Reputation, Entrenchment, and Dynamic Managerial Incentives [1]

Noah Lyman

[**Click here for most recent draft**]

ABSTRACT: CEOs are rarely fired. This fact, often attributed to entrenchment, may adversely impact managerial incentives; entrenched CEOs can advance their private interests at shareholders' expense while facing little risk of disciplinary termination. In this paper, I estimate a dynamic principal-agent model to assess the impact of entrenchment on managerial incentives. Firms hire CEOs of unknown quality and subsequently design the optimal compensation contract. Firms gradually learn about CEO quality and make replacement decisions based on their beliefs. CEOs are entrenched, so replacement is costly. The threat of termination compresses CEO pay in equilibrium, though this effect is weakened by entrenchment. Counterfactual experiments reveal a 10.3% reduction in average CEO compensation upon the elimination of entrenchment. Moreover, entrenchment is more costly for shareholders than moral hazard; on average, firm value increases by 6.3% when eliminating entrenchment compared to 1.3% when eliminating moral hazard. Entrenchment slows the rate at which low-quality CEOs are terminated and increases the cost of aligning incentives, exacerbating the severity of moral hazard.

KEYWORDS: Personnel Economics, Contracts, Executive Compensation, Turnover, Corporate Governance.
JEL CLASSIFICATION: M50, D86, D83, G30

[1]First version: August 7, 2023. This version: November 29, 2023. Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 CB 3305. Contact: **nmlyman@unc.edu**. I thank my dissertation advisors Andrés Hincapié, Barton Hamilton, and Fei Li for their guidance and support. I also thank Luca Flabbi and Qing Gong for their helpful comments. This work has also benefited from presentations at the Fuqua School of Business, Olin Business School, Kenan-Flagler Business School, the University of Mississippi, North Carolina State University, the College of Charleston, the University of South Carolina, the 2022 and 2023 Southeastern Economics Association meetings, the 2023 WashU Economics Graduate Student Conference, and the 2023 Summer Meeting of the Econometric Society.

1 Introduction

CEOs are rarely fired. In a given year, only 3% of CEOs are forcefully removed from their positions. This fact is often attributed to managerial entrenchment, a term encompassing a variety of inefficient mechanisms through which CEOs are protected from job loss.[2] As discussed by Taylor (2010), entrenchment leads boards to adopt inefficient firing policies at the expense of shareholder value. Furthermore, such protection from replacement can weaken the alignment of incentives between CEOs and shareholders, exacerbating moral hazard and increasing the incentive-aligning level of pay. While it is well understood that entrenchment increases the cost of CEO replacement, what impact this has on managerial incentives is an open question.

In this paper, I estimate a dynamic principal-agent model to quantify the cost of entrenchment and moral hazard, analyzing in detail the interaction between the two frictions. Boards hire CEOs of uncertain quality which is gradually learned as employment progresses. CEOs are privately informed of their actions and have limited liability, giving rise to moral hazard. Upon learning that their CEO is of sufficiently low quality, boards can fire their CEO and draw a replacement from a fixed population of executives. Firing a CEO subjects boards to both monetary and non-monetary costs, the latter reflecting entrenchment. Both incentive pay and the threat of termination help motivate the CEO to select efficient actions, though the incentive effects of termination are weakened by entrenchment.

I find that entrenchment is significantly more harmful for shareholders than moral hazard. Firm value increases by 6.32% on average upon the elimination of entrenchment compared with a 1.34% increase when eliminating moral hazard. Entrenchment more than doubles the cost of forced termination, pushing the firing rate far below its efficient level. This substantially increases CEO employment lengths, particularly for the lowest-quality CEOs. Furthermore, entrenchment weakens CEOs' termination incentives, the motivation to engage in efficient behavior else face risk of job loss, which increases the level of compensation necessary to achieve incentive compatibility. The model predicts a 10.3% reduction in average CEO compensation upon the elimination of entrenchment. The weakening of incentives under entrenchment more than doubles the cost of moral hazard. When CEOs are entrenched, moral hazard induces a 1.34% drop in firm value below its first-best level. In the absence of CEO entrenchment, this drop falls to 0.62% on average. This implies that measures taken to mitigate entrenchment, appointing in-

[2]Entrenchment may arise through a number of anti-takeover mechanisms (ex: poison pills, golden parachutes, voting agreements) or through personal relationships between managers and board members.

dependent board members for example, would significantly reduce the severity of moral hazard.

This paper makes several contributions to the literature. First, this paper makes a methodological contribution to the growing literature on CEO employment dynamics.[3] I leverage theoretical results from Demarzo and Sannikov (2017) to embed the equilibrium contract within the canonical model of CEO turnover first employed by Taylor (2010), allowing compensation and employment dynamics to be studied in tandem. Second, I contribute to the literature on corporate governance by quantifying the impact of entrenchment on managerial compensation. The role of disciplinary termination in CEO compensation contracts is often overlooked; I endogenize termination and provide empirical evidence that the threat of termination has considerable effects on the provision of managerial incentives. The level of entrenchment is often viewed as inversely related to a firm's quality of governance. Thus, the results of this paper may be interpreted as evidence that the cost of moral hazard declines with governance quality. Finally, this paper contributes to the broader literature in labor economics on job matching under uncertainty about worker quality (Jovanovic, 1979; Miller, 1984; Moscarini, 2005). This literature typically views firms and workers as symmetrically informed about worker quality, which is gradually learned over time. This framework provides a natural explanation of the increasing wage-tenure relationship commonly observed empirically; conditional on a job match surviving, posterior beliefs increase on average, driving wages up with tenure. I present a second explanation for the wage-tenure relationship in environments with asymmetric information between the firm and worker. Namely, as tenure increases, termination incentives fade, placing upward pressure on the incentive-aligning level of compensation.

With a panel of publicly traded North American firms spanning from 1995-2019, I estimate the model using the simulated method of moments. Firm-level information is obtained from Compustat while CEO-level information is obtained from Execucomp. Using data provided by Peters and Wagner (2014) and Jenter and Kanaan (2015), I classify cases of CEO turnover as either forced (being fired) or voluntary (retiring). As mentioned previously, the rate of forced CEO turnover is quite low; on average, only 3% of CEOs are fired in a given year. Using a similar structural approach, Taylor (2010) and Hamilton et

[3] The use of structural techniques is growing increasingly common in this literature. Taylor (2010) estimates a dynamic model of CEO turnover to measure the impact of entrenchment on firm value. Lippi and Schivardi (2014) use a similar approach to study the impact of concentrated ownership on executive selection. More recently, Ferraro (2021) and Hamilton et al. (2023) have extended the model of Taylor (2010) to respectively study female leadership and nepotism. Barry (2023) estimates a similar model to study the impact of shareholder voice on CEO pay.

al. (2023) find strong evidence that this rate is far below its efficient level. I confirm these conclusions, and delve into their implications for managerial incentives, which are not considered in their papers.

Misalignment of incentives between shareholders and CEOs has been an utmost concern since the separation of corporate ownership and control in the early 20^{th} century (Edmans et al., 2017). The objectives of shareholders and CEOs are unlikely to perfectly coincide, and perfect monitoring of the CEO is generally taken to be prohibitively costly (Hermalin and Weisbach, 1998). This gives rise to moral hazard, whose costs have received considerable empirical attention.[4] The literature generally agrees: conditional on employing a CEO, misalignment of incentives is harmful for firm value. However, securing a qualified CEO is no easy task. Firms face substantial uncertainty about CEO quality at time of hire,[5] and such information frictions at the hiring margin compound the moral hazard issue (Jovanovic and Prat, 2014; Demarzo and Sannikov, 2017). Given their limited information, firms cannot disentangle the effects of effort and quality when monitoring CEO performance, posing an identification problem for firms which CEOs can leverage for their private gain. Reminiscent of the ratchet effect in Laffont and Tirole (1988), CEOs who convey positive information to firms face more demanding incentive schemes later in their employment. CEOs thus have an incentive to convey negative information, which can be achieved by privately expropriating firm resources and increasing firm pessimism about future performance. Such behavior is inefficient, and the problem is especially severe when the rate of firm learning is slow. Indeed, the model estimates imply that firms learn about CEO quality quite slowly. After 10 years of CEO tenure, roughly 30% of the initial uncertainty remains. Furthermore, among CEOs with 10 years of tenure, roughly one in three has quality below that of the average replacement.

Retention of low-quality CEOs is detrimental for shareholders, as managerial quality has been found to be an important determinant of firm performance.[6] The existing literature has documented high variation in CEO quality, implying that the difference between a high and low-quality CEO is quite pronounced. I confirm this result; for the median-sized firm in my sample, the model estimates imply a \$73.6 million dollar increase in yearly cash flows following a one standard deviation increase in CEO quality. Thus, firms stand to gain a substantial amount of value through the prompt replacement of low-quality CEOs. Despite this, the firing option is rarely exercised. This fact, as

[4]See for example: Margiotta and Miller (2000), Gayle and Miller (2009), Gayle et al. (2015), Page (2018), Ai et al. (2022).

[5]See for example: Hermalin (2005), Taylor (2010), Hamilton et al. (2023)

[6]See for example: Bertrand and Schoar (2003), Allgood and Farrell (2003), Peréz-González (2006), Bennedsen et al. (2007), Mackey (2008)

discussed by Taylor (2010), is largely explained by CEO entrenchment. CEOs are said to be entrenched if boards retain them for longer than shareholders would prefer. CEO replacement duties are delegated to firms' board of directors, whose interests may be at odds with shareholders. In particular, boards may consider non-pecuniary factors, independent of shareholder value, when making CEO termination decisions. First, personal relationships with the CEO may lead the board to view forceful replacement as undesirable. Additionally, firing a CEO may reflect poorly on the board, who hired the CEO to begin with. In light of these considerations, I follow Taylor (2010) and represent CEO entrenchment as a non-pecuniary cost incurred by the board when firing their CEO. Through counterfactual experiments, I show that entrenchment substantially increases CEOs' average length of employment. Importantly, this effect is not uniform across the distribution of CEO quality; it is most stark for those of especially low quality. CEOs in the bottom quintile of the quality distribution see a roughly 60% extension in employment lengths on average due to entrenchment, compared with a roughly 10% increase for CEOs in the top quintile of quality.

A high degree of entrenchment is generally associated with weak corporate governance Gompers et al. (2003). In the corporate finance literature, managerial entrenchment has been argued to be an important determinant of a number of firm outcomes including investment decisions (Shleifer and Vishny, 1989), capital structure (Zwiebel, 1996; Berger et al., 1997), and dividend payout policy (Hu and Kumar, 2004). I contribute to this literature by quantifying the effect of entrenchment on equilibrium CEO compensation. Reduced-form analysis, which has proven to be challenging in this literature,[7] is poorly suited for my research question as compensation and termination decisions are endogenous. I tackle this challenge by instead taking a structural approach, modeling explicitly the relationship between turnover, entrenchment, and compensation. In the model, termination and compensation policies are jointly determined in equilibrium and serve as alternative levers through which the board can incentivize the CEO. Entrenchment weakens the credibility of the termination threat, increasing the level of compensation needed to align CEO incentives. The estimates imply an 10.3% reduction in average compensation in the absence of entrenchment. Importantly, entrenchment magnifies the cost of moral hazard through the weakening of managerial incentives. When CEOs are entrenched, moral hazard induces a 1.32% decrease in firm value on average relative to the first-best case. In the absence of entrenchment, these losses drop to 0.62%, implying that entrenchment more than doubles the losses associated with moral hazard. These results suggest that mitigating entrenchment is an effective remedy for the moral hazard

[7] See for example: Lehn et al. (2007), Bebchuk et al. (2009), Chang and Zhang (2015)

problem.

The results of this paper highlight the importance of turnover frictions when studying moral hazard in executive labor markets. Turnover is often overlooked in empirical studies on moral hazard, despite the consensus in the theoretical literature that turnover serves as a useful incentive device. The incentive effects of turnover have been studied as far back as Stiglitz and Weiss (1983), and more recently by Spear and Wang (2005) and DeMarzo and Fishman (2007) in discrete time settings. As discussed by Spear and Wang (2005), the income effect may lead termination of a risk-averse agent to be optimal when the agent's continuation payoff is sufficiently high. In other words, the agent may become "too rich" to effectively punish through compensation, thus leaving a role for termination in the optimal contract. Alternatively, termination may be optimal if the agent's continuation payoff becomes too low, particularly in the presence of limited liability. In this paper, I model CEOs as risk-neutral with limited liability, the latter constraint serving as the source of the moral hazard problem. Thus, termination serves as a punishment of last resort in the framework considered here. More recently, the incentive effects of turnover have received attention in a continuous time setting (Sannikov, 2008; Biais et al., 2010; Zhu, 2013; Demarzo and Sannikov, 2017; Grochulski and Zhang, 2023). Continuous-time methods improve the tractability of dynamic incentive problems, as the derivation of optimal incentives amounts to solving a partial differential equation, which can be done numerically with low computational burden. For improved tractability, I adopt a continuous-time approach in this paper.

On the theoretical front, the papers most similar to this one are Jovanovic and Prat (2014) (JP) and Demarzo and Sannikov (2017) (DS). Both papers develop dynamic models of moral hazard in which the principal and agent face symmetric uncertainty about agent quality. JP allows learning to be non-stationary while DS restricts attention to stationary learning. Uncertainty reduction over tenure is an important consideration for my research question, so I adopt the JP assumption of non-stationary learning. However, JP assumes agents are risk-averse with unlimited liability and does not consider turnover. I explicitly model turnover, both endogenous termination and exogenous retirement, and assume agents to be risk-neutral with limited liability. The assumption of limited liability is important in this paper, as it leaves a role for disciplinary termination in the equilibrium contract. Endogenous termination is included in DS, but agent replacement is not; rather, firms liquidate and cease operations when their agent is terminated. In my model, agents are replaced following an instance of turnover and firms continue operations. On the empirical front, the paper most similar to mine is Taylor (2010) who uses a structural approach to explain the low rate of forced CEO turnover observed empirically.

Taylor finds that the key determinant of this empirical regularity is CEO entrenchment, and offers a compelling argument suggesting entrenchment is detrimental for firm value. I build upon his paper by embedding the optimal contract into his framework, allowing for a detailed analysis of the equilibrium response of CEO compensation to turnover frictions. I find that the response is considerable, and show that turnover frictions have major implications for the severity of moral hazard in the executive labor market.

The remainder of the paper proceeds as follows. Section 2 outlines the sample and key empirical patterns motivating the structural model. Section 3 presents the theoretical environment and derivation of firms' optimal compensation and turnover policies. Section 4 discusses model identification and the estimation procedure. Section 5 discusses the model estimates. Counterfactual experiments are presented in Section 6. Closing discussion and concluding remarks are contained in Section 7.

2 Data

The sample is a matched CEO-firm yearly panel of North American publicly-traded firms spanning 1994-2019. The panel is constructed by linking three sources of data. First, I obtain CEO-level information from Execucomp, which provides detailed data on compensation packages and CEO tenure. I match this with Compustat which reports information on firm assets, industry classification, income, and other financial fundamentals. Firm performance is measured using their return on assets (ROA), defined as operating income per dollar in assets.[8] Firms with missing operating income or missing total assets are omitted from the final sample. Lastly, I match the sample with supplementary turnover data (Peters and Wagner, 2014; Jenter and Kanaan, 2015), which classifies instances of CEO turnover as forced or voluntary using the method outlined by Parrino (1997). The final sample consists of 42,513 firm-year observations, 3,627 distinct firms, and 8,191 distinct CEO employment spells. I observe 5,005 cases of CEO turnover, where 1,260 (25.2%) are forced and 3,745 (74.8%) are voluntary.[9] Throughout this section, I index CEOs by i, firms by j, and calendar years by t.

Summary statistics are reported in Table 1. CEOs in the sample are predominantly male; only 2.8% of CEOs are female. The average executive is 53 years in age. The majority of CEOs have some prior experience with their firms prior to becoming CEO. On average, an executive has roughly 9 years of firm-specific experience when appointed for

[8] ROA is defined as operating income ($oibdp_{ijt}$ in Compustat) divided by total assets (at_{ijt} in Compustat): $y_{ijt} = \frac{oibdp_{ijt}}{at_{ijt}} \times 100$

[9] See Appendix 8.1 for extended details on the turnover classification and construction of the sample.

Table 1: Summary Statistics

	Mean	Std. Dev.	5^{th} Percentile	Median	95^{th} Percentile
	Panel (a): Firm Characteristics				
Profitability (ROA)	11.9	11.8	-1.75	12.0	28.5
Total assets ($ Billions)	17.1	103.5	.178	2.22	51.5
Total revenue ($ Billions)	6.93	21.0	.136	1.66	27.4
	Panel (b): CEO Characteristics				
Spell length	8.29	7.12	1	6	22
Eventually fired	.259	.438	0	0	1
Eventually retired	.743	.437	0	1	1
	Panel (c): CEO Compensation ($ Millions)				
Total compensation	6.32	12.2	.601	3.72	18.9
Salary	.866	.457	.327	.809	1.56
Bonus	.549	1.88	0	0	2.40
Bonus (Conditional on > 0)	1.23	2.66	.058	.643	3.88
Other compensation	4.91	11.9	.015	2.44	16.2

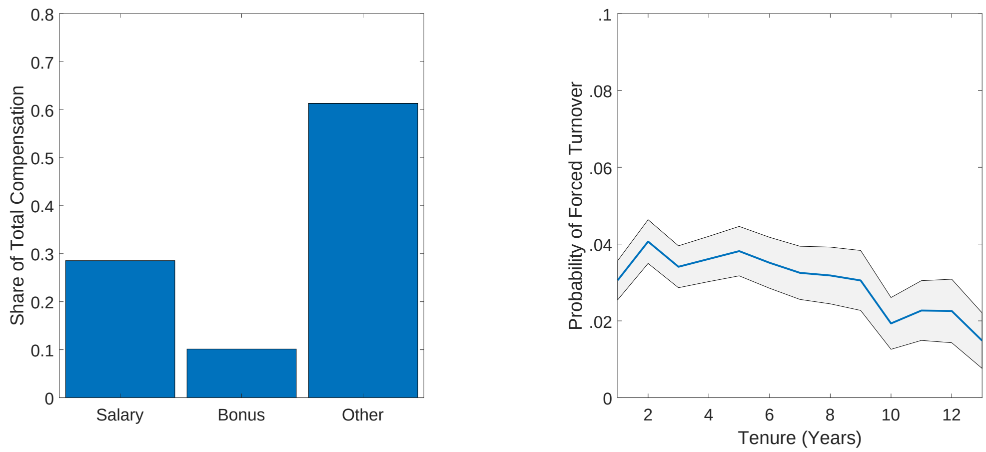
Notes: The unit of observation in Panels (a) and (b) is a firm-year. The unit of observation in Panel (c) is a CEO. All monetary values are expressed in 2015 dollars.

the CEO position. The average length of employment as CEO is 8.29 years, though the spell length distribution exhibits substantial rightward skewness. Of the CEO employment spells which are not right-censored, roughly 26% end in forced termination.

The level of CEO pay is substantial and largely attributable to performance pay, with fixed salary making up only about 30% of total compensation for the average CEO (Figure 2(a)). Performance pay is composed primarily of equity incentives, the use of which in executive compensation packages is well documented.[10] The use of performance pay is intended to align the interests of the CEO with those of the firm, mitigating the CEO's motivation to pursue private interests (Margiotta and Miller, 2000). In addition, the threat of forced termination provides incentives by serving as a punishment of last resort (Spear and Wang, 2005). Performance pay and the threat of termination thus complement each other in the incentive mix, rewarding the CEO in cases of positive performance and punishing the CEO in cases of persistent negative performance.

[10]Through the 1990s and beyond, stock and option packages surged to become the dominant component of CEO compensation. See Edmans et al. (2017) for a detailed discussion.

Figure 1: CEO Pay and Forced Turnover



(a) Average Pay Composition (b) Forced Hazard Rate

Notes: Panel (a) plots the composition of CEO compensation packages on average. Roughly 30% of total CEO pay can be attributed to salary, while the rest is split between cash bonuses and other forms of incentive pay. "Other" is composed of equity incentives including restricted stock grants, option grants, and long-term incentive payouts. Panel (b) plots the rate of forced turnover over the first 15 years of CEO tenure. The likelihood of forced termination is low and gradually declines with tenure.

2.1 Key Empirical Facts

Next, I outline the key empirical facts motivating my modeling decisions. The data suggest that variation in CEO quality is high; CEOs of relatively high quality generate substantially higher rates of profitability than their low-quality counterparts. Furthermore, firms are responsive to new information about the quality of their CEO, revealed by the firm's financial performance. As information is generated, firms adjust their pay contracts, and in the case of exceptionally poor performance, exercise their firing option. Low-quality CEOs are gradually forced out of employment, inducing positive selection of CEO quality over tenure. As tenure increases, firms substitute away from termination incentives towards monetary incentives. Thus, as CEO employment progresses, those who survive into later years of tenure are increasingly motivated by the use of incentive pay and decreasingly motivated by the risk of forced termination.

CEOs are rarely fired. Figure 2(b) plots the rate of forced turnover over the first 13 years of CEO tenure. As previously documented in the literature, CEOs are unlikely to be fired. The likelihood of forced termination is highest in early years of tenure, peaking at roughly 4% and otherwise generally declining with tenure. There are many potential explanations underlying the low rate of forced termination. As discussed by Taylor (2010), replacing a CEO is quite costly, so boards may only exercise their firing option as a last resort. Alternatively, if CEOs are relatively homogeneous in the population, CEO replacement may have minimal impact on the trajectory of firm performance. I argue next, however, that this second possibility is unlikely.

Variation in CEO quality is high. CEO quality has been shown to be extremely consequential for firm performance (Allgood and Farrell, 2003; Bertrand and Schoar, 2003; Bennedsen et al., 2020). While CEO quality is not directly observed in the data, I create a proxy for it using observed firm profitability. Specifically, I project firm profitability on a vector of firm and CEO characteristics, and obtain the shrinkage estimate of CEO quality given as a weighted average of the resulting profitability residuals. I begin by estimating the equation:

$$y_{ijt} = \lambda_0 + \lambda_1 \mathbf{C}_{ijt} + \lambda_2 \mathbf{F}_{ijt} + \tau_t + \gamma_j + \epsilon_{ijt} \tag{1}$$

where y_{ijt} denotes the return on assets for firm j employing CEO i at time t. $\mathbf{C}_{ijt}$ and $\mathbf{F}_{ijt}$ are a vector of CEO and firm characteristics, respectively.[11] τ_t and γ_j are year and firm fixed effects. Next, I use the residual component of Equation (1) to create a coarse proxy of CEO quality. For each CEO-firm match ij, denoting their length of employment by s_{ij}, the shrinkage estimate of CEO match quality θ_{ij} is defined by:

$$\hat{\theta}_{ij} = \frac{\omega_j}{1+\omega_j}\left(\frac{1}{s_{ij}}\sum_{ij'} \hat{\epsilon}_{ijt}\mathbf{1}[ij' = ij]\right) \tag{2}$$

$$\omega_j = Var(\hat{\epsilon}_{ij't} \mid j' = j\,) \tag{3}$$

$\hat{\theta}_{ij}$ is the James-Stein estimator of true match quality θ_{ij} and $\hat{\epsilon}_{ijt}$ fitted residuals obtained from estimating Equation (1).[12] Table 2, summarizes the distribution of $\hat{\theta}_{ij}$ across all CEOs in the sample. Its standard deviation is 5.4, which lies within the range found in

Table 2: Distribution of CEO Quality $\hat{\theta}_{ij}$

	Mean	Std. Dev.	5^{th} pct	50^{th} pct	95^{th} pct
Quality Proxy ($\hat{\theta}_{ij}$)	-.171	5.40	-7.15	0.00	6.51

previous literature.[13] For the median-sized firm in the sample, the implied difference between cash flows generated by a CEO at the 5^{th} and 95^{th} percentile of the distribution

[11] Specifically, I control for the CEO's age, gender, and tenure (as CEO), along with firm assets and total revenue.

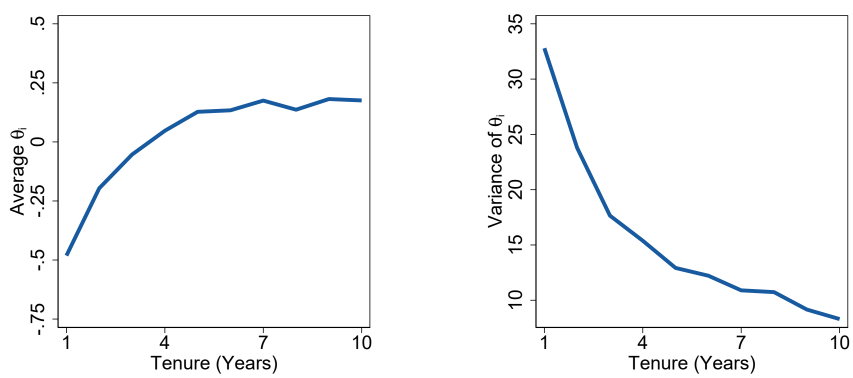
[12] Match quality θ_{ij} is clearly measured with error in the data. Given this measurement error, the James-Stein estimator, while biased, minimizes mean-squared error among the set of admissible M-estimators.

[13] For example: Taylor (2010) reports a SD of 2.42 while Bertrand and Schoar (2003) report a SD of 7.

of $\hat{\theta}_{ij}$ is roughly \$303 million per year.[14] This staggering difference suggests that the difference between a high and low-quality CEO is quite pronounced.

CEO quality is positively selected over tenure Notably, the distribution of $\hat{\theta}_{ij}$ evolves with CEO tenure. Figure 2 shows that the average of CEO quality increases over tenure (Panel (a)) while its variance decreases (Panel (b)). Such a pattern is consistent with positive selection on quality; firms exercise their firing option in response to poor signals of CEO quality, inducing gradual attrition of low-quality CEOs.[15] As low quality CEOs gradually exit the sample, the average quality of those who survive is pushed upwards. Furthermore, the termination of low quality CEOs compresses the variance of CEO quality among those employed, as the distribution becomes more concentrated around relatively high levels of quality.

Figure 2: Distribution of $\hat{\theta}_{ij}$ over Tenure



(a) Average $\hat{\theta}_{ij}$ (b) $Var(\hat{\theta}_{ij})$

Notes: Panel (a) plots the average value of the CEO quality proxy $\hat{\theta}_{ij}$ over the first 10 years of tenure while Panel (b) plots the variance. Together, the figures suggest that the distribution of CEO quality becomes increasingly concentrated among high values as tenure increases.

Termination incentives and monetary incentives are substitutes. Given the evolution of the CEO quality distribution, I next consider the co-evolution of the incentive mix. Let δ_{ijt} denote contract ij's pay-performance sensitivity in year t as calculated in Core and Guay (2002) and Coles et al. (2006).[16] Specifically, δ_{ijt} is defined as the year t monetary

[14] The median-sized firm in the sample has \$2.219 billion in assets. The approximate \$303 value is obtained by: $2219 \times \frac{6.51+7.15}{100} \approx 303$.

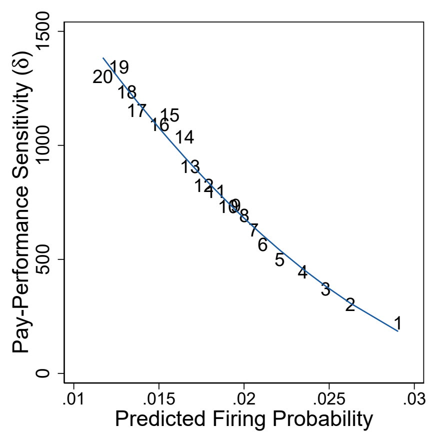
[15] I show in Appendix 8.1.2 that forced turnover decisions are sensitive to cumulative performance, while voluntary turnover decisions are statistically independent of cumulative performance.

[16] δ_{ijt} is calculated as the change in option portfolio value in response to a 1% increase in the firm's stock price, where options are valued using the standard model of Black and Scholes (1973) as modified by Merton (1973) to accommodate for dividend payouts. See Core and Guay (2002) for more details.

return (in \$1000s) that CEO i would receive in response to a 1% increase in firm j's stock price.

Figure 3 plots the evolution of pay sensitivity and the predicted firing probability over CEO tenure. Each point corresponds to a level of tenure. Long-tenure CEOs, who are of relatively high quality on average, face a very low risk of forced termination on average, and are motivated primarily by monetary incentives. The opposite is true for newly-tenured CEOs; monetary incentives are weaker relative to long-tenured CEOs while termination incentives are stronger. Figure 3 gives a sharp depiction of this gradual substitution away from termination threats towards monetary incentives. This substitution suggests a change in the firm's relative cost of providing termination versus monetary incentives. In particular, the observed pattern is consistent with termination incentives becoming increasingly costly relative to monetary incentives as CEO tenure increases. It is more costly to threaten to fire a proven, long-tenured CEO than a brand new CEO of uncertain quality. Such a phenomenon can have important implications for misalignment of incentives between firm and CEO, which I explore formally next.

Figure 3: Incentive Pay, Termination Risk, and CEO Tenure



Notes: For each of the first 20 years of CEO tenure, Figure 3 scatters the average pay-performance sensitivity over the predicted firing probability conditional on firm characteristics, tenure, and the CEO's history of performance as summarized by $\tilde{\theta}_{ijt}$. Pay-performance sensitivity is measured following Core and Guay (2002) and Coles et al. (2006); it gives the number of dollars (in thousands) a CEO would receive in response to a 1% increase in their firm's stock price. The figure reveals that CEO's risk of termination decreases with tenure, while their pay sensitivity increases. The risk of termination has a weaker impact on incentives late in CEOs' careers; these weakened termination incentives are gradually replaced with financial incentives.

3 Model

The model features two types of decision makers: boards, who act on behalf of their firm, and their respective CEOs. A firm's rate of profitability is determined by the their level of productivity, turnover costs, idiosyncratic shocks, and their CEO's quality and private actions.[17] Neither quality nor actions are directly observed by boards. Rather, quality is gradually learned by observing profitability and efficient actions are implemented through the board's design of a full-commitment contract. The contract specifies optimal compensation and termination policies which serve as alternative mechanisms through which incentives can be delivered to the CEO. Termination occurs when the board believes their CEO to be of sufficiently low quality, at which point boards hire a replacement and continue operations. The board designs contracts to maximize firm value plus non-pecuniary factors associated with CEO turnover. The presence of such non-pecuniary factors in boards' objective induces a wedge separating firm value and board utility. As such, the equilibrium contract may be inefficient in the sense that it does not purely maximize firm value.

Both boards and CEOs are risk-neutral. Each firm[18] j employs a CEO i whose current level of tenure is denoted by $t \geq 0$, which is continuous. CEO quality θ_i is fixed over time and drawn from population distribution $N(\theta_0, \delta_0^2)$. θ_i is not known by either party, but is gradually learned about over time by observing cash flows. $\tilde{\theta}_{ijt}$ and $\tilde{\theta}^a_{ijt}$ respectively denote the board and CEO's beliefs about θ_i, which do not coincide in general. If the board believes quality to be sufficiently low, they will fire their CEO and hire a replacement. Firing occurs at stopping time T. Otherwise, CEOs retire stochastically at Poisson rate λ. Retirement occurs at random time R. I let $\tau_{ij} = \min\{T, R\}$ denote the time at which CEO i's employment within firm j ends, whether by retiring or being fired. Firms incur cost c in the case of firing or retirement, representing the monetary costs associated with replacing a CEO. Additionally, boards incur non-pecuniary cost π when firing their CEO, reflecting CEO entrenchment. Both c and π are measured as a fraction of firm assets, allowing turnover costs to vary with firm size. The tenure index t resets to zero upon the replacement of a CEO. Boards operate over an infinite horizon, hiring successive CEOs whenever the previous one departs.

Firm Profitability Each firm j has two unique characteristics: total assets b_j and a productivity parameter $\gamma_j \sim N(0, \sigma_\gamma^2)$, both of which I assume to be known and constant over

[17] All firms operate independently of one another, so broader market equilibrium considerations are not addressed in this paper.

[18] Because each firm has a single board, the j subscript can be seen as indexing both firms and boards.

time.[19] Firm cash flows dX_{ijt} are represented as the increment of an Ito process following:

$$dX_{ijt} = b_j \overbrace{\left(\gamma_j dt + dY_{ijt}\right)}^{\text{Profitability Rate}} \tag{4}$$

$$dY_{ijt} = (\theta_i - a_{ijt})dt + \sigma_W dW_{ijt} \tag{5}$$

$$X_{ij0} = -b_j c + X_{(i-1)j\tau_{(i-1)j}} \tag{6}$$

For firm j with CEO i, the rate of profitability at tenure t is obtained by dividing cash flows dX_{ijt} by assets b_j. Profitability has two components. The drift term $\gamma_j dt$ represents the firm's known contribution to profitability, which is independent of the CEO. dY_{ijt} is the contribution of unobservables to profitability, which I henceforth refer to as *residual performance*. dY_{ijt} increases in the CEO's quality and decreases in their private action $a_{ijt} \geq 0$, representing the diversion of firm cash flows towards the CEO's private consumption. $W_{ijt} = \int_0^t dW_{ijs}$ is a standard Brownian Motion on probability space $\{\Omega, \mathcal{F}, P\}$,[20] and the parameter σ_W measures the volatility of contemporaneous performance. At the beginning of CEO i's employment spell, cumulative cash flows initialize at X_{ij0}. This initial condition is given by the cumulative level of cash flows at the time of departure for the previous CEO, denoted by $X_{(i-1)j\tau_{(i-1)j}}$, minus the monetary turnover cost $b_j c$. c is represented as a fraction of assets, so scaling by firm assets converts the units of the turnover cost to dollars.

Preferences and outside options CEOs are risk-neutral with flow utility:

$$u(w_{ijt}, a_{ijt}) = w_{ijt} + \phi_j a_{ijt} \tag{7}$$

$$\phi_j = b_j^\alpha \quad \alpha \in (0,1) \tag{8}$$

w_{ijt} is the tenure-t realization of compensation specified by the $\mathcal{F}_{ijt}$-adapted process $w : [0,T] \times \Omega \to \mathbb{R}_+$ and ϕ_j measures the size-dependent rate of cash flow diversion. Concretely, diverting a_{ijt}% of cash flows from the firm yields $\phi_j a_{ijt}$ dollars directly to the CEO, where ϕ_j increases with firm size. Hence, cash flow diversion is more profitable for CEOs in larger firms, implying that the misalignment of incentives between firm and

[19] The assumption of constant firm assets is equivalent to assuming that all profits are immediately paid as dividends to shareholders. This allows me to abstract from dividend payout decisions, which are outside of the scope of this paper. Assets are denoted in millions of dollars; I take as given that $b_j > 1$ for all j, so all firms have at least \$1 million in assets. The minimum value of the empirical distribution of assets is roughly \$57 million.

[20] Ω is the set of all sample paths of $\{W_{ijt}\}_{t\geq 0}$, with $\omega \in \Omega$ denoting an arbitrary sample path. P is a probability measure over $\mathcal{F}$, a $\sigma-$algebra over Ω.

CEO grows with firm size (Gayle and Miller, 2009). Imposing $\alpha < 1$ renders cash flow diversion inefficient and allows us to restrict attention to contracts which implement no diversion ($a_{ijt} = 0$) for all t.[21]

Upon exiting from the firm, the CEO receives outside option $C(\tilde{\theta}^a_{ij\tau})$ which depends explicitly on their private beliefs at time of departure. Specifically, I assume CEO outside options are an increasing, linear function of their perceived quality $\tilde{\theta}^a_{ijt}$:

$$C(\tilde{\theta}^a_{ijt}) = \mu + \frac{\phi_j}{r}\tilde{\theta}^a_{ijt} \tag{9}$$

Hence, high-quality CEOs will (on average) a high outside payoff, and thus will be more expensive to retain than their low-quality counterparts. The assumption that $\frac{dC}{d\tilde{\theta}^a} = \frac{\phi_j}{r}$ is not innocuous. As will be articulated later in this section, imposing this slope on the CEO's outside option implies that the marginal benefit associated with an increase in private beliefs is the same within and outside of the firm. This assumption vastly improves both the analytic and numerical tractability of the model. Furthermore, this slope assumption has little effect on the key qualitative results of the model.

Both parties discount the future at rate ρ. However, given the possibility of exogenous termination, the arrival rate λ of voluntary separations will be absorbed into the discount rate. For notational brevity, I therefore define $r \equiv \rho + \lambda$ as the effective discount rate. Given a strategy $a : [0, T] \times \Omega \to \mathbb{R}_+$, the CEO's net present value of employment at tenure t is given by:[22]

$$U^a_{ijt} = \mathbb{E}^a_t\left[\int_t^T e^{-r(s-t)} u(w_{ijs}, a_{ijs}) ds + \lambda \int_t^T e^{-r(s-t)} C(\tilde{\theta}^a_{ijs}) ds + e^{-r(T-t)} C(\tilde{\theta}^a_{ijT})\right] \tag{10}$$

The first term reflects the discounted flow payoffs accumulated during employment. The second term reflects the possibility of future retirement, in which case the CEO departs and collects their outside option. If the CEO is fired prior to the arrival of a retirement shock, they will also depart and collect their outside option, as reflected by the third term. Note that U^a_{ijt} is the continuation payoff given an arbitrary strategy a, which in general does not coincide with the board's recommended strategy, denoted by a^*.

I assume boards are risk-neutral and maximize the expected net present value of cash flows net of CEO pay and the non-pecuniary cost of termination. At the outset of the

[21] See Appendix 8.2 for a proof of this statement.

[22] Throughout the model, $\mathbb{E}^a_t[x] = \int_\Omega x dP^a_t$, where P^a_t is the tenure t probability measure arising from having observed the action process a. On the other hand, $\mathbb{E}_t[x] = \int_\Omega x dP_t$, where P_t is the tenure t probability measure having not observed the action process. Given that the CEO observes a and the firm does not, the two parties will condition their expectations on different information.

contractual relationship, the firm's optimal payoff is given by:

$$V_{ij0} = \max_{\mathscr{C}\in\mathbb{C}} \mathbb{E}_0\left[\int_0^T e^{-rt} dX_{ijt} - \int_0^T e^{-rt} w_{ijt} dt + \lambda \int_0^T e^{-rt} V_T dt + e^{-rT}\Big(V_T - b_j \pi\Big)\right] \tag{11}$$

The board maximizes over the space of admissible contracts $\mathbb{C}$, where a contract $\mathscr{C} = (w, a, T)$ is a triple specifying a compensation process w, action process a, and stopping time T, all of which are $\mathcal{F}_{ijt}$-adapted. Upon turnover, the board receives value V_T which is defined as:

$$V_T \equiv V_{(i+1)j0} \tag{12}$$

Following any instance of turnover, the board immediately draws a successor CEO $(i+1)$ from distribution $N(\theta_0, \delta_0^2)$ and continues operations. Note that given the initial condition for profitability (6), the monetary turnover cost c is reflected in the term $V_{i'j0}$. Additionally, in the case of firing the CEO, reflected by the last term of (11), the board incurs non-pecuniary cost $b_j\pi$. This represents CEO entrenchment, which raises the effective cost of forceful CEO replacement. Hence at stopping time T, the board incurs cost $b_j(\pi + c)$ and continues operations with new CEO i'. Note here that the board's problem is stationary and the turnover value V_T, the value of the subsequent CEO's employment, is independent of the current employment spell. When deriving the optimal contract, V_T can thus be treated as a constant.

Learning Though CEO quality is unknown, information about θ_i is continuously generated as firm profitability is realized. The distributions $N(\tilde{\theta}_{ijt}, \tilde{\delta}_{ijt})$ and $N(\tilde{\theta}^a_{ijt}, \tilde{\delta}_{ijt})$ represent the respective beliefs of the board and CEO given information up to t, where $\tilde{\theta}_{ijt} = \tilde{\theta}^a_{ijt}$ in equilibrium. Off the equilibrium path however, given their private knowledge of the process a, the CEO may form beliefs which differ from the board's. Considering deviations in the board and CEO's beliefs if thus necessary for establishing incentive compatibility. I assume rational expectations, so the initial beliefs of the board and CEO coincide with the population distribution:

$$\tilde{\theta}_{ij0} = \tilde{\theta}^a_{ij0} = \theta_0 \tag{13}$$

$$\tilde{\delta}^2_{ij0} = \delta_0^2 \tag{14}$$

As CEO tenure increases, beliefs adjust in response to realized performance. In particular, given the normality of both θ_i and W_{ijt}, the posterior mean will be an increasing, linear function of cumulative performance. Define cumulative performance Y_{ijt} and cumulative action A_{ijt} by:

$$Y_{ijt} = Y_{ij0} + \int_0^t (\theta_i - a_{ijs}) ds + \sigma_W \int_0^t dW_{ijs} \tag{15}$$

$$A_{ijt} = \int_0^t a_{ijs} ds \tag{16}$$

The parameters of the equilibrium belief distribution $N(\tilde{\theta}_{ijt}, \tilde{\delta}^2_{ijt})$ are then given by:

$$\tilde{\theta}_{ijt} = \frac{\delta_0^{-2}\theta_0 + \sigma^{-2}(Y_{ijt} - Y_{ij0} + A_{ijt})}{\tilde{\delta}^{-2}_{ijt}} \tag{17}$$

$$\tilde{\delta}^2_{ijt} = (\delta_0^{-2} + \sigma^{-2} t)^{-1} \tag{18}$$

Beliefs depend only on the CEO's cumulative contribution to profitability Y_{ijt}, as the firm productivity component of profitability γ_j is independent of the CEO. Additionally, beliefs are conditioned on the CEO's cumulative action A_{ijt}. In equilibrium, the CEO will always pick the board's recommended action, in which case the board correctly infers A_{ijt} and shares the same estimate of θ_i as the CEO. The equilibrium law of motion for $\tilde{\theta}_{ijt}$ follows from Ito's lemma:

$$d\tilde{\theta}_{ijt} = \frac{\tilde{\delta}^2_{ijt}}{\sigma_W^2}(dY_{ijt} - (\tilde{\theta}_{ijt} - a^*_{ijt})dt) \tag{19}$$

$$= \nu_{ijt}\sigma_W dZ_{ijt} \tag{20}$$

where $dZ_{ijt} = \sigma_W^{-1}(dY_{ijt} - (\tilde{\theta}_{ijt} - a^*_{ijt})dt)$ is the *innovation process*, tracking the realization of performance dY_{ijt} net of expectations $(\tilde{\theta}_{ijt} - a^*_{ijt})dt$. Beliefs adjust in response to these signals with sensitivity $\nu_{ijt} = \tilde{\delta}^2_{ijt}/\sigma_W^2$, which I define as the rate of learning. Equation (20) shows that in equilibrium, the posterior mean is a martingale with volatility $\nu_{ijt}\sigma_W$. The posterior variance on the other hand is deterministic and decreases monotonically with t:

$$d\tilde{\delta}^2_{ijt} = -\nu_{ijt}\tilde{\delta}^2_{ijt} dt \tag{21}$$

Note that the variance of beliefs depends only on tenure, so is unaffected by CEO action choices and hence will be the same on or off the equilibrium path for a given t.

The posterior mean for the board and CEO on the other hand will in general not coincide off the equilibrium path. In particular, deviations from the efficient action a^* will lead boards to misinterpret the realized signal of quality. Relative to the board's expectations, cash flow diversion induces a low realization of contemporaneous performance.[23] This leads the board's beliefs to drift downwards relative to the CEO's, inducing a gap in expectations about future performance. This is reminiscent of the ratchet effect as discussed by Laffont and Tirole (1988). The CEO benefits from conveying negative information to the board, as it eases their future incentive load. To prevent this in equilibrium, the board compensates the CEO via an information rent. Additionally, the CEO's incentive to convey negative information is limited by the risk of termination. If the boards' beliefs fall too low, the CEO will be fired.

CEO Turnover CEOs employment can end either through endogenous termination or exogenous retirement, where retirement shocks arrive at rate λ. Conditional on a retirement shock, the board immediately draws a replacement CEO at cost c and continues operations. Firing occurs when the board's beliefs $\tilde{\theta}_{ijt}$ drop below the endogenous threshold $\theta_f(t)$. Let $V(\tilde{\theta}_{ijt}, t, U_{ijt})$ denote the board's optimal payoff given state $(\tilde{\theta}_{ijt}, t, U_{ijt})$, where U_{ijt} denotes the CEO's promised equilibrium payoff. $\theta_f(t)$ is defined as the level of $\tilde{\theta}_{ijt}$ such that the board is indifferent between continuing with CEO i and drawing a new CEO $(i+1)$:

$$V(\theta_f(t), t, U_{ijt}) = V_T - b_j \pi \tag{22}$$

When CEOs are entrenched, i.e. $\pi > 0$, the board's termination threshold will be strictly lower than the firm-value-maximizing threshold, leading to inefficiently low levels of termination. The stopping time T denotes the first time that $\tilde{\theta}_{ijt}$ reaches $\theta_f(t)$. Concretely:

$$T = \inf\{t < \infty \,|\, \tilde{\theta}_{ijt} = \theta_f(t)\} \tag{23}$$

Board's Problem The optimal contract $\mathscr{C} = (w, a, t)$ delivers the CEO a payoff of U_{ij0} and maximizes the board's objective:

$$V(\tilde{\theta}_{ijt}, t, U_{ijt}) = \mathbb{E}_t\left[\int_0^T e^{-rt} dX_{ijt} - \int_0^T e^{-rt} w_{ijt} dt + \lambda \int_0^T e^{-rt} V_T dt + e^{-rT}\left(V_T - b_j \pi\right)\right] \tag{24}$$

[23] Given that the action process a^* attains its lower bound in equilibrium, it suffices to restrict attention to positive deviations. See Appendix 8.2 for more discussion.

subject to:

$$U_{ij0} = \mathbb{E}_0\left[\int_0^T e^{-rt}u(w_{ijt}, a_{ijt})dt + \lambda\int_0^T e^{-rt}C(\tilde{\theta}_{ijt})dt + e^{-rT}C(\tilde{\theta}_{ijT})\right] \tag{PK}$$

$$U_{ijt} \geq C(\tilde{\theta}_{ijt}) \ \forall \ t \leq T \tag{IR}$$

$$U_{ij0} \geq U_{ij0}^{\hat{a}} \text{ for any other } \hat{a} \tag{IC}$$

(PK) simply defines the CEO's promised value. (IR) and (IC) are the participation and incentive-compatibility constraints, respectively. $\mathbb{E}_t$ and $\mathbb{E}_t^a$ denote the expectation operators given information $\mathcal{F}_{ijt}$ and $\mathcal{F}_{ijt}^a$, respectively. $\mathcal{F}_{ijt}^a$ represents the information set when the action process $\{a_{ijt}\}$ is observed, while $\mathcal{F}_{ijt}$ is the information set under the assumption that true actions coincide with recommended actions. Rather than directly using (IC), I use a first-order approach following Williams (2011) and derive the CEO's first-order-incentive-compatibility condition (FOIC).

3.1 First-Best Case

Before deriving the optimal contract, it is useful to analyze the first-best case. Actions a_{ijt} are observable and the board ensures the first-best action $a_{ijt}^* = 0$ is selected for all t. Given that a_{ijt} is observable, the board and CEO's beliefs are identical (i.e. $\tilde{\theta}_{ijt} = \tilde{\theta}_{ijt}^a$). Here, the board solves a pure optimal stopping problem, monitoring performance and determining when to fire and replace their current CEO. Define $T^{FB} = \inf\{t < \infty | \tilde{\theta}_{ijt} = \theta_f^{FB}(t)\}$ as the first-best stopping time. $\theta_f^{FB}(t)$ is the first-best firing threshold; the lowest value of $\tilde{\theta}_{ijt}$ such that the board is willing to retain their CEO.

Proposition 1 (First-Best Firing Threshold)**.** *When actions are observable, the firing threshold which maximizes the board's payoff is given by:*

$$\theta_f^{FB}(t) = -r\pi + b_j^{-1}\left(\rho V_T - \frac{\nu_{ijt}^2\sigma_W^2}{2}V_{\theta\theta}(\theta_f^{FB}(t), t)\right) \tag{25}$$

Proof. See Appendix 8.2. □

Condition (25) implicitly defines the optimal firing boundary.[24] In the case of no entrenchment, $\pi = 0$ and (25) is shareholder-optimal in the sense that it maximizes firm value. Importantly, when $\pi > 0$, a wedge is induced which separates firm value from

[24]Obtaining a closed-form representation of $\theta_f^{FB}(t)$ is not feasible, but it can be computed numerically. See Appendix 8.2 for a detailed exposition of the numerical solution.

the board's optimal payoff V. Thus, in the presence of entrenchment, the board's enacted firing rule does not coincide with the shareholder-optimal firing rule. In particular, when CEOs are entrenched, the board retains some CEOs which shareholders would have preferred to see terminated ex-post. This also holds in the second-best case when CEO actions are unobservable, which I consider next.

3.2 Second-Best Case

In the second-best case, the CEO has private information a. As a result, off of the equilibrium path the board and CEO will not share the same beliefs about quality. It is convenient to define $\alpha_{ijt} = \tilde{\theta}^a_{ijt} - \tilde{\theta}_{ijt}$ as the gap in beliefs at tenure t. Because cash flow diversion is inefficient, the optimal contract will ensure that $a^*_{ijt} = 0$ for all t. Rewriting the innovation process in (20), we see that the board believes performance follows:

$$dY_{ijt} = \tilde{\theta}_{ijt}dt + \sigma_W dZ_{ijt} \tag{26}$$

whereas the CEO knows that profitability truly follows:

$$\begin{aligned} dY_{ijt} &= \tilde{\theta}_{ijt}dt + \sigma_W dZ^a_{ijt} \\ &= (\alpha_{ijt} - \hat{a}_{ijt} + \tilde{\theta}_{ijt})dt + \sigma_W dZ_{ijt} \end{aligned} \tag{27}$$

Hence, there is a disagreement about the data generating process, and the board and CEO will accordingly assign different probability measures to realizations of performance. Let P^0_{ijt} and P^a_{ijt} denote the probability measures arising from the board and CEO's information ($\mathcal{F}_{ijt}$ and $\mathcal{F}^a_{ijt}$), respectively. By Girsanov's Theorem:

$$dP^a_{ijt} = \Lambda_{ijt} dP^0_{ijt} \tag{28}$$

$$\Lambda_{ijt} = exp\left(\frac{1}{\sigma_W}\int_0^t (\alpha_{ijs} - \hat{a}_{ijs})dZ_{ijs} - \frac{1}{2\sigma_W^2}\int_0^t (\alpha_{ijs} - \hat{a}_{ijs})^2 ds\right) \tag{29}$$

$$d\Lambda_{ijt} = \Lambda_{ijt}\frac{\alpha_{ijt} - \hat{a}_{ijt}}{\sigma_W} dZ_{ijt} \tag{30}$$

Λ_{ijt} is the Radon-Nikodym derivative relating the measures P^0_{ijt} and P^a_{ijt}, where $\Lambda_{ijt} = 1$ in the case of no deviations. The relative density process Λ can be used to reformulate the agent's problem:

$$\max_a \mathbb{E}_t^a\left[\int_t^T e^{-r(s-t)}\left[u(w_{ijs},a_{ijs})+\lambda C(\tilde{\theta}^a_{ijs})\right]ds+e^{-r(T-t)}C(\tilde{\theta}^a_{ijT})\right]$$
$$=\max_a \mathbb{E}_t\left[\int_t^T e^{-r(s-t)}\Lambda_{ijt}\left[u(w_{ijs},a_{ijs})+\lambda C(\tilde{\theta}^a_{ijs})\right]ds+e^{-r(T-t)}\Lambda_{ijT}C(\tilde{\theta}^a_{ijT})\right] \tag{31}$$

Note the change in the expectation operator; the inclusion of Λ_{ijt} allows $\mathbb{E}_t^a$ and $\mathbb{E}_t$ to be interchanged. Using this representation, I solve the weak formulation of the agent's problem, where the choice of a_{ijt} corresponds to a choice of distribution P^a_{ijt} over Y_{ijt} (Cvitanic and Zhang, 2012). To derive necessary conditions for the incentive-compatibility of the efficient action path, I apply the stochastic maximum principle first proposed by Bismut (1973). The necessary conditions are summarized in the following theorem.

Proposition 2 (Necessary Conditions)**.** *Under the efficient strategy a^*, the CEO's promised value has equilibrium law of motion:*

$$dU_{ijt}=\left(rU_{ijt}-w_{ijt}-\lambda C(\tilde{\theta}_{ijt})\right)dt+\beta_{ijt}\sigma_W dZ_{ijt} \tag{32}$$

where β_{ijt} is a sensitivity process representing the incentives provided by the contract. The efficient strategy a^* is incentive compatible if:

$$\beta_{ijt}\geq \nu_{ijt}\Gamma_{ijt}+\phi_j \tag{33}$$

where $\Gamma_{ijt}=\frac{\partial U_t^a}{\partial\tilde{\theta}^a}$ is the CEO's information rent, the benefit of having marginally more optimistic beliefs relative to the board. Γ_{ijt} has equilibrium lower bound:

$$\Gamma_{ijt}\geq\frac{\phi_j}{r}\equiv\Gamma^*_{ijt} \tag{34}$$

which holds with equality when (33) *binds.*

Proof. See Appendix 8.2. □

Equation (32) is the standard representation of the CEO's promised value in continuous time, showing that U_{ijt} is an Ito process with respect to the standard Brownian Motion Z_{ijt}. U_{ijt} has drift $\left(rU_{ijt}-w_{ijt}-\lambda C(\tilde{\theta}_{ijt})\right)$, stating that the CEO's promised value accumulates at rate r net of the CEO's expected payoff $(w_{ijt}+\lambda C(\tilde{\theta}_{ijt}))$.[25] Promised value

[25]Speaking heuristically, prior to the realization of the retirement shock, the CEO's expected payoff over

has volatility $\beta_{ijt}\sigma_W$, measuring the sensitivity of the CEO's payoff to performance innovations dZ_{ijt}. The sensitivity process β_{ijt} is chosen implicitly by the board, and is the key instrument through which incentives are delivered to the CEO.

Condition (33) is the first-order counterpart of the incentive compatibility constraints (IC). The above theorem, however, says little about the participation constraint (IR). If the CEO's promised value falls below their outside option $C(\tilde{\theta}_{ijt})$, it is optimal for the CEO to leave the firm. The optimal compensation process ensures that $U_{ijT} = C(\tilde{\theta}_{ijT})$ while $U_{ijt} > C(\tilde{\theta}_{ijt})$ for all $t < T$, so CEO departure only occurs through firing when it is optimal for the board. Condition (33) implies that U_{ijt} changes with $\tilde{\theta}_{ijt}$ according to:

$$\frac{dU_{ijt}}{d\tilde{\theta}_{ijt}} = \frac{dU_{ijt}}{dY_{ijt}}\left(\frac{d\tilde{\theta}_{ijt}}{dY_{ijt}}\right)^{-1} = \frac{\beta_{ijt}}{\nu_{ijt}} \geq \frac{\phi_j}{\nu_{ijt}} + \Gamma_{ijt} \tag{35}$$

The condition (35) paired with the bound on the CEO's information rent (34) implies a lower bound on the CEO's continuation payoff U_{ijt} in any optimal contract. For a given level of tenure t and termination boundary $\theta_f(t)$, $U_{ijt} \equiv U(\tilde{\theta}_{ijt}, t, \theta_f(t))$ must exceed:

$$U_{ijt} = C(\theta_f(t)) + \int_{\theta_f(t)}^{\tilde{\theta}_{ijt}} \frac{dU_t}{d\tilde{\theta}_{ijt}} d\theta \tag{36}$$

$$\geq C(\theta_f(t)) + (\tilde{\theta}_{ijt} - \theta_f(t))\frac{\phi_j}{\nu_{ijt}} + \int_{\theta_f(t)}^{\tilde{\theta}_{ijt}} \Gamma_{ijt} d\theta \tag{37}$$

$$\geq C(\theta_f(t)) + (\tilde{\theta}_{ijt} - \theta_f(t))\left(\frac{\phi_j}{\nu_{ijt}} + \frac{\phi_j}{r}\right) \equiv U^*_{ijt} \tag{38}$$

The first inequality comes from (35), while the second comes from (34). U^*_{ijt} is thus the lower bound of the CEO's promised utility in any incentive-compatible contract implementing $a^*_{ijt} = 0$ for all t. The board seeks a compensation process w which minimizes U_{ijt} subject to the constraint $U_{ijt} \geq U^*_{ijt}$. Given this constraint, the following theorem presents the cost-minimizing compensation process.

Proposition 3 (Wage Determination)**.** *The firm's relaxed problem can be stated as follows. The board offers value U_{ijt} with volatility $\beta_{ijt}\sigma_W$ such that $U_{ijt} \geq C(\tilde{\theta}_{ijt})$ for all $t \leq T$, holding with equality only when $t = T$. The incentive-compatibility constraint* (33) *implies a lower*

interval Δt is:

$$e^{\lambda \Delta t} w_{ijt} + \lambda e^{\lambda \Delta t} C(\tilde{\theta}_{ijt}) + o(\Delta t)$$

The first two terms respectively represent the probability of zero and one retirement shocks arriving over interval Δt. The probability of > 1 shocks arriving is negligible, represented by the third term. Taking the limit as $\Delta t \to 0$ yields $w_{ijt} + \lambda C(\tilde{\theta}_{ijt}.)$

bound on U_{ijt} in equilibrium:

$$U_{ijt} \geq C(\theta_f(t)) + (\tilde{\theta}_{ijt} - \theta_f(t))\left(\frac{\phi_j}{\nu_{ijt}} + \frac{\phi_j}{r}\right) \equiv U^*_{ijt} \tag{39}$$

Thus, the board maximizes:

$$\mathbb{E}_t\left[\int_t^T e^{-r(s-t)} dX_{ijs} - \int_t^T e^{-r(s-t)} w_{ijs} ds + \lambda \int_t^T e^{-r(s-t)} V_T dt + e^{-r(T-t)}\left(V_T - b_j \pi\right)\right] \tag{40}$$

subject to $U_{ijt} \geq U^*_{ijt}$ and participation constraint (IR) *for all t. Given termination boundary $\theta_f(t)$, the cost-minimizing compensation process is given by:*

$$w_{ijt} = \rho\mu + \kappa_{1t}\tilde{\theta}_{ijt} + \kappa_{2t}\theta_f(t) \tag{41}$$

$$\kappa_{1t} = \phi_j\left(\frac{\rho}{r} + \frac{r}{\nu_{ijt}} + \frac{\nu_{ijt}}{r}\right) \tag{42}$$

$$\kappa_{2t} = \phi_j\left(1 - \frac{r}{\nu_{ijt}}\right) \tag{43}$$

Proof. See Appendix 8.2. □

The compensation process above ensures that $U_{ijt} = U^*_{ijt}$ for all t, so the board pays the CEO no more than is necessary to implement efficient actions. CEOs are compensated for their reputation according to the piece rate κ_{1t}. Additionally, the optimal level of compensation depends on the termination boundary $\theta_f(t)$. The coefficient κ_{2t} captures two conflicting effects of termination risk on the level of compensation. First, because CEOs prefer employment to unemployment, job security is valuable. Thus, CEOs must be compensated for decreases in job security (i.e. increases in $\theta_f(t)$) to maintain their equilibrium payoff. On the other hand, increasing the termination boundary decreases the CEO's incentive to deviate, as deviations increase the risk of job loss. Through this channel, financial incentives can be relaxed as the termination boundary rises. Which of these two effects dominates depends on the model's parameter values, so the question must ultimately be resolved empirically. What remains to be determined is the optimal termination boundary, which is summarized in the following theorem.

Proposition 4 (Second-Best Firing Threshold)**.** *When CEO actions are unobservable, the firing threshold which maximizes the firm's payoff is given by:*

$$\theta_f(t) = -r\pi + b_j^{-1}\left(\rho V_T - \frac{\nu^2_{ijt}\sigma^2_W}{2} V_{\theta\theta}(\theta_f(t), t, C(\theta_f(t))) - \frac{\beta^2_{ijt}\sigma^2_W}{2} V_{UU}(\theta_f(t), t, C(\theta_f(t)))\right) \tag{44}$$

Proof. See Appendix 8.2. □

The optimal firing threshold is derived by applying standard smooth-pasting and value-matching conditions to the firm's HJB equation. The second-best threshold retains many of the same features as the first-best counterpart. The key difference arises from the inclusion of the CEO's continuation payoff in the state. Given the CEO's limited liability, firms have limited ability to punish their CEO through financial means. Termination serves as an alternative method of punishment. In particular, if CEOs' continuation payoff falls to their outside option $C(\tilde{\theta}_{ijt})$, the board optimally fires and replaces the CEO. Thus, poor performance gradually drives down CEOs' continuation payoff, and in extreme cases results in the termination of their employment.

4 Identification and Estimation

Table 3: Summary of Model Parameters

Notation	Definition	Source of Identification:
θ_0	Mean of CEO quality distribution	Unconditional mean of ROA
δ_0	SD of CEO quality distribution	Unexplained variation in pay
σ_W	SD of profitability shocks	Within-CEO ROA variation
σ_γ	SD of firm productivity	Across-firm ROA variation
α	Rate of cash flow diversion	Correlation in pay & firm assets
μ	Outside option intercept	Unconditional mean of CEO pay
λ	Retirement arrival rate	Frequency of retirement
c	Monetary turnover cost	ROA variation around turnover
π	Non-pecuniary firing cost	Forced hazard rate

In Table 3, I summarize the model parameters and corresponding sources of identification. I fix the value of the discount rate $\rho = .05$ and estimate the remaining 9 parameters using the Simulated Method of Moments. I first discuss the identification of the parameters δ_0, α, and μ, whose key identifying information comes from the compensation data. Equation (41) shows that the unconditional expectation of model-implied compensation increases with μ. Because the mean of compensation is sensitive to the intercept term of CEOs' outside option, the empirical average of compensation is informative of μ. The second moment of the empirical compensation distribution, conditional on CEO tenure and firm size, is informative of the standard deviation of CEO quality δ_0. To see this directly, note that the optimal wage equation (41) implies that:

$$Var(w_{ijt}|b_j, t) = \gamma_{1t}^2 Var(\tilde{\theta}_{ijt}|t) \tag{45}$$

Thus, after conditioning on firm size and CEO tenure, the remaining variation in CEO compensation is attributed to variation in beliefs $\tilde{\theta}_{ijt}$. Unexplained variation in the compensation data therefore helps identify the standard deviation of the CEO quality distribution. Furthermore, defining $w^*_{ijt} \equiv w_{ijt} - \rho\mu$ as CEO compensation net of its intercept $r\mu$, equation (41) implies:[26]

$$log(w^*_{ijt}) = \alpha \, log(b_j) + log(f(\tilde{\theta}_{ijt}, t)) \tag{47}$$

which reveals that the diversion rate α is given exactly by the elasticity of w^*_{ijt} with respect to firm assets b_j. The quantity w^*_{ijt} is of course not observed in the data, so equation (47) is not feasible to estimate. Rather, I approximate it using the auxiliary model:

$$log(w_{ijt}) = \beta_0 + \beta_1 \, log(assets_{ijt}) + \beta_2 \, t_{ij} + \epsilon_{ijt} \tag{48}$$

The auxiliary parameters β_0 and β_1 are informative of μ and α, respectively. The variance of the fitted residual $\hat{\epsilon}_{ijt}$ carries information about the parameter δ_0. I thus match the parameters of equation (48) along with the variance of the fitted residuals $\hat{\epsilon}_{ijt}$ to recover the structural parameters μ, α, and δ_0.

The parameters σ_W, σ_γ, θ_0, and c are identified off of firm profitability data. I first discuss the disentangling of the volatility parameters σ_γ and σ_W. Under the assumption that both CEO quality θ_i and firm productivity γ_j are time-invariant, variation in ROA within a given employment spell is generated entirely by idiosyncratic shocks. Thus, within-spell variation in ROA helps to identify σ_W. Across-firm variation in ROA on the other hand is informative of σ_γ. Let y_{ijt} denote observed ROA, and let $\mathbb{E}_j$ and Var_j denote the mean and variance operators conditioned on firm j. To separately identify σ_γ and σ_W, I target the following two moments:

$$Var(\mathbb{E}_j[y_{ijt}]) \tag{49}$$

$$\mathbb{E}[Var_j(y_{ijt})] \tag{50}$$

The moment (50) is the within-firm variance of y_{ijt} averaged across all firms. This is informative of the idiosyncratic volatility σ_W. The moment (49) is the across-firm variance of

[26] Recalling that $\phi_j = b_j^\alpha$, the expression (47) is obtained from equation (41) by first subtracting $r\mu$ from both sides, pulling b_j^α to the front of the right hand side, and taking logs. The function $f(\tilde{\theta}_{ijt}, t)$ is given by:

$$f(\tilde{\theta}_{ijt}, t) = \left(\frac{\rho}{r} + \frac{r}{\nu_{ijt}} + \frac{\nu_{ijt}}{r}\right)\tilde{\theta}_{ijt} + \left(1 - \frac{r}{\nu_{ijt}}\right)\theta_f(t) \tag{46}$$

the within-firm average of ROA, and carries information about the standard deviation of firm productivity σ_γ. The mean of match quality θ_0 affects the mean of profitability, so is pinned down by the empirical average of ROA. The monetary cost of turnover, measured by c, is identified by variation in ROA around episodes of CEO turnover.

The remaining parameters, π and λ, are identified off of CEO turnover data. The probability of forced termination strictly decreases in the non-pecuniary cost of forced turnover π. The empirical forced termination rate thus carries information about the level of entrenchment. I target the parameters of the auxiliary model:

$$d_{ijt} = \lambda_0 + \lambda_1 \, tenure_{ijt} + \lambda_2 \, tenure_{ijt}^2 + \xi_{ijt} \tag{51}$$

where $d_{ijt} \in \{0,1\}$ is an indicator for forced turnover. Finally, the arrival rate of retirement shocks λ is recovered from the empirical rate of retirement. In total, I estimate the model using a 12×1 vector of moments denoted by $\hat{M}$. I obtain the optimal weighting matrix as the inverse of the covariance matrix of $\hat{M}$.[27]

The estimation algorithm proceeds as follows. Let $\Theta \in \mathbb{R}^9$ denote an arbitrary vector of structural parameters. Given Θ, I obtain the value function V by numerically solving the firm's HJB equation, from which the optimal termination boundary and compensation process can be computed. Given the optimal policies, I simulate 5000 firms 20 times each. Firms draw an initial CEO, and thereafter performance, beliefs, turnover, and compensation evolve as specified in the previous section. The simulation proceeds for 50 periods, where a period corresponds to a calendar year. Using the simulated data, I compute the same 12 moments as were computed in the empirical sample. If the simulated moments are sufficiently close to their empirical counterparts, the algorithm halts and returns the estimate $\hat{\Theta}$. Otherwise, a new candidate parameter vector is chosen and the procedure repeats. Standard errors for the estimates are computed based upon the asymptotic distribution of the SMM estimator as presented by Duffie and Singleton (1993). Details on model fit can be found in Appendix 8.4.

5 Results

5.1 Firm Profitability and Turnover Costs

The estimated standard deviation of the CEO quality distribution (σ_θ) is .033, or 3.3% of total assets per year (Table 4). This estimate in comparable to what has previously been

[27]See Appendix 8.3 for details on the computation of the optimal weighting matrix, estimation algorithm, and standard error computation.

found in the literature; Taylor (2010) for example reports an estimate of 2.42%. To give a dollar interpretation to the estimates of σ_θ and σ_γ, a one standard deviation increase in CEO quality implies a \$73.6 million dollar increase in average yearly cash flows for the median-sized firm in the sample.[28] The estimate of σ_γ, the standard deviation of firm productivity, implies a \$228.6 million increase in yearly cash flows resulting from a one standard deviation increase in productivity for a median-sized firm. Using the volatility estimates σ_θ, σ_γ, and σ_W, I decompose the variance of firm ROA: 6.8% and 65.4% of the variation of firm ROA can be respectively attributed to variation in CEO quality and firm productivity. The remaining 27.8% is attributed to idiosyncratic variation orthogonal to CEO and firm characteristics.[29]

Table 4: Structural Estimates

θ_0	δ_0	σ_W	σ_γ	λ	α	c	π	μ
.110	.033	.067	.103	.088	.476	.015	.097	12.2
(.001)	(.001)	($2.6e^{-4}$)	($1.5e^{-4}$)	(.001)	(.001)	(.008)	(.050)	(.274)

Notes: Standard errors are included in parenthesis.

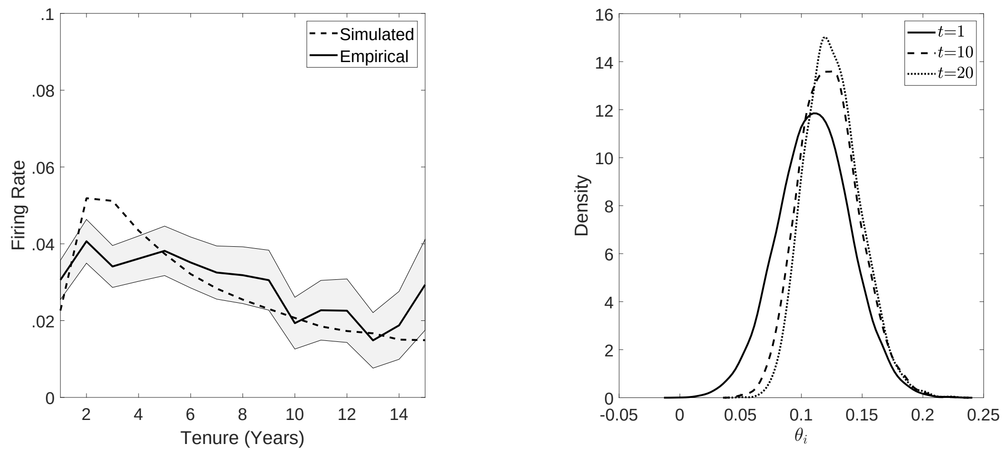
The model replicates well the low rate of forced CEO turnover in the data (Figure 5(a)). Such a low rate is explained in part by the high estimated turnover costs. The estimated monetary cost of turnover (c) is 1.5% of total firm assets (Table 4). For the median-sized firm, this amounts to roughly \$33.1 million. This cost reflects the expenses associated with finding a replacement CEO, severance payouts, and general disruptions to profitability resulting from onboarding new personnel. Compared to the monetary cost c, firms' utility cost of forced turnover (π) is estimated to be higher. While turnover entails a monetary cost of 1.5% of total assets, boards' effective cost of turnover is $(c+\pi)\times 100 = 11.2\%$ of total assets. This added non-monetary cost of forced turnover induces a substantial wedge separating the value-maximizing firing policy from firms' enacted policy, reflecting CEO entrenchment.

As employment progresses and uncertainty is resolved, low-quality CEOs are terminated. CEO quality is positively selected over tenure, leading its respective density to shift rightward as tenure increases (Figure 5(b)). The initial ($t = 1$) distribution of CEO quality, prior to any turnover decisions, is symmetric and matches the population distribution $N(\theta_0, \delta_0^2)$. Thus, among CEOs with one year of tenure, 50% are of below-average

[28] The median firm in the sample has approximately \$2.2 billion in assets.

[29] Interpreting the increment dy_{ijt} as firm ROA, under the assumption that CEO quality θ_i and firm productivity γ_j are independent we have that $Var(dy_{ijt}) = (\sigma_\theta^2 + \sigma_\gamma^2 + \sigma_W^2)dt$. The shares of the variation in ROA attibuted to each component $k \in \{\theta, \gamma, W\}$ are then computed as $\frac{\sigma_k^2}{\sigma_\theta^2+\sigma_\gamma^2+\sigma_W^2}$.

Figure 4: Forced Turnover and Positive Selection



(a) Firing Rate Fit

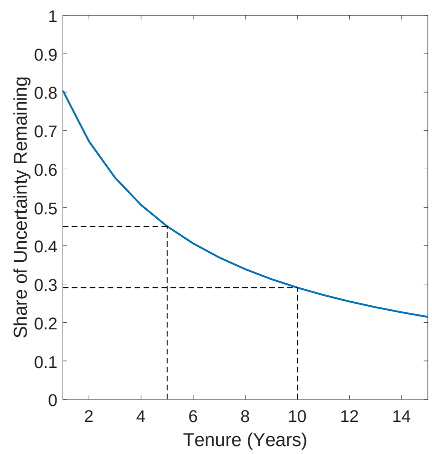
(b) Evolution of CEO Quality

Notes: Panel (a) compares the simulated and empirical firing rates. Solid lines denote the empirical forced turnover rate while dashed lines denote the simulated counterpart. The gray shaded region indicates the 95 percent confidence intervals around the empirical forced turnover rate. Panel (b) plots the simulated density of quality among CEOs who survive to tenure $t \in \{1, 10, 20\}$. All CEOs make it through one year of tenure, so the $t = 1$ distribution coincides with the population distribution $N(\theta_0, \delta_0^2)$. As tenure progresses, low-quality CEOs are filtered out, inducing attrition of CEOs with quality in the left tail of the distribution.

quality. As tenure increases, CEOs with quality in the left tail of the distribution are gradually forced out, concentrating the distribution around increasingly high levels of quality. Among CEOs with 10 years of tenure, the share of those with quality below the population mean θ_0 drops to 34.8%. It is striking that among the CEOs who have survived 10 years of filtering, roughly one in three of them are below average, suggesting that the rate at which firms filter out low-quality CEOs is fairly slow.

Two mechanisms generate this slow rate of selection. First, entrenchment increases the cost of terminating the CEO, protecting them from job loss. A firm may suspect that their CEO is of below-average quality, and thus would generate higher cash flows in expectation by drawing a replacement CEO, but turnover costs prevent the firm from doing so. Second, given the substantial uncertainty over CEO quality, firms may be unaware that their CEO is of relatively low quality. Given the costs of exercising their turnover option, firms prefer to wait to replace their CEO until they have sufficient certainty that their CEO is of low quality. The high estimate of σ_W suggests that ROA is a very noisy signal of θ_i, limiting firms' ability to make precise inferences about the quality of their CEO. To illustrate this, in Figure 5, I plot the share of quality uncertainty remaining over the first 15 years of CEO tenure. Consistent with Hamilton et al. (2023), the rate of learning is quite slow. Roughly 45% and 30% of the quality uncertainty remains after 5 and 10 years of tenure, respectively. The low rate of information acquisition leads firms to postpone their firing decisions, increasing the employment lengths of low-quality CEOs.

Figure 5: Speed of Learning



Notes: Figure 5 plots the share of uncertainty remaining over the first 15 years of CEO tenure. I define the share of quality uncertainty remaining at tenure t as $(\tilde{\delta}_{ijt}/\delta_0)^2$.

5.2 CEO Compensation

The key model-implied determinants of CEO compensation are CEO reputation, tenure, firm size, and outside options. The estimates along with parametric assumption (9) imply an outside option of \$43.6 million in net present value for a CEO of average quality.[30] This value is reasonable in light of the fact that CEOs are typically among the most highly-skilled and well-connected workers in the labor market (Rajgopal et al., 2006; Liu, 2014). This tightens the participation constraint, decreasing firms' share of the surplus generated over their CEO's employment. Relative to a lower-ranked employee, CEOs are thus quite expensive to retain, and the high level of CEO compensation observed in the data can in part be explained simply by CEOs' lucrative outside options. Moreover, outside options are sensitive to CEOs' reputation; a CEO employed in a median-sized firm will see a \$9.43 million dollar increase in their outside option in response to a one standard deviation increase in their perceived quality $\tilde{\theta}_{ijt}$.[31] This reward is substantial, and must be matched by employers to ensure that high-quality CEOs are not drawn out of their positions into the outside market.

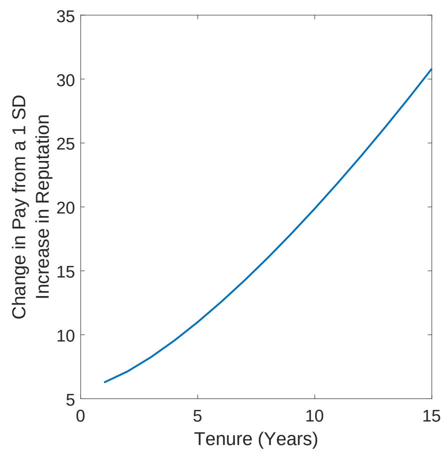
In Figure 6, I plot the model-implied increase in compensation resulting from a one standard deviation increase in CEO reputation. For early-tenured CEOs, the sensitivity of

[30] Setting firm assets to their median value and evaluating (9) at the estimates of θ_0, μ, α, and λ yields $C(\theta_0) = 43.6$.

[31] In the population, the standard deviation of the outside option $C(\tilde{\theta}_{ijt})$ is equal to $\frac{\phi_j \sigma_\theta}{r}$

compensation to their reputation is comparable to that of the outside market. However, this sensitivity quickly increases with CEO tenure. For a CEO with 10 years of tenure, the model implies a roughly \$20 million increase in total compensation resulting from a one standard deviation increase in perceived quality. Because CEOs at high levels of tenure are of relatively high quality on average, they face minimal risk of termination and thus must be motivated through financial means. This preserves incentive compatibility; as termination incentives weaken over tenure, financial incentives must increase in response.

Figure 6: CEO Reputation and Compensation



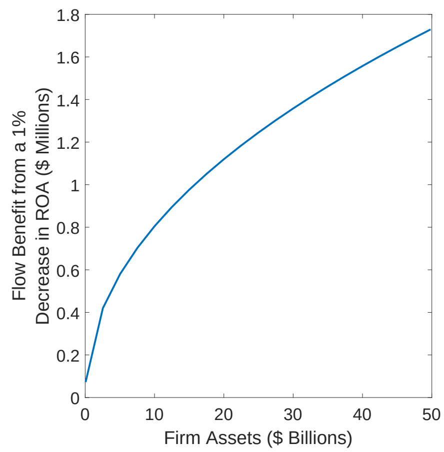
Notes: Figure 6 plots the model-implied increase in pay resulting from a one standard deviation (SD) increase in CEO reputation ($\tilde{\theta}_{ijt}$) as a function of tenure. This is obtained as the square root of the variance of optimal compensation (41) conditional on CEO tenure: $Var(w_{ijt}|t) = (\delta_{ijt}\gamma_{1t})^2$.

Lastly, as discussed extensively by Gayle and Miller (2009), firm size is another major source of variation CEO pay. As firm size increases, CEOs' private incentive to divert cash flows also increases. This is illustrated in Figure 7, which plots the CEO's private benefit associated with a 1% decrease in firm ROA. For example, a CEO employed by a firm with \$20 billion in assets would privately receive roughly \$1.12 million in response to decreasing the firm's ROA by 1%. However, this action would result in a \$200 million decrease in cash flows for such a firm, highlighting the extreme inefficiency of cash flow diversion. To prevent such inefficient behavior, large firms must increase compensation along the equilibrium path to preserve the incentive compatibility of efficient actions.

6 Moral Hazard and Entrenchment

In this section, I explore the economic burden of moral hazard, CEO entrenchment, and match quality uncertainty, analyzing in detail the impact of these three frictions on firm

Figure 7: Firm Size and the Rate of Cash Flow Diversion



Notes: Figure 7 plots the CEO's flow benefit of diverting 1% of firm assets towards their private consumption as a function of firm size. The incentive to divert cash flows grows with firm size, as does the inefficiency of such an action.

value and CEO employment dynamics. First, I compute firm value in the baseline case and compare it to the first-best case with no moral hazard. In all cases, firm value is obtained by simulating 5000 firms for 50 years each and computing the net present value (NPV) of cash flows net of CEO pay. Similarly, I compare firm value with and without CEO entrenchment. Results in Table 5 reveal substantial gains upon the elimination of entrenchment in both the baseline and first-best case.

Table 5: Moral Hazard, Entrenchment, and Firm Value

	Entrenchment ($\pi = .097$)	No Entrenchment ($\pi = 0$)	Net Present Value (NPV) % Change
2^{nd} Best (Hidden actions)	$22.36	$23.77	6.32%
1^{st} Best (Observable actions)	$22.66	$23.92	5.57%
NPV % Change	1.34%	.62%	6.98%

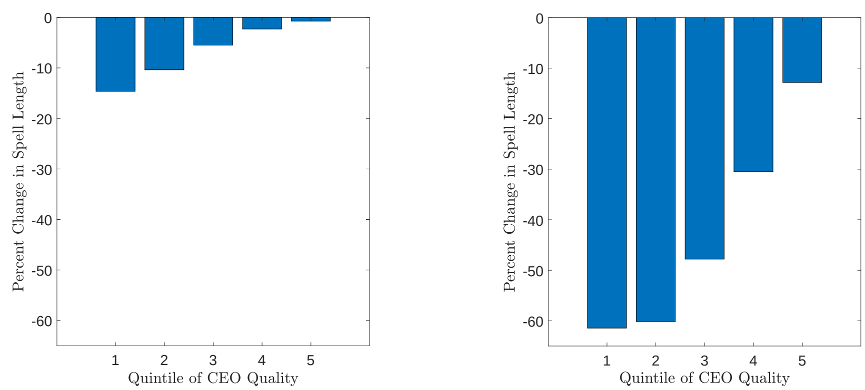
Notes: Table 5 reports firm value (in billions of dollars), computed as the net present value of cash flows net of CEO pay, in four separate environments. The top left cell is the baseline case with both moral hazard and entrenchment. The top right cell is a counterfactual environment with moral hazard, but no entrenchment. The second row contains the environments with no moral hazard, with and without entrenchment. Percent changes in NPV when eliminating entrenchment (moral hazard) are reported in the rightmost column (bottom row).

6.1 First-Best Case

In the first-best case, CEO actions are observed by the firm. Hence, there is no information asymmetry and moral hazard is eliminated. Table 5 shows that firm value increases by

1.34% on average upon the elimination of moral hazard. This increase in value is driven by two mechanisms. First, firms need not satisfy the incentive-compatibility constraint in the first-best case, so their compensation expense decreases substantially. Second, in the absence of moral hazard, low-quality CEOs are terminated quicker. To understand this second point, note that firms have two means of delivering value to their CEOs: direct compensation and job security. Expanding job security increases the CEO's expected length of employment, increasing their equilibrium payoff. In the presence of moral hazard, firms decrease their firing threshold, thereby increasing expected employment lengths, in exchange for decreased compensation. Thus, CEOs are employed longer on average in the second-best case relative to first best. I illustrate this in Figure 9(a), which plots the percent change in average spell lengths for CEOs upon the elimination of moral hazard, across the quality distribution.

Figure 8: Counterfactual Changes in Employment Length



(a) No Moral Hazard (b) No Entrenchment

Across the board, spell lengths decline upon the elimination of moral hazard. However, the effect is most pronounced for CEOs of relatively low quality. Low-quality CEOs are thus filtered out of employment more quickly in the absence of moral hazard, positively impacting firm value. However, the relative change in employment lengths is fairly small, given the substantial level of CEO entrenchment. In the baseline simulation, CEOs in the first quintile of quality are employed for 5.32 years on average. This drops to 4.54 years in the first-best case, less than a one year decline in employment length. While the elimination of moral hazard does improve CEO filtering, entrenchment substantially limits the realized improvements.

In Figure 10(a), I compare average compensation across the CEO quality distribution in the first and second-best cases of the model. In the absence of moral hazard, CEOs' participation constraints remain relevant while the incentive compatibility constraint becomes immaterial. Hence, compensation is specified to exactly offset CEOs' outside option. As such, compensation falls substantially, with high-quality CEOs bearing the majority of the loss in pay. Additionally, the gradient of compensation with respect to CEO quality also falls. In the first-best case, compensation increases with perceived quality only through its dependence on the outside option. Thus, the CEO quality premium falls substantially. The figure highlights that relative to the case of perfect monitoring, high-quality CEOs are substantially more expensive to retain when actions are not observable by boards.

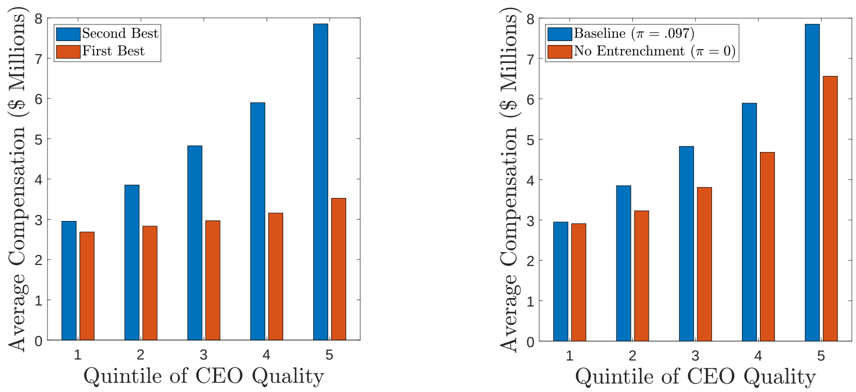
6.2 CEO Entrenchment

Because firms' non-pecuniary cost of forced turnover (π) is large, entrenchment has major consequences for CEO employment dynamics. Entrenchment decreases the firing boundary, increasing the expected length of CEO employment and decreasing the strength of termination incentives. Table 5 shows that upon the elimination of entrenchment, firm value increases by 6.32% on average, significantly outweighing the gains generated when eliminating moral hazard. There are two main channels through which entrenchment harms firm value. First, entrenchment slows the rate at which low-quality CEOs are terminated. Low-quality CEOs remain employed longer in the presence of entrenchment, negatively impacting performance. Second, entrenchment weakens termination incentives, increasing the necessary amount of compensation to satisfy incentive compatibility. This weakening of termination incentives exacerbates the cost of moral hazard, as can be seen in Table 5; the gains in NPV from eliminating moral hazard drop substantially in the absence of entrenchment.

Notably, the effect of entrenchment on CEO employment lengths is not uniform across the distribution of quality (Figure 9(b)). CEOs in the left tail of the quality distribution see a roughly 60% decline in length of employment on average. This enhances firm value, as low-quality CEOs are employed for shorter spells. CEOs in the right tail see a decline in employment lengths of roughly 10%. Firms increase their firing boundary, strengthening termination incentives, which thereby decreases the necessary amount of compensation in equilibrium. While high quality CEOs see the smallest changes in employment lengths, they see the largest changes in compensation upon the elimination of entrenchment (Figure 10(b)).

As discussed previously, termination incentives and monetary incentives act as substitutes. In the absence of entrenchment, CEOs' risk of job loss is substantially higher. This termination risk incentivizes the CEO to remain on path, decreasing the level of compen-

Figure 9: Counterfactual Changes in Compensation



(a) No Moral Hazard

(b) No Entrenchment

sation needed to satisfy the IC constraint. Entrenchment mitigates termination incentives, which increases the equilibrium level of pay, particularly for high-quality CEOs. Thus, the high level of compensation observed in the data can be attributed in part to the weakening of incentives induced by entrenchment. Because entrenchment weakens termination incentives, it is more costly for firms to align the incentives of their CEOs in the presence of entrenchment relative to the case with no entrenchment. Table 5 shows that with entrenchment, the gains from eliminating moral hazard more than double relative to the case with no entrenchment. Thus, decreased CEO entrenchment, brought upon for instance by improved corporate governance, is one channel through which the severity of moral hazard can be mitigated.

7 Conclusion

This paper shines light on the impact of turnover frictions on dynamic managerial incentives. The theoretical model, motivated by a set of empirical facts, provides a tractable framework in which moral hazard, entrenchment, and reputation can be studied comprehensively. The results highlight the substitutability of financial incentives and termination incentives when motivating managers. Entrenchment increases the cost of CEO replacement, weakening termination incentives and increasing the level of compensation needed to align incentives between shareholders and the CEO. The model predicts a considerable decrease in managerial pay upon the elimination of entrenchment.

Counterfactual experiments show that entrenchment more than doubles the cost of moral hazard. A practical takeaway is that one remedy for the moral hazard problem is to lessen to extent of managerial entrenchment. This can be achieved, for example, by refraining from the use of anti-takeover tactics such as poison pills, or precluding constitutional limits on shareholder voting power. Such practices add frictions to the CEO replacement process, weakening managerial incentives and decreasing firm value.

There are several extensions of this paper which may be fruitful avenues for future research. First, rather than taking the level of entrenchment as given, one could treat the level of entrenchment as a contractible object. This is similar to work by Grochulski et al. (2020), who analyze the effects of golden-parachute type mechanisms on the dynamic allocation of incentives. Additionally, entrenchment may extend CEOs' time horizons, mitigating the extent of managerial short-termism. Estimating the impact of entrenchment on CEOs' willingness to invest in long-term projects, R&D endeavors for instance, may be an interesting task for future work. In addition, while this paper restricts attention to the market for executives, the framework considered here may be insightful when applied to other labor markets. One could, for example, apply a similar model to study the impact of teacher tenure on student outcomes. This may be an interesting and important task for future research.

References

Ai, Hengjie, Dana Kiku, and Rui Li, "A Quantitative Model of Dynamic Moral Hazard," *The Review of Financial Studies*, 08 2022, *36* (4), 1408–1463.

Allgood, Sam and Kathleen A. Farrell, "The Match between CEO and Firm," *The Journal of Business*, 2003, *76* (2), 317–341.

Barry, John, "Shareholder Voice and Executive Compensation," *Working Paper*, 2023.

Bebchuk, Lucian, Alma Cohen, and Allen Ferrell, "What matters in corporate governance?," *The Review of financial studies*, 2009, *22* (2), 783–827.

Bennedsen, Morten, Francisco Perez-Gonzalez, and Daniel Wolfenzon, "Do CEOs Matter? Evidence from Hospitalization Events," *The Journal of Finance*, 2020, *75* (4), 1877–1911.

_, Kasper Meisner Nielsen, Francisco Perez-Gonzalez, and Daniel Wolfenzon, "Inside the Family Firm: The Role of Families in Succession Decisions and Performance," *The Quarterly Journal of Economics*, 2007, *122* (2), 647–691.

Berger, Philip G., Eli Ofek, and David L. Yermack, "Managerial Entrenchment and Capital Structure Decisions," *The Journal of Finance*, 1997, *52* (4), 1411–1438.

Bertrand, Marianne and Antoinette Schoar, "Managing with Style: The Effect of Managers on Firm Policies," *The Quarterly Journal of Economics*, 2003, *118* (4), 1169–1208.

Biais, Bruno, Thomas Mariotti, Jean-Charles Rochet, and Stéphane Villeneuve, "Large Risks, Limited Liability, and Dynamic Moral Hazard," *Econometrica*, 2010, *78* (1), 73–118.

Bismut, Jean-Michel, "Conjugate convex functions in optimal stochastic control," *Journal of Mathematical Analysis and Applications*, 1973, *44* (2), 384–404.

Black, Fischer and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 1973, *81* (3), 637–654.

Brandimarte, P., *Numerical Methods in Finance and Economics: A MATLAB-Based Introduction* 04 2006.

Chang, Xin and Hong Feng Zhang, "Managerial Entrenchment and Firm Value: A Dynamic Perspective," *The Journal of Financial and Quantitative Analysis*, 2015, *50* (5), 1083–1103.

Coles, Jeffrey L., Naveen D. Daniel, and Lalitha Naveen, "Managerial incentives and risk-taking," *Journal of Financial Economics*, 2006, *79* (2), 431–468.

Core, John and Wayne Guay, "Estimating the Value of Employee Stock Option Portfolios and Their Sensitivities to Price and Volatility," *Journal of Accounting Research*, 2002, *40* (3), 613–630.

Cvitanic, Jakša and Jianfeng Zhang, *Contract theory in continuous-time models*, Springer Science & Business Media, 2012.

Cvitanic, Jaksa, Xuhu Wan, and Jianfeng Zhang, "Optimal Compensation with Hidden Action and Lump-Sum Payment in a Continuous-Time Model," *Applied Mathematics and Optimization*, 02 2009, *59.*

DeMarzo, Peter M. and Michael J. Fishman, "Optimal Long-Term Financial Contracting," *The Review of Financial Studies*, 09 2007, *20* (6), 2079–2128.

Demarzo, Peter M. and Yuliy Sannikov, "Learning, Termination, and Payout Policy in Dynamic Incentive Contracts," *The Review of Economic Studies*, 07 2017, *84* (1), 182–236.

Duffie, Darrell and Kenneth J. Singleton, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 1993, *61* (4), 929–952.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter, "Executive Compensation: A Survey of Theory and Evidence," 2017, *1*, 383–539.

Ferraro, Valeria, "Media Focus and Executive Turnover: Consequences for Female Leadership," *Working Paper*, 2021.

Gayle, George-Levi and Robert A. Miller, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?," *American Economic Review*, December 2009, *99* (5), 1740–69.

_, Limor Golan, and Robert A. Miller, "Promotion, Turnover, and Compensation in the Executive Labor Market," *Econometrica*, 2015, *83* (6), 2293–2369.

Gibbons, Robert and Kevin Murphy, "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 02 1992, *100*, 468–505.

Gompers, Paul, Joy Ishii, and Andrew Metrick, "Corporate Governance and Equity Prices*," *The Quarterly Journal of Economics*, 02 2003, *118* (1), 107–156.

Grochulski, Borys and Yuzhe Zhang, "Termination as an incentive device," *Theoretical Economics*, 2023, *18* (1), 381–419.

_, Russell Wong, and Yuzhe Zhang, "Optimal incentive contracts with job destruction risk," *Federal Reserve Bank of Richmond Working Papers*, 2020.

Hamilton, Barton, Andrés Hincapié, and Noah Lyman, "The Bright Side to Nepotism? Family CEOs, Turnover, and Firm Performance," *Working Paper*, 2023.

Hermalin, Benjamin and Michael Weisbach, "Endogenously Chosen Boards of Directors and Their Monitoring of the CEO," *The American Economic Review*, 1998, *88* (1), 96–118.

Hermalin, Benjamin E., "Trends in Corporate Governance," *The Journal of Finance*, 2005, *60* (5), 2351–2384.

Hu, Aidong and Praveen Kumar, "Managerial Entrenchment and Payout Policy," *Journal of Financial and Quantitative Analysis*, 2004, *39* (4), 759–790.

Jenter, Dirk and Fadi Kanaan, "CEO Turnover and Relative Performance Evaluation," *The Journal of Finance*, 2015, *70* (5), 2155–2184.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 1979, *87* (5), 972–90.

_ and Julien Prat, "Dynamic contracts when the agent's quality is unknown," *Theoretical Economics*, 2014, *9* (3), 865–914.

Laffont, Jean-Jacques and Jean Tirole, "The Dynamics of Incentive Contracts," *Econometrica*, 1988, *56* (5), 1153–1175.

Lehn, Kenneth, Sukesh Patro, and Mengxin Zhao, "Governance indexes and valuation: Which causes which?," *Journal of Corporate Finance*, 2007, *13* (5), 907–928.

Lippi, Francesco and Fabiano Schivardi, "Corporate control and executive selection," *Quantitative Economics*, 2014, *5* (2), 417–456.

Liu, Yun, "Outside options and CEO turnover: The network effect," *Journal of Corporate Finance*, 2014, *28* (C), 201–217.

Mackey, Alison, "The Effect of CEOs on Firm Performance," *Strategic Management Journal*, 2008, *29* (12), 1357–1367.

Margiotta, Mary M and Robert A Miller, "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 2000, *41* (3), 669–719.

Merton, Robert C., "Theory of Rational Option Pricing," *The Bell Journal of Economics and Management Science*, 1973, *4* (1), 141–183.

Miller, Robert A., "Job Matching and Occupational Choice," *Journal of Political Economy*, 1984, *92* (6), 1086–1120.

Moscarini, Giuseppe, "Job Matching and the Wage Distribution," *Econometrica*, 2005, *73* (2), 481–516.

Page, T. Beau, "CEO attributes, compensation, and firm value: Evidence from a structural estimation," *Journal of Financial Economics*, 2018, *128* (2), 378–401.

Parrino, Robert, "CEO Turnover and Outside Succession, A Cross-Sectional Analysis," *Journal of Financial Economics*, 1997, *46* (2), 165–197.

Peréz-González, Francisco, "Inherited Control and Firm Performance," *American Economic Review*, 2006, *96* (5), 1559–1588.

Peters, Florian S. and Alexander F. Wagner, "The Executive Turnover Risk Premium," *The Journal of Finance*, 2014, *69* (4), 1529–1563.

Rajgopal, Shivaram, Terry Shevlin, and Valentina Zamora, "CEOs' outside Employment Opportunities and the Lack of Relative Performance Evaluation in Compensation Contracts," *The Journal of Finance*, 2006, *61* (4), 1813–1844.

Sannikov, Yuliy, "A Continuous- Time Version of the Principal: Agent Problem," *The Review of Economic Studies*, 2008, *75* (3), 957–984.

Shleifer, Andrei and Robert W. Vishny, "Managerial Entrenchment: The Case of Manager-Specific Investments," *Journal of Financial Economics*, 1989, *25* (1), 123–139.

Spear, Stephen and Cheng Wang, "When to fire a CEO: optimal termination in dynamic contracts," *Journal of Economic Theory*, 2005, *120* (2), 239–256.

Stiglitz, Joseph E. and Andrew Weiss, "Incentive Effects of Terminations: Applications to the Credit and Labor Markets," *The American Economic Review*, 1983, *73* (5), 912–927.

Taylor, Lucian, "Why Are CEOs Rarely Fired? Evidence from Structural Estimation," *The Journal of Finance*, 2010, *65* (6), 2051–2087.

Williams, Noah, "Persistent Private Information," *Econometrica*, 2011, *79* (4), 1233–1275.

Zhu, John Y., "Optimal Contracts with Shirking," *The Review of Economic Studies*, 2013, *80* (2 (283)), 812–839.

Zwiebel, Jeffrey, "Dynamic Capital Structure under Managerial Entrenchment," *The American Economic Review*, 1996, *86* (5), 1197–1215.

8 Appendix

8.1 Data Appendix

Execucomp I obtain data on CEO pay and tenure from Execucomp. Each CEO-firm match is uniquely identified by the variable *co_per_rol*. The key compensation variable I use in estimation is *tdc1*, defined as "Total Compensation (Salary + Bonus + Other Annual + Restricted Stock Grants + LTIP Payouts + All Other + Value of Option Grants)." I convert *tdc1* to millions of dollars in estimation; all nominal variables are denoted in 2015 dollars. I winsorize the distribution of *tdc1* at its 1st and 99th percentiles.

Compustat I obtain company fundamentals data from Compustat North America, which contains a rich set of financial information on publicly held companies in Canada and the U.S. Each firm is uniquely identified by the variable *gvkey*. Using operating income before depreciation (item *oibdp*) and total assets (item *at*) I compute return on assets (ROA_{ijt}) for each firm-year as:

$$ROA_{ijt} = \frac{oibdp_{ijt}}{at_{ijt}}$$

I drop firms with values of ROA_{ijt} outside of the range $[-1, 1]$ (70 observations). Industries are defined using the Global Industry Classification Standard (GICS) codes, corresponding to the Compustat variable *gind*.

Forced turnover data Data on forced CEO turnover was graciously shared by Florian Peters. He and a team of researchers gathered these data for CEOs listed in Execucomp from years 1995 to 2015. The criteria used to classify turnover as forced are described in detail in Peters and Wagner (2014) and Jenter and Kanaan (2015). Both methodologies follow the three-step criteria to classify successions as forced from Parrino (1997):

1. "All successions for which the *Wall Street Journal* reports that the CEO is fired, forced from the position, or departs due to unspecified policy differences are classified as forced."
2. "All other successions in which the departing CEO is under age 60 are reviewed to identify cases in which the *Wall Street Journal* announcement of the succession either (1) does not report the reason for departure as involving death, poor health, or the acceptance of another position (elsewhere or within the firm), or (2) reports that the CEO is retiring, but does

not announce the retirement at least six months before the succession. These cases are also classified as forced successions."

3. "The circumstances surrounding departures that are classified as forced in the previous step are further investigated by searching the business and trade press for relevant articles. These successions are reclassified as voluntary if the incumbent takes a comparable position elsewhere or departs for previously undisclosed personal or business reasons that are unrelated to the firm's activities."

If turnover is not classified as forced in Florian Peters' data, it is assumed to be voluntary. For a small number of cases, forced turnover is reported in year t, but the executive is still listed as CEO in year $t+1$. To avoid inconsistencies, all indicators of turnover are moved to the last year of the CEOs tenure as reported in Execucomp. In my final sample, I observe 908 instances of forced turnover and 2,667 instances of voluntary turnover.

8.1.1 Profitability and CEO Tenure

Table 6: Profitability and CEO Tenure

	Profitability and CEO Tenure			
	(1)		(2)	
	Coefficient	SE	Coefficient	SE
	CEO Characteristics:			
Tenure	.059***	(.008)	.181	(.183)
Age	-.009	(.009)	-.036	(.080)
Female	.085	(.273)	-	-
	Firm Characteristics:			
Log(Assets)	-4.61***	(.062)	-8.71***	(.534)
Log(Revenue)	5.94***	(.087)	11.7***	(.654)
	Fixed Effects:			
Year	✓		✓	
CEO-Firm Match			✓	
Observations	41,415		41,415	

Notes: Column (1) reports pooled OLS estimates while column (2) reports within-match estimates. CEO gender is omitted from column (2) since this is fixed within match. The tenure effect disappears within match.

In principle, profitability may rise with tenure as a result of learning by doing on part of the CEO. To test this, I report in Table 6 estimates from an ROA regression with and without CEO-firm match effects. Across matches, there is a positive and significant relationship between CEO tenure and firm performance. However, the tenure effect vanishes within match. This is evidence favoring selection on CEO quality as the key determinant of the tenure-profitability relationship as opposed to learning by doing.

8.1.2 Turnover and Cumulative CEO Performance

Table 7: Profitability and CEO Tenure

	Forced Turnover		Voluntary Turnover	
	Marginal Effect	SE	Marginal Effect	SE
	CEO Reputation:			
$\tilde{\theta}_{ijt}$ (Standardized)	-.003***	(.001)	-3.5e^{-4}	(.002)
	CEO Characteristics:			
Tenure	-.001***	(1.4e^{-4})	-.001***	(1.7e^{-4})
Age	-4.1e^{-4}***	(1.1e^{-4})	.007***	(2.2e^{-4})
Female	.008**	(.004)	-.016**	(.008)
	Firm Characteristics:			
Log(Assets)	-.002**	(.001)	-.005***	(.001)
Log(Revenue)	.002	(.001)	.006***	(.002)
Observations	41,202		41,402	

Notes: Columns (1) and (2) respectively report marginal effects obtained from a logit regression of forced and voluntary turnover indicators on vectors of CEO and firm characteristics.

Taylor (2010) and Hamilton et al. (2023) suggest that firms make CEO replacement decisions in response to new information about CEO quality. To proxy for firms' evolving information set, I define the *adaptive* shrinkage estimator $\tilde{\theta}_{ijt}$ as the best estimate of θ_{ij} given information at time t. Concretely, $\tilde{\theta}_{ijt}$ is defined as:

$$\tilde{\theta}_{ijt} = \frac{\tilde{\theta}_{ijt-1} + \omega_j \hat{\epsilon}_{ijt}}{1 + \omega_j} \tag{52}$$

$\tilde{\theta}_{ijt}$ is the cumulative weighted average of the performance residuals $\hat{\epsilon}_{ijt}$ implied by estimating Equation (1). I refer to the quantity $\tilde{\theta}_{ijt}$ as CEO i's *reputation* with firm j at time t.[32]

Table 7 reports marginal effects obtained from regressing forced and voluntary turnover indicators on sets of CEO and firm characteristics, along with the reputation proxy $\tilde{\theta}_{ijt}$. The probability of forced turnover significantly declines in response to positive cumulative performance, as proxied by $\tilde{\theta}_{ijt}$. On the other hand, the probability of voluntary

[32] $\tilde{\theta}_{ijt}$ only conditions on performance residuals up to time t, whereas the James-Stein estimator $\hat{\theta}_{ij}$ conditions on the CEO's complete performance history. The terminal value of $\tilde{\theta}_{ijt}$ is the baseline James-Stein estimate $\hat{\theta}_{ij}$.

turnover is unaffected by cumulative performance. In light of these results, I assume in the model section that voluntary turnover is independent of CEOs' reputation.

8.2 Model Appendix

Throughout the model appendix, I suppress the ij subscripts to ease notational burden.

8.2.1 Belief Manipulation

Suppose that at tenure t the firm recommends $a^*_{ijt} = 0$, but the CEO deviates to $\hat{a}_{ijt} > 0$. The firm, assuming the CEO took the recommended action, updates beliefs according to:

$$\begin{aligned} d\tilde{\theta}_{ijt} &= \nu_{ijt}(dY_{ijt} - \tilde{\theta}_{ijt}dt)d\tilde{\theta}_{ijt} \\ &= \nu_{ijt}\sigma_W dZ_{ijt} \end{aligned} \tag{53}$$

The CEO, knowing their true action choice, will update according to:

$$\begin{aligned} d\tilde{\theta}^a_{ijt} &= \nu_{ijt}(dY_{ijt} - (\tilde{\theta}_{ijt} - \hat{a}_{ijt})dt) \\ d\tilde{\theta}^a_{ijt} &= \nu_{ijt}\sigma_W dZ^a_{ijt} \end{aligned} \tag{54}$$

where $dZ^a_{ijt} = \sigma_W^{-1}(dY_{ijt} - (\tilde{\theta}_{ijt} - \hat{a}_{ijt})dt)$ is the innovation process observed by the CEO. The asymmetry in information induces a discrepancy in the incremental belief update:

$$d\tilde{\theta}^a_{ijt} - d\tilde{\theta}_{ijt} = \hat{a}_{ijt}\nu_{ijt}dt > 0 \tag{55}$$

The CEO, who is perfectly informed of their action choices, can always update beliefs "correctly" in the sense that their estimate of θ_i is unbiased. The firm on the other hand lacks knowledge of the CEO's action choices, and instead assumes that the recommended action a^*_{ijt} has been selected. The firm thus misinterprets the information generated under a deviation. In particular, positive deviations decrease realized performance relative to the firm's expectations $(\tilde{\theta}_{ijt} - a^*_{ijt})dt$, leading the firm's beliefs to drift downwards relative to those of the CEO. Figure 10 illustrates an example.

In essence, the firm faces an identification problem in which the effects of CEO quality and actions on performance cannot be disentangled, since neither are observable. The CEO can take advantage of this by deviating and inducing a gap in beliefs between firm and CEO. This benefits the CEO because the optimal contract will reward them if performance surpasses the firm's expectations. The lower the firm's expectations relative to the CEO's, the higher the probability that realized performance exceeds expectations. Hence,

Figure 10: Belief Manipulation

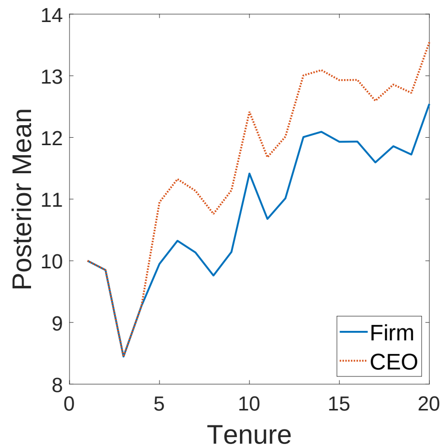


Figure 10 illustrates the gap in beliefs induced following a deviation by the CEO. I simulate a sample path of profitability and record the corresponding beliefs. In this simulation, the firm recommends $a_{ijt} = 0$ for all t. The CEO deviates to $\hat{a}_{ijt} = .1$ at $t = 5$. This leads performance to fall short (on average) of the firm's expectations, decreasing the firm's beliefs relative to the CEO's. The CEO is informed of their action choice, so updates beliefs accurately.

manipulating the firm's beliefs downwards increases the CEO's likelihood of accumulating rewards later in their employment spell. The gap in beliefs $\alpha_{ijt} = \tilde{\theta}^a_{ijt} - \tilde{\theta}_{ijt}$ has law of motion:

$$d\alpha_{ijt} = \nu_{ijt}((\hat{a}_{ijt} - a^*_{ijt}) - \alpha_{ijt})dt \tag{56}$$

Note that in the absence of further deviations, the gap in beliefs converges to zero at rate ν_{ijt}.

8.2.2 Proofs of Key Theorems

Theorem 1. *Given any IC contract $\mathscr{C} = (a, c, T)$ recommending $a_t > 0$ for some t, there is an alternative contract $\mathscr{C}'$ recommending $\{a'_t = 0\}_{t\geq 0}$ in which the CEO's payoff is unchanged and the firm's payoff is weakly greater.*

Proof. This proof is essentially a restatement of Lemma A in Demarzo and Sannikov (2017). Let ω_t denote an arbitrary sample path of profitability Y_{ijt} up until tenure t. The compensation process c_t and stopping time T under the original contract $\mathscr{C}$ map from sample paths to $\mathbb{R}_+$:

$$c_t\Big(Y_s; s \in [0,t]\Big) : \omega_t \to \mathbb{R}_+$$
$$T\Big(Y_s; s \in [0,t]\Big) : \omega_t \to \mathbb{R}_+$$

Consider the alternative contract $\mathscr{C}' = (a', c', T')$ with compensation and stopping time defined by:

$$c'_t \equiv c_t\left(Y_s - \int_0^s a_l dl\,;\, s \in [0,t] \right) + \phi \int_0^t a_s ds$$
$$T' \equiv T\left(Y_s - \int_0^s a_l dl; s \in [0,t] \right)$$

which adjust compensation and the stopping time according to the cash flows the CEO would have diverted given original action recommendation $\{a_t > 0\}$. If the CEO selected strategy $\{a'_t \geq 0\}$ under contract $\mathscr{C}'$, their flow payoff would be identical to their payoff under $\mathscr{C}$ and $\{a'_t + a_t\}$ for a given path ω_t. Furthermore, under contract $\mathscr{C}$ and strategy $\{a'_t + a_t\}$, the path of:

$$Y_t = Y_0 + \int_0^t (\tilde{\theta}^a_{ijt} - a'_s - a_s)ds + \sigma_W \int_0^t dZ_s \tag{57}$$

coincides with the path of:

$$Y_t = Y_0 + \int_0^t (\tilde{\theta}^a_{ijt} - a'_s)ds + \sigma_W \int_0^t dZ_s \tag{58}$$

under contract $\mathscr{C}'$ and strategy $\{a'_t\}$. Given the definition of c'_t, the CEO's flow payoff under $\mathscr{C}'$ is greater by $\phi a_t dt$, the amount they would have diverted, so their payoff under both contracts and respective strategies are identical. The incentive compatibility of original contract $\mathscr{C}$ implies that $\{a'_t = 0\}$ is optimal for the CEO under $\mathscr{C}'$. Additionally, given $\{a_t > 0\}$ under $\mathscr{C}$ the firm's payoff strictly increases under $\mathscr{C}'$ as long as $\phi = b^\alpha < b$, which holds when $\alpha < 1$. □

Proposition 1. Given state $(\tilde{\theta}_{ijt}, t)$, the value of the firm continuing with their current CEO is:

$$V(\tilde{\theta}_{ijt}, t) = b\tilde{\theta}_{ijt}\Delta t + e^{-r\Delta t}\mathbb{E}[dV(\tilde{\theta}_{ijt+\Delta t}, t+\Delta t)] + \lambda e^{-r\Delta t} V_T + o(\Delta t) \tag{59}$$

Taking the limit as $\Delta t \to 0$ and applying Ito's lemma:

$$rV(\tilde{\theta}, t) = b\tilde{\theta} + \underbrace{V_t(\tilde{\theta}, t) + \frac{\nu_t^2 \sigma_W^2}{2} V_{\theta\theta}(\tilde{\theta}, t)}_{\mathbb{E}[dV(\tilde{\theta}_{ijt+\Delta t}, t+\Delta t)]/\Delta t} + \lambda V_T \tag{60}$$

(60) is the firm's Hamilton-Jacobi-Bellman (HJB) equation, the standard recursive repre-

sentation of the value function in continuous time. To derive the firing threshold, I apply the value matching and smooth pasting conditions along the stopping boundary:

$$V(\theta_f^{FB}(t), t) = V_T - b\pi \tag{61}$$

$$V_t(\theta_f^{FB}(t), t) = \frac{\partial V_T}{\partial t} = 0 \tag{62}$$

Inserting these conditions into the HJB equation (60) evaluated at $\theta_f^{FB}(t)$ yields:

$$\theta_f^{FB}(t) = -r\pi + b^{-1}\left(\rho V_T - \frac{\nu_{ijt}^2 \sigma_W^2}{2} V_{\theta\theta}(\theta_f^{FB}(t), t)\right) \tag{63}$$

□

Proposition 2. I apply the stochastic maximum principle of Bismut (1973) to derive necessary conditions. For the Hamiltonian function H and pair of states (Λ_t, α_t), define the corresponding adjoint processes $(p_t^\Lambda, p_t^\alpha)$ as the solutions to the backward SDEs:

$$dp_t^\Lambda = rp_t^\Lambda dt - H_\Lambda dt + q_t^\Lambda dZ_t \tag{64}$$

$$dp_t^\alpha = rp_t^\alpha dt - H_\alpha dt + q_t^\alpha dZ_t \tag{65}$$

where $(q_t^\Lambda, q_t^\alpha)$ are the volatility processes corresponding to the adjoint pair $(p_t^\Lambda, p_t^\alpha)$. Conditions (64) and (65) are the stochastic counterparts of the deterministic Pontryagin's maximum principle. The CEO's (current value) Hamiltonian reads:

$$H(t, \Lambda, \alpha, a, p^\Lambda, p^\alpha, q^\Lambda, q^\alpha) = \Lambda\left(w + \phi a + \lambda C(\tilde{\theta}_t^a)\right) + p^\alpha(\nu(a - \alpha)) + q^\Lambda \Lambda \frac{\alpha - a}{\sigma_W} \tag{66}$$

Incentive compatibility of the first-best action $a_t = 0$ requires $\frac{\partial H}{\partial a} \leq 0$. Differentiating the Hamiltonian yields:

$$\Lambda_t \phi + p_t^\alpha \nu_t - \frac{q_t^\Lambda}{\sigma_W} \Lambda_t \leq 0 \tag{67}$$

(67) is the necessary condition for the incentive compatibility of the efficient strategy a^*. Before discussing the condition's economic interpretation, I'll first find expressions for the multipliers p_t^α and q_t^Λ. Under $\{a_t = 0\}$, $\Lambda_t = 1$ and $\alpha_t = 0$:

$$dp_t^\Lambda = \left(rp_t^\Lambda - w_t - \lambda C(\tilde{\theta}_t)\right)dt + q_t^\Lambda dZ_t \tag{68}$$

$$dp_t^\alpha = (r + \nu_t)p_t^\alpha dt - \Lambda \frac{q_t^\Lambda}{\sigma_W} dt + q_t^\alpha dZ_t \tag{69}$$

with terminal values $p_T^\Lambda = C(\tilde{\theta}_T)$ and $p_T^\alpha = C_\alpha(\tilde{\theta}_T)$. Letting $\beta_t \equiv \frac{q_t^\Lambda}{\sigma_W}$, the pair (p_t^Λ, β_t) is a weak solution to (68) where:

$$p_t^\Lambda = \mathbb{E}_t\left[\int_t^T e^{-r(s-t)} w_s ds + \lambda \int_t^T e^{-r(s-t)} C(\tilde{\theta}_s) ds + e^{-r(T-t)} C(\tilde{\theta}_T)\right] \tag{70}$$

Notice that $p_t^\Lambda = U_t$, so the CEO's continuation payoff U_t becomes a state variable of the firm's problem. This is a standard result in the recursive contracts literature. Including U_t in the state preserves the history-dependence of the contract. However, as discussed in Williams (2011), additional information is needed in environments with persistent private information. This additional information is represented by the second co-state variable p_t^α. The pair (p_t^α, q_t^α) is a weak solution to (69) where:

$$p_t^\alpha = \mathbb{E}_t\left[\nu_t^{-1} \int_t^T e^{-r(s-t)} \beta_s \nu_s ds + \nu_t^{-1} \nu_T e^{-r(T-t)} C_\alpha(\tilde{\theta}_T)\right] \tag{71}$$

Let $p_t^\alpha \equiv \Gamma_t$ denote the CEO's *information rent*, which serves as an additional necessary state variable. Substituting into the IC constraint, and noting again that $\Lambda_t = 1$ under the recommended action path, the constraint becomes:

$$\beta_t \geq \Gamma_t \nu_t + \phi \tag{72}$$

(72) coincides exactly with Jovanovic and Prat (2014), who consider a similar environment with non-stationary learning, but use the alternative approach of Cvitanic et al. (2009) to derive necessary conditions. It is also identical to Demarzo and Sannikov (2017) with the exception of ν_t being constant in their model.

Using (68), we know that along the equilibrium path, the CEO's promised value evolves according to:

$$dU_{ijt} = \left(rU_t - w_t - \lambda C(\tilde{\theta}_t)\right)dt + \beta_t \sigma dZ_t \tag{73}$$

where $\frac{dU_t}{dY_t} = \beta_t$ is the sensitivity of U_t to profitability dY_t. The constraint (72) implies a lower bound on the sensitivity process β_t. Noting the presence of β_t in the expression for Γ_t, it also implies a lower bound on the CEO's information rents. Substituting the constraint into (69) yields:

$$d\Gamma_t = (r+\nu_t)\Gamma_t dt - \beta_t dt + q_t^\alpha dZ_t \tag{74}$$
$$\geq (r+\nu_t)\Gamma_t dt - (\Gamma_t \nu_t + \phi)dt + q_t^\alpha dZ_t \tag{75}$$
$$= (r\Gamma_t - \phi)dt + q_t^\alpha dZ_t \equiv d\Gamma_t^* \tag{76}$$

Solving (76) using the terminal value $\Gamma_t^* = C_\alpha(\tilde{\theta}_T) = \frac{\phi}{r}$ yields:

$$\Gamma_t^* = \mathbb{E}\left[\int_t^T e^{-r(s-t)}\phi ds + e^{-r(T-t)}\frac{\phi}{r}\right] \tag{77}$$
$$= \frac{\phi}{r} \tag{78}$$

Γ_t^* is the minimum value of Γ_t in equilibrium, which is attained when the IC binds. Under the assumed functional form of the CEO's outside option, Γ_t^* is a constant, so can be dispensed with as a state variable as long as the IC binds. □

Proposition 3. Given an arbitrary termination boundary $\theta_f(t)$, the CEO's equilibrium value of continuing employment satisfies the HJB equation:

$$rU(\tilde{\theta}_t, t, \theta_f(t)) = w_t + \frac{\partial U}{\partial t} + \theta_f'(t)\frac{\partial U}{\partial \theta_f} + \frac{\nu_t^2\sigma_W^2}{2}\frac{\partial^2 U}{\partial \theta^2} + \lambda C(\tilde{\theta}_t) \tag{79}$$

The cost-minimizing compensation process ensures that $U(\tilde{\theta}_t, t, \theta_f(t)) = U_t^*$ for all t. This process is obtained by substituting the expression for U_{ijt}^* into the HJB equation above. Doing so and collecting terms yields:

$$w_t = \rho\mu + \kappa_{1t}\tilde{\theta}_t + \kappa_{2t}\theta_f(t) \tag{80}$$
$$\kappa_{1t} = \phi\left(\frac{\rho}{r} + \frac{r}{\nu_t} + \frac{\nu_t}{r}\right) \tag{81}$$
$$\kappa_{2t} = \phi\left(1 - \frac{r}{\nu_t}\right) \tag{82}$$

□

Proposition 3. Applying similar arguments as in the first-best case, Ito's lemma implies that the firm's HJB equation is given by:

$$\begin{aligned} rV(\tilde{\theta}_{ijt}, t, U_{ijt}) = b\tilde{\theta}_{ijt} - w_{ijt} + V_t(\tilde{\theta}_{ijt}, t, U_{ijt}) + \Big(rU_{ijt} - w_{ijt} - \lambda C(\tilde{\theta}_{ijt})\Big) V_U(\tilde{\theta}_{ijt}, t, U_{ijt}) \\ + \frac{\nu_{ijt}^2 \sigma_W^2}{2} V_{\theta\theta}(\tilde{\theta}_{ijt}, t, U_{ijt}) + \frac{\beta_{ijt}^2 \sigma_W^2}{2} V_{UU}(\tilde{\theta}_{ijt}, t, U_{ijt}) + \lambda V_T \end{aligned} \tag{83}$$

To derive the optimal firing boundary $\theta_f(t)$, I apply the following conditions:

$$V(\theta_f(t), t, U_{ijt}) = V_T - b\pi \tag{84}$$

$$V_t(\theta_f(t), t, U_{ijt}) = \frac{\partial V_T}{\partial t} = 0 \tag{85}$$

$$V_U(\theta_f(t), t, U_{ijt}) = \frac{\partial V_T}{\partial U} = 0 \tag{86}$$

Condition (84) is the value matching condition, while conditions (85) and (86) are smooth pasting conditions. Substituting these conditions into the HJB equation evaluated along the firing boundary yields:

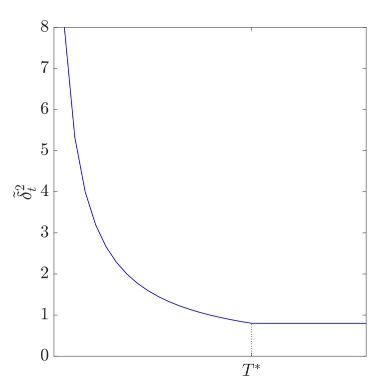
$$\theta_f(t) = -r\pi + b^{-1}\left(\rho V_T - \frac{\nu_{ijt}^2 \sigma_W^2}{2} V_{\theta\theta}(\theta_f(t), t, U_{ijt}) - \frac{\beta_{ijt}^2 \sigma_W^2}{2} V_{UU}(\theta_f(t), t, U_{ijt})\right) \tag{87}$$

□

8.2.3 Numerical Solution for First-Best Case

To obtain a numerical solution, I first assume that at some $T^* < \infty$, the posterior variance $\tilde{\delta}_t^2$ remains constant. Thus, $\tilde{\delta}_{ijt} = \tilde{\delta}_{ijs} \equiv \tilde{\delta}$ and $\nu_{ijt} = \nu_{ijs} \equiv \nu$ for all $T^* \leq t < s$. Note that this approximation can be arbitrarily accurate, as T^* can be arbitrarily large. At T^*, the firm's

Figure 11: Evolution of Posterior Variance



value can be derived analytically. Consider a firm who has retained their CEO until T^*, their value is given by:

$$V(\tilde{\theta}_{ijT^*}, T^*) \equiv V(\tilde{\theta}_{ijT^*}) = \mathbb{E}\left[\int_{T^*}^{T} e^{-r(s-T^*)}\tilde{\theta}_{ijT^*} ds + e^{-r(T-T^*)} V_T\right] \tag{88}$$

Importantly, time is no longer a relevant state variable for $t \geq T^*$. Applying Ito's lemma in this case yields the simplified HJB equation:

$$rV(\tilde{\theta}_{ijT^*}) = \tilde{\theta}_{ijT^*} + \frac{\nu^2\sigma^2}{2}\frac{d^2V}{d\theta^2} \tag{89}$$

with boundary condition $V(\theta_f) = V_T$. This is a straightforward second-order ODE with solution:

$$V(\tilde{\theta}_{ijT^*}) = \frac{\tilde{\theta}_{ijT^*}}{r} + exp\left(-\sqrt{2r}\left(\frac{\tilde{\theta}_{ijT^*} - \theta_f}{\nu\sigma}\right)\right)\left(V_T - \pi - \frac{\theta_f}{r}\right) \tag{90}$$

Hence, conditional on reaching tenure T^*, the firm's value is given exactly by Equation (90). I solve the model backwards from this point. I'll first introduce some notation largely following Brandimarte (2006). Define the discrete grids of state variables:

$$\Theta = \{\mu + \Delta\theta, \mu + 2\Delta\theta, \ldots, \mu + (M-1)\Delta\theta, \mu + M\Delta\theta\} \tag{91}$$

$$\mathbb{T} = \{\Delta t, 2\Delta t, \ldots, (N-1)\Delta t, N\Delta t\} \tag{92}$$

where $\mu + \Delta\theta$ is the smallest value of θ_{ijt} contained in its discrete grid. Define $V_{i,j} \equiv V(\mu + i\Delta\theta, j\Delta t)$ as the discretized counterpart of the firm's value function evaluated at grid points $\mu + i\Delta\theta$ and $j\Delta t$. I use a finite difference approach to approximate the derivatives in the HJB equation (60). I use the backward and standard approximations of V_t and $V_{\theta\theta}$ respectively:

$$\frac{\partial V(\tilde{\theta}_t, t)}{\partial t} \approx \frac{V_{i,j} - V_{i,j-1}}{\Delta t} \tag{93}$$

$$\frac{\partial^2 V(\tilde{\theta}_t, t)}{\partial \theta^2} \approx \frac{V_{i+1,j-1} - 2V_{i,j-1} + V_{i-1,j-1}}{(\Delta\theta)^2} \tag{94}$$

Plugging these approximations into (60) and doing some algebra yields the discretized HJB equation:

$$V_{i,j} = A_{i,j}V_{i-1,j-1} + B_{i,j}V_{i,j-1} + C_{i,j}V_{i+1,j-1} + D_i \tag{95}$$
$$A_{i,j} = \frac{\nu_j^2\sigma^2}{2}\rho \tag{96}$$
$$B_{i,j} = \nu_j^2\sigma^2\rho + r\Delta t + 1 \tag{97}$$
$$C_{i,j} = -\frac{\nu_j^2\sigma^2}{2}\rho \tag{98}$$
$$D_i = -(\mu + i\Delta\theta)\Delta t \tag{99}$$

where $\rho = \frac{\Delta t}{(\Delta\theta)^2}$. This can be represented as an $M-1 \times M-1$ system of linear equations:

$$\begin{pmatrix} V_{1,j} \\ V_{2,j} \\ V_{3,j} \\ \vdots \\ V_{M-1,j} \\ V_{M,j} \end{pmatrix} = \begin{pmatrix} B_{1,j} & C_{1,j} & 0 & 0 & 0 & 0 & 0 \\ A_{2,j} & B_{2,j} & C_{2,j} & 0 & 0 & 0 & 0 \\ 0 & A_{3,j} & B_{3,j} & C_{3,j} & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & A_{M-1,j} & B_{M-1,j} & C_{M-1,j} \\ 0 & 0 & 0 & 0 & 0 & A_{M,j} & B_{M,j} \end{pmatrix} \begin{pmatrix} V_{1,j-1} \\ V_{2,j-1} \\ V_{3,j-1} \\ \vdots \\ V_{M-1,j-1} \\ V_{M,j-1} \end{pmatrix} + \begin{pmatrix} D_1 + A_{1,j}V_{0,j-1} \\ D_2 \\ D_3 \\ \vdots \\ D_{M-1} \\ D_M + C_{M,j}V_{M+1,j-1} \end{pmatrix}$$

Note that the values $V_{0,j-1}$ and $V_{M+1,j-1}$ are not defined. These are instead given by the boundary conditions $V_{0,j-1} = R$ and $V_{M+1,j-1} = V(M\Delta\theta,(j-1)\Delta t)$. Rewriting the system slightly:

$$\underbrace{\begin{pmatrix} V_{1,j} - (D_1 + A_{1,j}V_{0,j-1}) \\ V_{2,j} - D_2 \\ V_{3,j} - D_3 \\ \vdots \\ V_{M-1,j} - D_{M-1} \\ V_{M,j} - (D_M + C_{M,j}V_{M+1,j-1}) \end{pmatrix}}_{\mathbf{D}} = \underbrace{\begin{pmatrix} B_{1,j} & C_{1,j} & 0 & 0 & 0 & 0 & 0 \\ A_{2,j} & B_{2,j} & C_{2,j} & 0 & 0 & 0 & 0 \\ 0 & A_{3,j} & B_{3,j} & C_{3,j} & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & A_{M-1,j} & B_{M-1,j} & C_{M-1,j} \\ 0 & 0 & 0 & 0 & 0 & A_{M,j} & B_{M,j} \end{pmatrix}}_{\mathbf{Q}} \underbrace{\begin{pmatrix} V_{1,j-1} \\ V_{2,j-1} \\ V_{3,j-1} \\ \vdots \\ V_{M-1,j-1} \\ V_{M,j-1} \end{pmatrix}}_{\mathbf{V}}$$

This can be expressed more succinctly in matrix form: $\mathbf{D} = \mathbf{QV}$. Note that the matrix $\mathbf{Q}$ is tridiagonal, and thus can be decomposed into an upper triangular matrix $\mathbf{U}$, a lower triangular matrix $\mathbf{L}$, and a diagonal matrix with positive elements $\mathbf{P}$:

$$\mathbf{Q} = \mathbf{L} + \mathbf{U} + \mathbf{P} \tag{100}$$

With this insight, the successive over relaxation (SOR) iterative method can be applied to solve for firm values for $t < T^*$. To apply this method, I rewrite the linear system as:

$$(\mathbf{P} + \omega\mathbf{L})\mathbf{V}^n = [(1-\omega)\mathbf{P} - \omega\mathbf{U}]\,\mathbf{V}^{n-1} + \omega\mathbf{D} \tag{101}$$

n is the iteration counter and the parameter $\omega \in (0,2)$ is the relaxation parameter whose optimal value is given by:

$$\omega = \frac{2}{1 + \sqrt{1 - |\rho(\mathbb{G})|^2}} \tag{102}$$

where $\rho(\mathbb{G})$ is the spectral radius of the matrix $\mathbb{G} = \mathbf{P}^{-1}(\mathbf{L}+\mathbf{U})$. When $(\mathbf{P}+\omega\mathbf{L})$ is invertible, (101) can be written as:

$$\mathbf{V}^n = (\mathbf{P} + \omega\mathbf{L})^{-1}\Big[[(1-\omega)\mathbf{P} - \omega\mathbf{U}]\,\mathbf{V}^{n-1} + \omega\mathbf{D}\Big] \tag{103}$$

whose elements can be computed via forward substitution:

$$V_i^n = (1-\omega)V_i^{n-1} + \frac{\omega}{Q_{i,i}}\left[D_i - \sum_{j=1}^{i-1} Q_{i,j}V_j^n - \sum_{j=i+1}^{n} Q_{i,j}V_j^{n-1}\right] \tag{104}$$

8.2.4 Numerical Solution for Second-Best Case

The solution method for the second-best case works largely the same as in the first-best, with the addition of U_t as a state variable. As before, I assume some arbitrarily large T^* such that $\tilde{\delta}_t = \tilde{\delta}_s \equiv \tilde{\delta}$ for all $t \geq T^*$.

For $t < T^*$, the firm's HJB equation is given by:

$$rV(\tilde{\theta}_t, t, U_t) = \tilde{\theta}_{ijt} - w_{ijt} + V_t + \Big(rU_t - w_t - \lambda C(\tilde{\theta}_t)\Big)V_U \tag{105}$$

$$+ \frac{\nu_t^2\sigma^2}{2}V_{\theta\theta} + \frac{\beta_t^2\sigma^2}{2}V_{UU} + \lambda V_T \tag{106}$$

I approximate this with a linear system of equations using an upwind finite-difference scheme. The approximations of the relevant derivatives are given by:

$$
\begin{aligned}
\frac{\partial V(\tilde{\theta}_t, t, U_t)}{\partial t} &\approx \frac{V_{i,j}^{t} - V_{i,j}^{t-1}}{\Delta t} && (107)\\
\frac{\partial^2 V(\tilde{\theta}_t, t, U_t)}{\partial \theta^2} &\approx \frac{V_{i+1,j}^{t-1} - 2V_{i,j}^{t-1} + V_{i-1,j}^{t-1}}{(\Delta\theta)^2} && (108)\\
\frac{\partial^2 V(\tilde{\theta}_t, t, U_t)}{\partial U^2} &\approx \frac{V_{i,j+1}^{t-1} - 2V_{i,j}^{t-1} + V_{i,j-1}^{t-1}}{(\Delta U)^2} && (109)
\end{aligned}
$$

Consistent with the upwind scheme, I approximate the partial derivative V_U by:

$$
\frac{\partial V(\tilde{\theta}_t, t, U_t)}{\partial U} \approx \begin{cases} \frac{V_{i,j+1}^{t-1} - V_{i,j}^{t-1}}{\Delta U} & \text{if} \quad rU_{ijt} \geq w_{ijt} + \lambda C(\tilde{\theta}_{ijt}) \\ \frac{V_{i,j}^{t-1} - V_{i,j-1}^{t-1}}{\Delta U} & \text{if} \quad rU_{ijt} < w_{ijt} + \lambda C(\tilde{\theta}_{ijt}) \end{cases} \tag{110}
$$

Substituting these approximations into the firm's HJB yields the implicit scheme:

$$
\begin{aligned}
V_{i,j}^{t} &= A^t V_{i-1,j}^{t-1} + B_{i,j}^{t} V_{i,j}^{t-1} + C^t V_{i+1,j}^{t-1} + D_{i,j}^{t} V_{i,j-1}^{t-1} + E_{i,j}^{t} V_{i,j+1}^{t-1} + F_{i,j}^{t} && (111)\\
A^t &= -\frac{1}{2}\nu_t^2 \sigma^2 \rho_{\theta\theta} && (112)\\
B_{i,j}^{t} &= 1 + r\Delta t + \Big(rU_{i,j}^{t-1} - w_{i,j}^{t-1} - \lambda C(i\Delta\theta)\Big)\rho_u \Big(\mathbf{1}[drift \geq 0] - \mathbf{1}[drift < 0]\Big) \\
&\quad + \rho_{\theta\theta}\nu_t^2\sigma^2 + \rho_{uu}\beta_t^2\sigma_t^2 && (113)\\
C^t &= -\frac{1}{2}\nu_t^2 \sigma^2 \rho_{\theta\theta} && (114)\\
D_{i,j}^{t} &= \Big(rU_{i,j}^{t-1} - w_{i,j}^{t-1} - \lambda C(i\Delta\theta)\Big)\rho_u \mathbf{1}[drift < 0] - \frac{1}{2}\beta_t^2\sigma^2\rho_{uu} && (115)\\
E_{i,j}^{t} &= -\Big(rU_{i,j}^{t-1} - w_{i,j}^{t-1} - \lambda C(i\Delta\theta)\Big)\rho_u \mathbf{1}[drift \geq 0] - \frac{1}{2}\beta_t^2\sigma^2\rho_{uu} && (116)\\
F_{i,j}^{t} &= -\Big(i\Delta\theta - w_{i,j}^{t-1} + \lambda V_T\Big) && (117)
\end{aligned}
$$

where $\rho_u = \frac{\Delta t}{\Delta U}$, $\rho_{uu} = \frac{(\Delta t)^2}{(\Delta U)^2}$, and $\rho_{\theta\theta} = \frac{(\Delta t)^2}{(\Delta\theta)^2}$. This is represented as an $(M-1)(N-1) \times (M-1)(N-1)$ linear system:

$$
\begin{pmatrix} V_{1,1}^t \\ V_{2,1}^t \\ \vdots \\ V_{M,1}^t \\ V_{1,2}^t \\ \vdots \\ V_{M-1,N}^t \\ V_{M,N}^t \end{pmatrix} = \begin{pmatrix} B_{1,j} & C_{1,j} & 0 & 0 & \dots & E_{1,1}^t & 0 & \dots & \dots \\ A_{2,j} & B_{2,j} & C_{2,j} & 0 & \dots & 0 & E_{2,2}^t & \dots & \dots \\ \vdots & & & & & & & & \\ \dots & A_{3,j} & B_{3,j} & C_{3,j} & \dots & E_{M,2}^t & \dots & 0 & \dots \\ D_{1,1}^t & \dots & A_{3,j} & B_{3,j} & C_{3,j} & \dots & \dots & E_{1,3}^t & \dots \\ \vdots & & & & & & & & \\ 0 & D_{M-1,N-1}^t & \dots & \dots & 0 & 0 & A_{M-1,j} & B_{M-1,j} & C_{M-1,j} \\ 0 & 0 & D_{M,N-1}^t & \dots & 0 & 0 & 0 & A_{M,j} & B_{M,j} \end{pmatrix} \begin{pmatrix} V_{1,1}^{t-1} \\ V_{2,1}^{t-1} \\ \vdots \\ V_{M,1}^{t-1} \\ V_{1,2}^{t-1} \\ \vdots \\ V_{M-1,N}^{t-1} \\ V_{M,N}^{t-1} \end{pmatrix} + \begin{pmatrix} F_{1,1}^t + A_{1,j}V_{0,j-1} \\ F_{2,1}^t \\ \vdots \\ F_{M,1}^t \\ F_{1,2}^t \\ \vdots \\ F_{M-1,N}^t \\ F_{M,N}^t + C_{M,j}V_{M+1,j-1} \end{pmatrix}
$$

With this representation in hand, I proceed by using the same method as in the first-best case.

8.3 Estimation Appendix

8.3.1 Weighting Matrix

From the empirical sample I obtain a $K \times 1$ vector of moments $\hat{M}$. Let Ψ denote the corresponding $N \times K$ matrix of influence functions, N being the number of observations in the sample. Each element Ψ_{nk} is the influence function describing observation n's contribution to moment k. The covariance matrix of the vector of moments can then be estimated as:

$$a\hat{v}ar(\hat{M}) = \Psi'\Psi \tag{118}$$

The weighting matrix $\hat{W}$ is then obtained as the inverse of matrix 118. Let $\Theta \in \mathbb{R}^P$ denote an arbitrary vector of structural parameters. Define the moment residual $g : \mathbb{R}^P \to \mathbb{R}^M$ as:

$$g(\Theta) = \hat{M} - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta) \tag{119}$$

Where $\hat{M}$ is the vector of empirical moments, $\hat{m}^s(\Theta)$ is the vector of simulated moments given parameter values Θ in simulation s, and S is the total number of simulations. The vector of estimates $\hat{\Theta}$ minimizes the SMM objective function:

$$\hat{\Theta} = \arg\min_{\Theta} \; g(\Theta)\hat{W}g(\Theta)' \tag{120}$$

8.3.2 Model Estimation Algorithm

I use the particle swarm algorithm to minimize the SMM objective function (120). The model is estimated as follows:

1. *Set initial guesses for model parameters:* I set initial values for the structural parameters Θ. The initial guess is chosen manually, while subsequent guesses are selected by the particle swarm algorithm.

2. *Compute the firm's value function:* Given a vector of parameters Θ, I compute the value function V using the procedure outlined in Appendix 8.2.4, from which the optimal firing boundary and compensation process can be computed.

3. *Simulate model:* Given the optimal firing and compensation policies, I simulate 5000 firms 20 times each. Firms draw an initial CEO from distribution $N(\theta_0, \delta_0^2)$. Perfor-

mance evolves according to Equation (5) and beliefs evolve according to Equation (20). Firms make compensation and firing decisions based upon the optimal policies outlined in the second-best case of the model.

4. *Construct simulated panel and compute moments:* Using the simulated data, I construct a panel resembling the empirical sample and compute the same moments as described in Section 4.

5. *Evaluate objective function:* Given the set of simulated moments, I evaluate the SMM objective function (120). If the objective function value satisfies the particle swarm stopping criterion, the algorithm halts. Otherwise, a new candidate parameter vector Θ' is selected and steps 2-5 repeat. This continues until the algorithm halts.

8.3.3 Standard Errors for Parameter Estimates

For true parameter vector Θ and consistent estimate $\hat{\Theta}$, we have the following asymptotic distribution (Duffie and Singleton, 1993):

$$\sqrt{n}(\hat{\Theta} - \Theta) \rightarrow^{d} N(0, avar(\hat{\Theta})) \tag{121}$$

$avar(\hat{\Theta})$ can be expressed as:

$$avar(\hat{\Theta}) = \left(1 + \frac{1}{S}\right)\left(\frac{\partial g(\Theta)}{\partial \Theta} W \frac{\partial g(\Theta)}{\partial \Theta'}\right)^{-1} \tag{122}$$

where $\frac{\partial g(\Theta)}{\partial \Theta}$ is the Jacobian of the moment residual (119) with respect to the structural parameters, W is the optimal weighting matrix, and S is the number of simulations. I approximate the Jacobian using:

$$\frac{\partial \hat{g}_m(\Theta)}{\partial \Theta_p} = \frac{g_p(\hat{\Theta} + h_p) - g_p(\hat{\Theta})}{h_p} \tag{123}$$

for each moment m and parameter p. h_p is the perturbation size for parameter which I set to 1% of the absolute value of the parameter estimate. The standard errors are then obtained as the square root of the diagonal elements of the matrix:

$$\left(1 + \frac{1}{S}\right)\left(\frac{\partial \hat{g}(\Theta)}{\partial \Theta} \hat{W} \frac{\partial \hat{g}(\Theta)}{\partial \Theta'}\right)^{-1} \tag{124}$$

where $\hat{W}$ is the sample counterpart of the optimal weighting matrix.

8.4 Further Details on Results

8.4.1 Model Fit

Figure 12: Fit of Estimation Moments

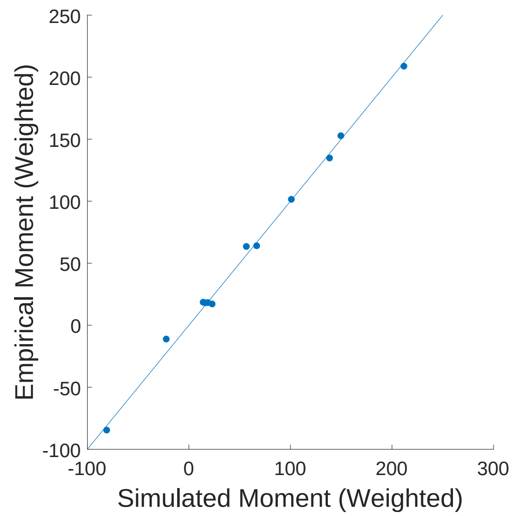
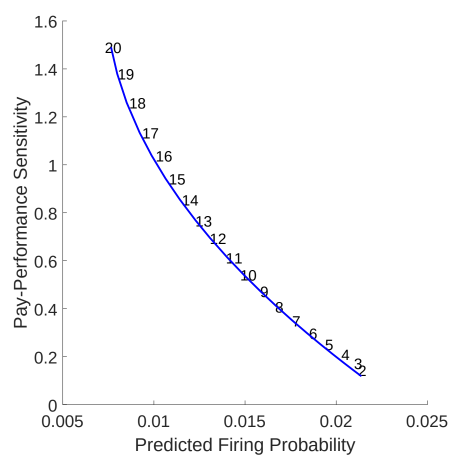


Figure 12 scatters the empirical moments over their simulated counterparts using the 45-degree line as a reference point. All moments are scaled by the corresponding empirical standard error. Additionally, the model replicates well the substitution of financial and termination incentives as illustrated in Figure 3. Figure 13 plot the simulated counterpart.

Figure 13: Simulated Incentive Pay and Termination Risk



Notes: For the first 20 years of CEO tenure, Figure 13 plots the average pay sensitivity (γ_{1t}) over the average rate of forced termination. The rate of forced termination decreases with tenure, while the pay-performance sensitivity increases.

Pay sensitivity here is measured by the parameter γ_{1t}, the marginal effect of CEOs' perceived contribution to ROA ($\tilde{\theta}_{ijt}$) on total compensation. As tenure increases, firms become increasingly certain about CEO quality, and low-quality CEOs are forced out. The high-quality CEOs which survive into later years of tenure face relatively low risk of job loss, so are incentivized alternatively via financial instruments. Such a pattern is consistent with Gibbons and Murphy (1992) for example, who also find that CEO pay contracts become increasingly sensitive to performance as tenure increases.