. In addition, we provide a novel estimate of the extent

of involuntary termination that does not rely on assumptions about the correlation between executive demographics and termination likelihood. Using counterfactual exercises, we decompose welfare losses due to moral hazard and match value uncertainty. We also quantify the profit gains firms achieve by committing to a long-term contract. Results from these counterfactual exercises will be included in a future version of the paper.

References

Antonovics, K. and L. Golan (2012): "Experimentation and job choice," *Journal of Labor Economics*, 30, 333–366.

Atal, J. P., H. Fang, M. Karlsson, and N. R. Ziebarth (2020): "Long-term health insurance: Theory meets evidence," Tech. rep., National Bureau of Economic Research.

Bhaskar, V. and G. J. Mailath (2019): "The curse of long horizons," *Journal of Mathematical Economics*, 82, 74–89.

Cisternas, G. (2018): "Two-sided learning and the ratchet principle," *The Review of Economic Studies*, 85, 307–351.

Clementi, G. L. and H. A. Hopenhayn (2006): "A theory of financing constraints and firm dynamics," *The Quarterly Journal of Economics*, 121, 229–265.

DeMarzo, P. M. and Y. Sannikov (2006): "Optimal security design and dynamic capital structure in a continuous-time agency model," *The journal of Finance*, 61, 2681–2724.

——— (2016): "Learning, termination, and payout policy in dynamic incentive contracts," *The Review of Economic Studies*, 84, 182–236.

Demsetz, H. and K. Lehn (1985): "The structure of corporate ownership: Causes and consequences," *Journal of political economy*, 93, 1155–1177.

Fudenberg, D., B. Holmstrom, and P. Milgrom (1990): "Short-term contracts and long-term agency relationships," *Journal of economic theory*, 51, 1–31.

Gayle, G.-L., L. Golan, and R. A. Miller (2015): "Promotion, turnover, and compensation in the executive labor market," *Econometrica*, 83, 2293–2369.

Gayle, G.-L. and R. A. Miller (2009): "Has moral hazard become a more important factor in managerial compensation?" *American Economic Review*, 99, 1740–69.

——— (2015): "Identifying and testing models of managerial compensation," *The Review of Economic Studies*, 82, 1074–1118.

Ghili, S., B. Handel, I. Hendel, and M. D. Whinston (2021): "Optimal long-term health insurance contracts: characterization, computation, and welfare effects," .

Hendel, I. and A. Lizzeri (2003): "The role of commitment in dynamic contracts: Evidence from life insurance," *The Quarterly journal of economics*, 118, 299–328.

Hermalin, B. E. and M. S. Weisbach (1998): "Endogenously chosen boards of directors and their monitoring of the CEO," *American Economic Review*, 96–118.

Hofmann, A. and M. Browne (2013): "One-sided commitment in dynamic insurance contracts: Evidence from private health insurance in Germany," *Journal of Risk and Uncertainty*, 46, 81–112.

Holmström, B. (1999): "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 66, 169–182.

Jenter, D. and 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.

Racine, J. and Q. Li (2004): "Nonparametric estimation of regression functions with both categorical and continuous data," *Journal of Econometrics*, 119, 99–130.

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 Proofs

A.1 Proof of Proposition 1

Notice that, from Eq. (1):

$$-(\lambda_0^{IC}+\lambda_0^{IR})(1-(A_1|d_{1,0}=1,\mathcal{S}_1))v_1+\lambda_0^{IC}g_1(1-(A_1|d_{0,0}=1,\mathcal{S}_1))v_1 \tag{9}$$

$$\geq -\lambda_0\frac{1-(A_1|d_{1,0}=1,\mathcal{S}_1)}{(1-q_1)(A_1|d_{1,0}=1,\mathcal{S}_1)+q_1}\geq 0 \tag{10}$$

where the inequality follows from

$$\frac{1-(A_1|d_{0,0}=1,\mathcal{S}_1)}{1-(A_1|d_{1,0}=1,\mathcal{S}_1)}\geq 1\geq \frac{(1-q_1)(A_1|d_{0,0}=1,\mathcal{S}_1)+q_1}{(1-q_1)(A_1|d_{1,0}=1,\mathcal{S}_1)+q_1} \tag{11}$$

and $A_1|(\cdot) \leq 1$.

Therefore, combining with (2), we establish that if $\lambda_1 p_1 \mathbb{E}\left[V_{0,1} - (\tilde{V}_2 - w_2 + \lambda_2/\lambda_1 V_2)\right] \geq 0$, then $\lambda_1^{q,1} > 0$.

Next, we show that at $\bar{V}_1$, $\lambda_1^{q,0} > 0$. Notice that from the FOC w.r.t p_1 in (8), an executive is retained with some probability only if keeping the executive yields weakly higher future payoff. Therefore, conditional on $q_1 < 1$:

$$\mathbb{E}\left[V_{0,1} - (\hat{V}_2 - w_2 + \lambda_2/\lambda_1 V_1)\right] < 0$$

which implies

$$-(\lambda_0^{IC}+\lambda_0^{IR})\alpha_1^{(b_0 01)^{-1}}(1-(A_1|d_{1,0}=1,\mathcal{S}_1))v_1+\lambda_0^{IC}g_1\alpha_0^{(b_0-1)^{-1}}(1-(A_1|d_{0,0}=1,\mathcal{S}_1))v_1+\lambda_1^{q,0}>0$$

Suppose, for a contradiction, that $\lambda_1^{q,0} = 0$ at $\tilde{V}_1 = \bar{V}$. Therefore, when $q_1 \in [0, 1)$

$$\begin{aligned}(\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}}(A_1|d_{1,0}=1,\mathcal{S}_1) - \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}}(A_1|d_{0,0}=1,\mathcal{S}_1)\\ > (\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}} - \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1 \qquad (12)\end{aligned}$$

Combined this with (1), for $q \in (0, 1)$:

$$\begin{aligned}\lambda_0 - (\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}}(A_1|d_{1,0}=1,\mathcal{S}_1) + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}}(A_1|d_{0,0}=1,\mathcal{S}_1)\\ < 0 < \lambda_0 - (\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}} + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1 \qquad (13)\\ \implies \lim_{\tilde{V}_1 \to \bar{V}_1} \lambda_0 - (\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}}(A_1|d_{1,0}=1,\mathcal{S}_1) + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}}(A_1|d_{0,0}=1,\mathcal{S}_1)\\ < 0 < \lim_{\tilde{V}_1 \to \bar{V}_1} \lambda_0 - (\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}} + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1 \qquad (14)\end{aligned}$$

However, this is a contradiction, as $0 = \lim_{\tilde{V}_1 \to \bar{V}_1} \lambda_0 - (\lambda_0^{IR} + \lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}}(A_1|d_{1,0}=1,\mathcal{S}_1) + \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}}(A_1|d_{0,0}=1,\mathcal{S}_1)$ (Eq. 1). Therefore, $\lambda_1^{q,1} > 0$ at $\tilde{V}_1 = \bar{V}_1$. This implies that q_1 exhibits a kink point at $\tilde{V}_1$.

Lastly, we show that $\tilde{V}_1$ is unique, which follows from the assumption that g_1^0 is decreasing in $\tilde{V}_1$. As $\tilde{V}_1$ increases, $\mathbb{E}\left[V_{0,1} - (\hat{V}_2 - w_2 + \lambda_2/\lambda_1 V_1)\right]$ is more negative, and g_1^0 also decreases. Therefore, $\lambda_1^{q,0} > 0$ for $\tilde{V}_1 > \bar{V}_1$. □

A.2 Proof of Proposition 3

We organize the proof of this proposition into several lemmas.

Lemma A.1. *α_1 and ρ are identified.*

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

$$\underline{u}(p_t|\mathcal{S}_t^*) = \mathbb{E}\left(\alpha_1^{(b_t-1)^{-1}} \exp\left(-\rho w_{t+1}/b_{t+1}\right)\right) \tag{15}$$

$$\underline{u}(p_t|\mathcal{S}_t^*) = \mathbb{E}\left(\tilde{\alpha_1}^{(b_t-1)^{-1}} \exp\left(-\tilde{\rho} w_{t+1}/b_{t+1}\right)\right) \tag{16}$$

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

$$\mathbb{E}\left[\alpha_1^{(b_t-1)^{-1}} \exp\left(-\rho w_{t+1}/b_{t+1}\right) - \tilde{\alpha_1}^{(b_t-1)^{-1}} \exp\left(-\tilde{\rho} w_{t+1}/b_{t+1}\right) \middle| \mathcal{S}_{t-1}^*, \tilde{V}_t\right] = 0 \tag{17}$$

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

Let $f_w(\cdot|(\mathcal{S}_{t-1}^*, \tilde{V}_t)$ denote the distribution of compensation. Under Assumption 1, there exists s, s' such that $\frac{f_w(\cdot|(\mathcal{S}_t^*=s')}{f_w(\cdot|(\mathcal{S}_t^*=s)}$ is monotone. Let $\bar{w}$ denote the upper bound of the support of $w_{t+1}|\mathcal{S}_{t+1} \in \{s, s'\}$. WLOG, we assume that the likelihood ratio is decreasing, which implies that $\bar{w}$ is the upper bound of the support of $w_{t+1}|\mathcal{S}_{t+1} = s'$. Consider (17) at $(\mathcal{S}_t^* = s')$:

$$0 = \int_{\underline{w}}^{\bar{w}} \Delta(w) f_w(w|(\mathcal{S}_{t-1}^*, \tilde{V}_t) \frac{f_w(w|(\mathcal{S}_{t-1}^*, \tilde{V}_t')}{f_w(w|(\mathcal{S}_{t-1}^*, \tilde{V}_t)} dw \tag{18}$$

$$\implies 0 = -\int_{\underline{w}}^{\bar{w}} \int_{\underline{w}}^{w} \Delta(w') f_w(w'|(\mathcal{S}_{t-1}^*, \tilde{V}_t) dw' \frac{\partial}{\partial w} \frac{f_w(w|(\mathcal{S}_{t-1}^*, \tilde{V}_t')}{f_w(w|(\mathcal{S}_{t-1}^*, \tilde{V}_t)} dw \tag{19}$$

where (19) uses integration by parts and (17). Because $\Delta(w)$ crosses 0 only once from below, $\int_{\underline{w}}^{w} \Delta(w')\ f_w(w'|(\mathcal{S}_{t-1}^*, \tilde{V}_t) dw' < 0$ for any $w' < \bar{w}$. The likelihood ratio is decreasing; therefore, the LHS of (19) is strictly positive. This is a contradiction. □

Lemma A.2. *For any t, $A_t - (1 - p_t)$ is increasing in p_t.*

Proof of Lemma A.2. Let ϵ_t^* denote the threshold of ϵ_t at which the executive is indifferent

between staying and (voluntarily) leaving. Therefore, $\epsilon_t^* = F_\epsilon^{-1}(1-p_t)$.

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

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

Lemma A.3. $\{\alpha_0, g_1, q_1, A_1(d_{1,0}=1), A_1(d_{0,0}=1)\}$ *are identified.*

Proof. We omit the conditioning belief $\mathcal{S}_0$ for notational brevity.
Taking expectation of Eq. (1) multiplied by v_1 over v_{t+1} implies $\lambda_0^{IR} = \lambda_0$. Taking the limit of Eq. (1) as $\tilde{V}_1 \to \infty$ implies

$$0 = \lambda_0 \lim_{\tilde{V}_1\to\infty} v_1^{-1} - (\lambda_0^{IC}+\lambda_0^{IR})\alpha_1^{(b_0-1)^{-1}} \lim_{\tilde{V}_1\to\infty} A_1(d_{1,0}=1)$$

where we have utilized the assumption $\lim_{\tilde{V}_1\to\infty} g_1 = 0$ and the result from Proposition 1 that $\lim_{\tilde{V}_1\to\infty} q_1 = 0$. Because α_1 and ρ are identified, $\lim_{\tilde{V}_1\to\infty} v_1^{-1}$ is identified and equal to the annuitized compensation at the upper bound of the wage distribution. Similarly, $\lim_{\tilde{V}_1\to\infty}$ is identified. Therefore, λ_0^{IC} is identified.

Now, we consider $\tilde{V}_1 > \bar{V}_1$, which implies that $q_1 = 0$, and $p_1 = \hat{p}_1$. Therefore, $A_1(d_{1,0}=1)$ is identified.
Taking expectation of (4) over $\tilde{v}_2$ multiplied by v_2 yields:

$$\begin{aligned}0 =& \lambda_1 - \lambda_1^{IR}\underline{u}(p_1) \\ &+(1-q_1)\left[-\left(\lambda_0^{IR}+\lambda_0^{IC}\right)\alpha_1^{(b_0-1)^{-1}} v_1(A_1(d_{1,0}=1)-(1-p_1))\right] \\ &+(1-q_1)\left[\lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} v_1(A_1(d_{0,0}=1)-(1-p_1|d_{0,0}=1))g_1\right] \qquad (20)\end{aligned}$$

Because (8) identifies λ_1^{IR}, $\alpha_0^{(b_0-1)^{-1}} v_1(A_1(d_{0,1}=1)-(1-p_1|d_{0,1}=1))g_1(p_1|d_{0,1}=1)$ is thus identified. From Lemma A.2, $p_1|d_{0,0}=1$ is an identified decreasing function of $\alpha_0^{(b_0-1)^{-1}} g_1$. On the other hand, (1) also identifies $p_1|d_{0,0}=1$ ($A_1|d_{0,0}=1$) as an increasing (decreasing) function of $\alpha_0^{(b_0-1)^{-1}} g_1$. Therefore, $p_1|d_{0,0}=1$ is identified, and $\alpha_0^{(b_0-1)^{-1}} g_1$ is identified..

We now consider the case in which $\tilde{V}_1 < \bar{V}_1$. In this case, $q_1 \in (0,1)$. Eq. (3) thus identifies $\alpha_0^{(b_0-1)^{-1}} g_1$. Therefore, combined with the result from the previous paragraph, $\mathbb{E}(\alpha_0^{(b_0-1)^{-1}} g_1)$

is identified, and this expectation is equal to $\alpha_0^{(b_0-1)^{-1}}$. Therefore, α_0 is identified. Subtracting Eq. (20) from Eq. (1):

$$\begin{aligned} 0 = \lambda_1 - \lambda_1^{IR}\underline{u}(p_1) + \left(\lambda_0^{IR} + \lambda_0^{IC}\right)\alpha_1^{(b_0-1)^{-1}} v_1 \left(q_1 + (1-q_1)(1-p_1(d_{1,0}=1))\right) \\ -\lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} v_1 \left(q_1 + (1-q_1)(1-p_1|d_{0,0}=1)\right) g_1 \qquad (21) \\ \implies 0 = \lambda_1 - \lambda_1^{IR}\underline{u}(p_1) + \left(\lambda_0^{IR} + \lambda_0^{IC}\right)\alpha_1^{(b_0-1)^{-1}} v_1 \left(1-\hat{p}_1\right) \\ -\lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} v_1 \left(1-(1-q_1)(p_1|d_{0,0}=1)\right) g_1 \qquad (22) \end{aligned}$$

Using the assumption that ϵ follows a logistic distribution, we obtain:

$$\underline{u}(p,b) = \frac{1-p}{p}(b-1)^{-1}$$

Therefore, $\frac{\underline{u}(p,b)}{\underline{u}'(p,b)} = -p(1-p)(b-1)$. From (8),

$$\lambda^{IR}\underline{u}(p_1,b_1) = \lambda_1\hat{p}_1\mathbb{E}\left[(\tilde{V}_2 - w_2 + \lambda_2/\lambda_1 V_2) - V_{0,1}\right](1-p_1)(b_1-1)$$

Combine the above equation with Eq. (22) implies that $(1-q_1)p_1(d_{0,0}=1)$ is an identified decreasing function of $p_1(d_{1,0}=1)$, which further implies that $p_1(d_{0,0}=1)$ is an identified decreasing function of $p_1(d_{1,0}=1)$.
From (1):

$$\begin{aligned} 0 = \lambda_0 v_1^{-1} - (\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{(b_0-1)^{-1}} \left[(A_1|d_{1,0}=1,\mathcal{S}_1)(1-q_1) + q_1\right] \\ +\lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1 \left[(A_1|d_{0,0}=1,\mathcal{S}_1)(1-q_1) + q_1\right] \qquad (23) \\ \implies \left[(\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{(b_0-1)^{-1}}(A_1|d_{1,0}=1,\mathcal{S}_1) - \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1((A_1|d_{0,0}=1,\mathcal{S}_1)\right] \\ = \lambda_0 v_1^{-1}(1-q_1)^{-1} - \left[(\lambda_0^{IC} + \lambda_0^{IR})\alpha_1^{(b_0-1)^{-1}} - \lambda_0^{IC}\alpha_0^{(b_0-1)^{-1}} g_1\right] q_1(1-q_1)^{-1} \qquad (24) \end{aligned}$$

Because $1-q_1 = \frac{\hat{p}}{p_1(d_{1,0}=1)}$, the RHS of Eq. 24 is increasing in $p_1(d_{1,0}=1)$, whereas the LHS is decreasing in $p_1(d_{1,0}=1$. Therefore, $p_1(d_{1,0}=1)$ is uniquely identified by Eq. 24. Thus, $p_1(d_{0,0}=1)$ is also identified. □

Lemma A.4. $\{g_2(d_{0,1}=1)A_2(d_{0,1}=1), g_2(d_{0,0}=1)A_2(d_{0,0}=1)\}$ *are identified.*

Proof. From the proof of Lemma A.3, λ_1^{IR} is identified. We now show that λ_1^{IC} is identified. Because $\lim_{\tilde{V}_2\to\infty} f(\mathcal{S}_1|\mathcal{S}_2) = 0$, and $\lim_{\tilde{V}_2\ rightarrow\infty} g_2^0 < \infty$, the limit of (4) as $\tilde{V}_2 \to \infty$ is

given by:

$$0 = \lambda_1 v_2^{-1} - (\lambda_1^{IR} + \lambda_1^{IC})\alpha_1^{(b_1-1)^{-1}} A_2(d_{1,1} = 1, \mathcal{S}_1)$$

where we have also established that $\lim_{\tilde{V}_2 \to \infty} q_2 = 0$. Therefore, λ_1^{IC} is identified.
Now, consider $\tilde{\mathcal{S}}_1$ such that $f(\tilde{\mathcal{S}}_1, \mathcal{S}_2|\mathcal{S}_0)$ is the same as $f(\mathcal{S}_1, \mathcal{S}_2|\mathcal{S}_0)$, and $g_2^0|\mathcal{S}_1$ is the same as $g_2^1|\tilde{\mathcal{S}}_1$ (and vice versa). For notational convenience, the Lagrangian multipliers, compensation, and termination probability associated with $\tilde{\mathcal{S}}_1$ are denoted with a tilde. Because contracts at period t depend only on $\mathcal{S}_t$ and $\mathcal{S}_{t+1}$, the on-path continuation value A_2 depends only on $\mathcal{S}_2$ (but not $\mathcal{S}_1$).
We consider three separate cases.

Case 1 In the first case, if $q_2 = \tilde{q}_2 = 0$, the FOC w.r.t v_2 at $\mathcal{S}_1$ and $\tilde{\mathcal{S}}_1$ are sufficient to identify $A_2 g_2|d_{0,1} = 0$ and $A_2 g_2|d_{0,0} = 0$.

Case 2 Now, consider the case where $\tilde{q}_2 = 0$ but $q_2 \in (0, 1)$. Because $p_2 = \tilde{p}_2$ as both are functions of only $\mathcal{S}_2$, and $\tilde{p}_2$ is already identified and is equal to $\hat{p}_2$, q_2 is identified. Next, Subtracting the FOC w.r.t q_2 with q_2 from the FOC w.r.t v_2 multiplied by v_2 yields:

$$\begin{aligned}
0 = \lambda_1 - \left(\lambda_1^{IR} + \lambda_1^{IC}\right) v_2 \alpha_1^{(b_1-1)^{-1}} (A_2|d_{1,1} = 1, \mathcal{S}_2) \\
+\lambda_1^{IC} \alpha_0^{(b_1-1)^{-1}} g_2^1 (A_2|d_{0,1} = 1, \mathcal{S}_1) \\
-(1-q_1) f(\mathcal{S}_1|\mathcal{S}_2) \Bigg[\left(\lambda_0^{IR} + \lambda_0^{IC}\right) \alpha_1^{(b_0-1)^{-1}} v_1 \frac{A_1 - (1 - (p_1|d_{1,0} = 1, \mathcal{S}_1)}{\mathbb{E}((1-q_1)A_1 + q_1|d_{1,0} = 1, \mathcal{S}_1} (A_2|d_{1,0} = 1) \\
-\lambda_0^{IC} \alpha_0^{(b_0-1)^{-1}} v_1 g_2^0 \frac{A_1 - (1 - (p_1|d_{0,0} = 1, \mathcal{S}_1)}{\mathbb{E}((1-q_1)A_1 + q_1|d_{0,0} = 1, \mathcal{S}_1} (A_2|d_{0,0} = 1) \Bigg] \\
-\lambda_2 (p_2 - \hat{p}_2) \mathbb{E}\left[V_{0,2} - (\tilde{V}_3 - w_3 + \lambda_3/\lambda_2 V_3) \right]
\end{aligned} \tag{25}$$

Therefore, Eq. (25) and the FOC w.r.t v_2 at $\mathcal{S}_1$ jointly identifies $\{g_2(d_{0,1} = 1)A_2(d_{0,1} = 1), g_2(d_{0,0} = 1)A_2(d_{0,0} = 1)\}$.

Case 3 Now, consider the case where q_2 and $\tilde{q}_2$ are both interior. Adding the FOC w.r.t q_2 multiplied by $(1 - q_2)$ to the FOC w.r.t v_2 multiplied by v_2 yield:

$$0 = \lambda_2 \hat{p}_2 \mathbb{E}\left[V_{0,2} - (\tilde{V}_3 - w_3 + \lambda_3/\lambda_2 V_3) \right] + \lambda_1^{IC} g_2^1 v_2 +$$

$$(1-q_1)f(\mathcal{S}_1|\mathcal{S}_2)\lambda_0^{IC}v_1 g_2^0 \frac{A_1-(1-p_1)|d_{0,0}=1}{\mathbb{E}\left[v_2(1-q_2)A_2+q_2|d_{0,0}=1\right]}(1-b_1^{-1})v_2 \tag{26}$$

The only unknown terms in the above equation are g_2^1 and g_2^0. Similarly, using the combined FOC w.r.t $\tilde{q}_2$ and $\tilde{v}_2$, we obtain 2 linear equations with 2 unknowns (g_2^1 and g_2^0). Therefore, g_2^1 and g_2^0 are identified.

Recall Eq. (25). The RHS of the above equation is increasing in p_2. We obtain a similar equation at $\tilde{S}_1$. Therefore, conditional on p_2, $p_2|d_{0,0}=1$ and $p_2|d_{1,0}=1$ are both identified decreasing functions of p_2.

Finally, if there exists $\mathcal{S}_1^*$ such that the firing probability at $\mathcal{S}_2$ conditional on $\mathcal{S}_1^*$ is equal to 0, p_2 is identified from the observed retention probability at this belief. Alternatively, if such state does not exist, conditional on $\mathcal{S}_2$, taking the expectation of the FOC w.r.t v_3 multiplied by v_3 yields:

$$0=\lambda_2-\lambda_2^{IR}\underline{u}(p_2)+\mathbb{E}\left[(1-q_2)\left[-(\lambda_1^{IR}+\lambda_1^{IC})\alpha_1^{(b_1-1)^{-1}}v_2(A_2(d_{1,1}=1)-(1-(p_2|d_{1,1}=1))\right]|\mathcal{S}_2\right] \tag{27}$$

$$-\mathbb{E}\left[(1-q_2)\left[\lambda_1^{IC}\alpha_0^{(b_1-1)^{-1}}v_2(A_2(d_{0,1}=1)-(1-p_2|d_{0,1}=1)\right]g_2^1|\mathcal{S}_2\right] \tag{28}$$

$$+\mathbb{E}\left[(1-q_1)(1-q_2)\left[-(\lambda_0^{IR}+\lambda_0^{IC})\alpha_1^{(b_0-1)^{-1}}v_1\frac{\partial A_1|d_{1,0}=1}{\partial A_2|d_{1,0}=1}v_2(A_2(d_{1,0}=1)-(1-p_2|d_{1,0}=1))\right]|\mathcal{S}_2\right] \tag{29}$$

$$+\mathbb{E}\left[(1-q_1)(1-q_2)\left[\lambda_0^{IC}\alpha_1^{(b_0-1)^{-1}}v_1\frac{\partial A_1|d_{0,0}=1}{\partial A_2|d_{0,0}=1}v_2(A_2(d_{0,0}=1)-(1-p_2|d_{0,0}=1))g_2^0\right]|\mathcal{S}_2\right] \tag{30}$$

As shown in the proof of Lemma A.3, $\lambda_2^{IR}\underline{u}(p_2)$ is an identified decreasing function of p_2. Because $p_2|d_{0,0}=1$ and $p_2|d_{1,0}=1$ are identified decreasing function of p_2, combined with Lemma A.1, the RHS of (30) is an increasing function of p_2. Therefore, p_2, $p_2|d_{0,1}=0$ and $p_2|d_{0,0}=0$ are identified. □

Proof of Proposition 3. Using induction, it's straightforward to see that $\{(A_t g_t|d_{0,\tau}=0)\}_{\tau<t}$ are identified for any t. At $t=T$, $(A_T|d_{0,\tau}=0)=1$, therefore, $\mathbb{E}(A_T g_T|d_{0,\tau}=0)=(g_{T-1}|d_{0,\tau}=0)$ is identified, and $(A_{T-1}|d_{0,\tau})$ is also identified. The same argument applies to all previous periods. □

B Additional tables and figures

Figure B.1: Distribution of gross return

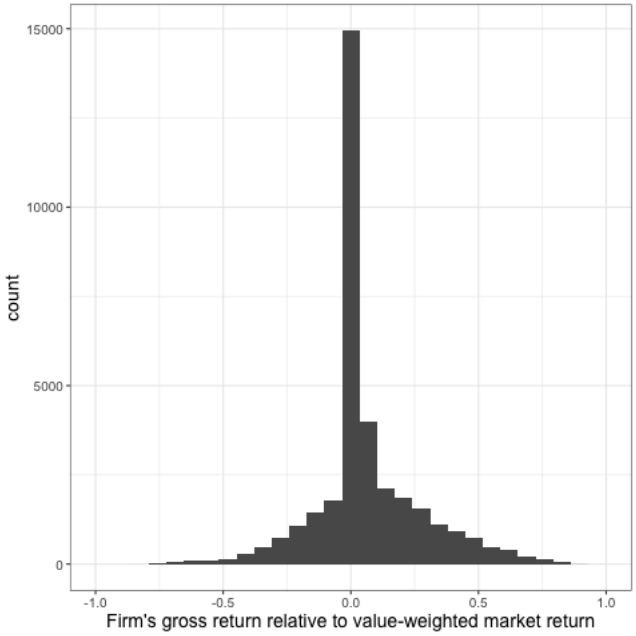


Figure B.2: Distribution of constructed first and second moment of predicted market value

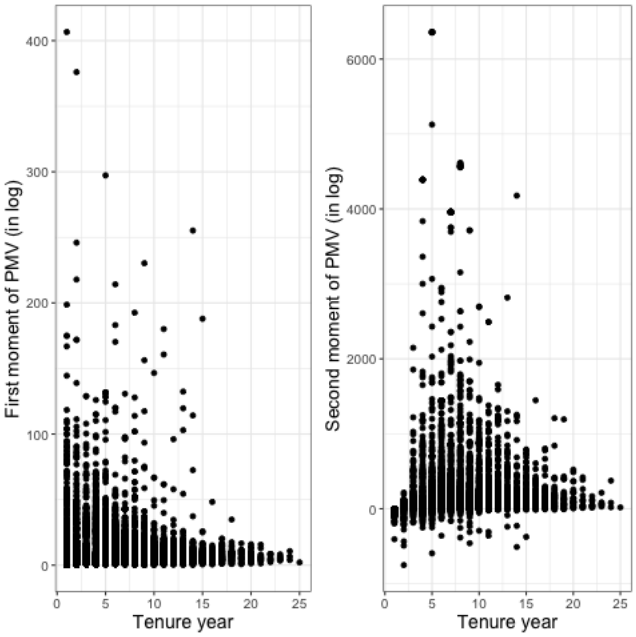
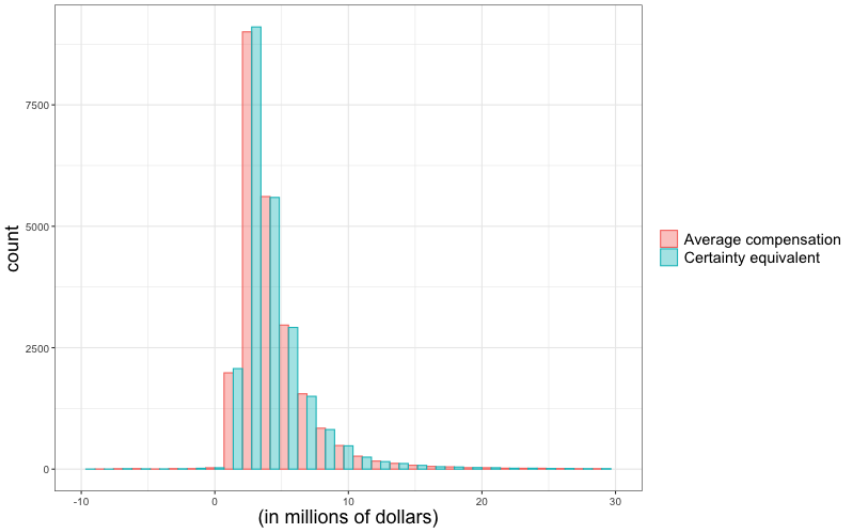


Figure B.3: Distribution of estimated certainty equivalent and average compensation



Note: We use kernel estimation to compute the average compensation across different realizations of the prior belief $\mathcal{S}_t$ and bond price in the data.

Figure B.4: Distribution of likelihood ratio and ratchet ratio

