Why Are CEOs Rarely Fired?
Evidence from Structural Estimation

August 5, 2008

Abstract

Two percent of CEOs are fired per year on average. To evaluate this magnitude, I solve and estimate a dynamic model of forced CEO turnover. The model features costly turnover and learning about CEO ability. To rationalize the two percent firing rate, boards must behave as if replacing the CEO costs shareholders 5.9% of the firm's assets. This cost mainly reflects CEO entrenchment and poor governance rather than a real cost for shareholders. In terms of both direction and magnitude, the model helps explain the relation between CEO firings and tenure, profitability, and stock returns.

Previous empirical work has established that CEOs are rarely fired, and that firm performance is a poor predictor of forced CEO turnover. On average, 2% of CEOs at large U.S. corporations are fired each year (e.g., Kaplan and Minton (2006), Huson, Parrino, and Starks (2001)). While many authors find a statistically significant relation between turnover and measures of firm performance, a recent survey concludes that "performance continues to explain very little of the variation in CEO turnover" (Brickley (2003)). For instance, Kaplan and Minton (2006) forecast CEO turnover using lagged profitability and stock returns, and they obtain R-squared values of only 2 to 11%.

It is tempting to conclude from these patterns that boards do not act in shareholders' interests. However, the literature provides little guidance for making such a judgment. For example, it is not clear what rate of forced CEO turnover we should expect from boards which do act in shareholders' interests. Therefore, it is difficult to judge whether the observed 2% rate is low or high.

This paper's goal is to provide a benchmark for evaluating the frequency of forced CEO turnover and the relation between turnover and firm performance. The benchmark is a dynamic model featuring a rational board that maximizes shareholder value. In the model, the board decides at each point in time whether to fire or keep its current CEO. Some CEOs are more skilled than others, meaning they can produce higher average firm-specific profits. Firing an unskilled CEO is not always in shareholders' interests, because CEO turnover entails a real cost to shareholders. For instance, the firm may have to pay a severance package and executive search fees. Complicating matters, the board cannot directly observe CEO skill, but instead learns about it over time. The board learns in part by observing the CEO's industry-adjusted profits. However, profits are not the only information boards have about CEO skill. For example, stock returns, market share, the CEO's strategic plan, and the CEO's specific actions may also be informative. The model aggregates all this other information into a single additional signal. At each point in time, the board observes profits and the additional signal, assesses the CEO's skill, and then decides optimally whether to replace him with a new CEO of uncertain skill.

The model offers an explanation for why profitability poorly predicts forced CEO turnover: Boards rely heavily on information besides profits when evaluating CEO ability. The model predicts that boards rationally rely less on signals of CEO skill which are less precise. Profitability is an imprecise signal, because it is noisy (it fluctuates for reasons unrelated to CEO skill) and persistent (it responds slowly to CEOs' actions). If the profitability signal is much less precise than the board's additional signal, then the board pays little attention to profitability when evaluating the CEO's ability. As a result, profitability poorly predicts forced CEO turnover. The weak predictive relation is in shareholders' best interests, since the board uses the most reliable information to make firing decisions.

The model suggests three potential reasons why boards rarely fire CEOs. First, firing a CEO may entail large, real costs to shareholders. Second, CEO skill may not matter much. The model predicts that fewer CEOs are fired if there is less variation in skill across CEOs. Intuitively, if CEOs all have roughly the same ability, then there is little incentive to replace one CEO with another, especially if doing so is costly. Third, boards may learn slowly about CEO skill, so that unskilled CEOs survive longer in office and possibly retire before the board decides to fire them. The model predicts that boards learn more slowly when profitability is more volatile or persistent, when the board's additional signal is less precise, or when prior beliefs about CEO skill are stronger.

Outside the good-governance benchmark, there is a fourth potential reason why CEOs are rarely fired: Boards dislike firing CEOs for reasons unrelated to shareholder value. For instance, directors may have personal or professional ties to the CEO. Also, firing the CEO may put the directors' own jobs at risk, may require uncompensated effort to find a new CEO, or may hurt directors' chances of being nominated to other boards (Hermalin and Weisbach (1998)). To capture these effects, the model assumes the board incurs a personal utility cost each time CEO turnover occurs. Unlike the real turnover costs discussed in the previous paragraph, the personal costs do not directly affect profits. To the extent that boards honor these personal costs, they deviate from maximizing shareholder value, at least in an ex post sense.

It is a challenge to measure the importance of these four potential reasons why CEOs are rarely fired. The board's firing choices are endogenous, which generates endogenous patterns in firm performance. There are no obvious instruments. Several elements of the model are unobservable, including a CEO's actual and perceived skill, the CEO talent pool, the board's additional signals of CEO skill, and the board's personal costs of CEO turnover. Also, some of the reasons interact. For example, variation in CEO skill (reason two) affects the speed of learning (reason three) via the uncertainty about CEOs' ability. Finally, measuring the reasons' magnitudes is difficult. While we can measure directional effects using reduced-form empirical techniques, evaluating magnitudes requires estimating or calibrating an economic model.

These challenges lend themselves to a structural estimation approach. The structural approach infers unobservable quantities from endogenous patterns in firing decisions and firm performance. It also takes into account interactions between the four reasons above. Finally, with a structural approach we can assess not only the reasons' directional effects, but also their magnitudes.

I estimate the model's parameters by applying the simulated method of moments to data on firm profitability and both forced and voluntary CEO turnover in large U.S. firms from 1971 to 2006. The estimated parameters include the real cost of CEO turnover to shareholders, the variation in skill across potential CEOs, the volatility and persistence of profitability, the precision of boards' additional information about CEO skill, and the personal cost of turnover to the board.

Estimates imply that extremely high CEO turnover costs are needed to rationalize the observed rate of forced CEO turnover. Boards behave as if the total cost of firing a CEO is an estimated 5.9% of the firm's assets, or \$236 million for the median sample firm. This estimate provides a metric for evaluating the empirical rate of forced CEO turnover: the 2% rate is indeed low, in the sense that an extremely high CEO turnover costs is needed to explain it.

The total CEO turnover cost is the sum of real costs to shareholders and personal utility costs to the board. This study disentangles these costs using two assumptions: the personal costs do not affect profits, and the real costs show up in profits in the two years around the succession. I find that the personal cost makes up 4.6% of the total 5.9% turnover cost. In dollar terms for the median firm, the total \$236 million turnover cost breaks down into \$183 million of personal costs and \$53 million of real costs. The board behaves as if firing the CEO costs shareholders \$236 million, whereas it really only costs them \$53 million. The result is that the board keeps some CEOs whom shareholders would rather see fired. The large personal cost indicates either that CEO turnover is extremely costly to directors (in a utility sense), or that directors do not care much about shareholder value. The model cannot distinguish between these two possibilities. Either way, the results imply CEOs are highly entrenched.

One interpretation of this entrenchment is that it reflects poor governance by boards. Consistent with this interpretation, personal turnover costs are significantly smaller in situations with better governance, as proxied by a higher fraction of directors who are not also officers, CEOs who left office in 1990-2006 instead of 1971-1989, or in larger firms where shareholders have greater incentives to monitor the board. A second interpretation is that the level of entrenchment I measure is optimal for shareholders ex ante, for instance, because it allows shareholders to pay the CEO less or hire a better CEO. Consistent with this interpretation, I find no significant relation between personal costs and the fraction of shares owned by the board, another measure of governance quality. On balance, results indicate that the personal turnover costs reflect bad governance. However, not all the evidence above supports this view.

As mentioned earlier, profitability poorly predicts forced CEO turnover in the data. To rationalize this pattern, I find that boards' additional signal of CEO skill (the signal unrelated to profits) must have a 5.3 times larger influence on the board's beliefs than profitability has. Essentially, the board's non-earnings signal must be extremely precise, so the board does not put much weight on earnings when evaluating a CEO's ability. As a result, the board's firing decisions are only weakly related to profitability. This result provides a metric for evaluating the profitability-turnover relationship in the data: the relationship is indeed weak, in the sense that to rationalize it, boards must have extremely precise additional information about CEO ability.

The estimated model fits the data well. The model predicts that 2.2% of CEOs are fired per year on average, which is close to the 2.3% rate in the data. Because of its dynamic setup, the model can also explain the timing of forced CEO turnover, which depends on how fast the boards learn about CEO skill. The model predicts that the median fired CEO spends a total of four years in office, which exactly matches the empirical median. Also, the model generates a hump-shaped relation between forced turnover rates and tenure, which is similar to the pattern in the data. The model can fit several aspects of the turnover-profitability relationship. In predictive probit models of forced CEO turnover on lagged profitability, the model generates a pseudo R-squared of only 2%, which is close to the 3% value from the empirical sample. The model can also closely match the empirical probit slopes. The model produces a V-shaped pattern in average profitability around CEO dismissals, which closely matches the empirical pattern. The model also makes predictions about stock returns. The model predicts a –18% abnormal stock return over the five years before forced CEO turnover, on average. On this dimension, the model misses in terms of magnitude: the corresponding empirical return is –35%. However, the model gets the shape right. In both the data and the model, stock prices drop gradually leading up to CEO dismissal and are essentially flat after.

While there is a large empirical[1] and theoretical[2] literature on CEO turnover, this is the first study that estimates a structural model of CEO turnover. Therefore, this is one of the first attempts at using an economic model to evaluate the magnitudes in the CEO turnover data. As I discuss later, my model is consistent with several existing empirical findings, and the model also produces new, untested predictions. My goal is not to make a theory contribution, however, but to adapt and estimate existing models, and use the parameter estimates to evaluate empirical magnitudes.

Hermalin and Weisbach (1998) is the most closely related model in the CEO turnover literature. Unlike their model, mine has a fully dynamic setup and a persistent performance signal. My model is simpler in that it does not feature endogenous monitoring, board composition, or CEO pay.

Eisfeldt and Rampini (2008) also use a model to evaluate magnitudes in the CEO turnover data. Their model focuses on CEOs' incentives for revealing private information, whereas mine focuses on learning about manager ability with symmetric information. They calibrate their model to match business cycle variation in CEO turnover and compensation, which I do not attempt to explain.

Like my model, the model of Dangl, Wu, and Zechner (2007) features learning about a manager's skill from his output, optimal firing decisions, costly turnover, and a dynamic setup. Unlike my contribution, theirs is purely theoretical, and their focus is mutual fund managers.

Miller (1984) estimates a model similar to mine using labor market data from multiple occupations, such as farm workers. Our models both feature optimal worker separation, a dynamic setup, and learning about skill from the worker's output. However, there are several differences between our models, and we use different data, identification strategies, and estimation methods.

[1]E.g., Coughland and Schmidt (1985); Warner, Watts, and Wruck (1988); Weisbach (1988); Murphy and Zimmerman (1993); Weisbach (1995); Kim (1996); Parrino (1997); Huson, Parrino, and Starks (2001); Parrino, Sias, and Starks (2003); Engel, Hayes, and Wang (2003); Huson, Malatesta, and Parrino (2004); Kaplan and Minton (2006); Jenter and Kanaan (2006); Lehn and Zhao (2006).

[2]E.g., Herschleifer and Thakor (1994, 1998); Hermalin and Weisbach (1998, 2007); Fisman, Khurana, and Rhodes-Kropf (2007); Eisfeldt and Rampini (2007).

I present the model in Section 1. Section 2 describes the data and estimation method. Section 3 presents estimation results, and Section 4 discusses robustness. Section 5 concludes.

I. Model

In this section I develop and solve a model of forced CEO turnover. In the model, a board decides at the beginning of each period whether to fire or keep the current CEO. Some CEOs are more skilled than others, meaning they can generate higher average firm-specific profits. The board faces a tradeoff when deciding whether to fire an unskilled CEO. On one hand, the board cares about future profits and therefore prefers to have a skillful CEO in office. On the other hand, firing the CEO entails a real cost to shareholders as well as a personal utility cost to the board. The board cannot directly observe CEO ability, but instead learns gradually by observing two signals, firm-specific profitability and an additional, unrelated signal. At each point in time, the board makes its best possible assessment of the CEO's ability and then makes an optimal firing decision.

A. Assumptions

The model features a firm which lives for an infinite number of periods, a large pool of potential CEOs, and a board which makes CEO firing decisions. I set one period equal to a year in the empirical implementation. The board of directors can fire the CEO and hire a new one at the beginning of each period. In addition, a CEO who has already served τ periods voluntarily leaves the firm (he either quits or retires) with exogenous probability $f(\tau)$.[3] The firm's book value of assets equals B_t at the beginning of period t.[4] The firm generates profits equal to $Y_t B_t$ at the end of period t, so Y_t is the firm's profitability.[5] Profitability has three components:

$$Y_t = v_t + y_t - c_t^{(firm)}. \tag{1}$$

Component v_t is the industry average profitability at time t. Component $c_t^{(firm)}$ is the real cost of CEO turnover, which I define later. Firm-specific profitability y_t mean-reverts around α, the current CEO's skill level:

$$y_t = y_{t-1} + \phi\left(\alpha - y_{t-1}\right) + \epsilon_t. \tag{2}$$

The shock ϵ_t is independently and normally distributed with mean zero and variance σ_ϵ^2. To be precise, equation (2) should include a CEO-specific subscript on ability α, because different CEOs can have different ability levels. A given CEO's ability is constant over his tenure in office. Equation (2) defines the model's

[3] Making voluntary turnover exogenous and only dependent on tenure improves tractability, is consistent with the lack of an empirical correlation between firm performance and voluntary turnover (Huson, Malatesta, and Parrino (2004)), and still allows the timing of voluntary turnover to affect the firing decision.

[4] For tractability, and since investment and dividend policy are not a focus of this paper, I assume all profits (including potential negative profits) are immediately paid out as dividends.

[5] Profitability is net of CEO pay, which is outside the model. Therefore, the model takes no stand on how the surplus from CEO ability is shared between the CEO and shareholders.

notion of CEO skill: A CEO is considered highly skilled (i.e. high α) if he can achieve profitability higher than the industry, on average and in the long run.

Parameter ϕ determines the persistence in firm-specific profitability, y_t. y_t is a random walk when $\phi = 0$, is iid when $\phi = 1$, and is mean reverting for $0 < \phi < 1$. I allow persistence in firm-specific profitability for two reasons. First, there is empirical evidence of persistence (e.g. Fama and French (2000)). More importantly, persistence allows a CEO to have long-lasting effects on profitability, which is plausible and affects the firing decision. For instance, after a CEO is fired for poor earnings performance, earnings may continue to be low for a few years even if the new CEO is highly skilled, because it takes time to undo the old CEO's mistakes. During those years, the board would not want to penalize the new CEO for the old CEO's mistakes. In other words, persistence in profitability affects the way the board evaluates CEO ability.

I define component $c_t^{(firm)}$ next. There are two types of CEO turnover costs in the model: real costs to shareholders ("firm costs"), denoted $c_t^{(firm)}$, and costs to the board ("personal costs"), denoted $c_t^{(pers)}$. Firm costs include severance or retirement packages, fees to executive search firms, disruption costs, and any other CEO turnover costs which directly affect profits. Personal costs do not affect the firm's profits, but do affect the board members. Examples of personal costs include the loss of the CEO as an ally both within the firm and in the directors' careers outside the firm, any uncompensated effort and stress from the succession process, and in the case of forced turnover, reputation costs from "rocking the boat," which may damage directors' chances of being nominated to other boards (Hermalin and Weisbach (1998)). I assume forced and voluntary CEO turnover are equally costly. For robustness, in Section 4 I solve and estimate the model assuming voluntary turnover is costless. I assume all turnover costs are incurred in the turnover period, so the firm cost $c_t^{(firm)} = c^{(firm)}$ if CEO turnover occurs in period t, and otherwise $c_t^{(firm)}$ equals zero. I define the personal cost $c_t^{(pers)}$ similarly. Both turnover costs are a constant fraction of the firm's book assets. This assumption is motivated from empirical evidence that costs such as separation pay and executive search fees increase with firm size[6]. For robustness, later I allow the costs to depend on firm size.

The board makes firing choices $d_t \in$\{fire CEO, keep CEO\} that maximize lifetime utility U_t :

$$\max_{\{d_{t+s}\}_{s=0}^{\infty}} U_t \equiv E_t\left[\sum_{s=0}^{\infty} \beta^s u_{t+s}\right]. \tag{3}$$

Parameter β is the board's discount factor, with $0 < \beta < 1$. Expectation E_t is with respect to the board's information at time t. The board's time-t utility is

$$u_t \equiv \kappa B_t Y_t - B_t c_t^{(pers)}. \tag{4}$$

The board prefers higher profits (the $\kappa B_t Y_t$ term) and experiences a personal cost from CEO succession (the $B_t c_t^{(pers)}$ term). The constant $\kappa > 0$ controls the degree to which the board internalizes shareholder value.

[6]Yermack (2006) finds that separation pay to CEOs is increasing in firm size. Executive search fees are proportional to CEO compensation (e.g., the Association of Executive Search Consultants, http://www.aesc.org/), which increases in firm size (Gabaix and Landier (2006)).

For instance, we might believe κ is higher when directors have a greater sense of fiduciary responsibility, own more shares or options, or receive greater reputation benefits from their firm's success. CEO firing choices affect the board's utility in two ways. First, they affect profitability Y_t, because profitability depends on the acting CEO's ability as well as the firm turnover costs $c^{(firm)}$. Second, the board incurs an additional personal cost $c^{(pers)}$ each time it fires the CEO. There is an indeterminacy between κ and $c^{(pers)}$, since the utility function is defined up to an affine transformation only. I discuss this indeterminacy more later.

Substituting (4) into (3), we can decompose the board's objective function into two terms:

$$U_t = \kappa E_t \left[\sum_{s=0}^{\infty} \beta^s B_{t+s} Y_{t+s} \right] - E_t \left[\sum_{s=0}^{\infty} \beta^s B_{t+s} c_{t+s}^{(pers)} \right] \tag{5}$$

If we interpret discount factor β as the firm's cost of capital, then the first term in equation (5) is κ times the net present value of future dividends, or the board's assessment of shareholder value. The second term is the net present value of personal turnover costs. Therefore, the board maximizes shareholder value when and only when there are no personal costs ($c^{(pers)} = 0$), at least in an ex post sense[7].

The board can observe all parameters, but cannot observe CEOs' skill levels α. Therefore, when the board observes high firm-specific profitability, it cannot be sure whether this is due to CEO skill (i.e. high α) or luck (i.e. high ϵ_t). When the board hires a new CEO, it starts with normally distributed prior beliefs about his ability:

$$\alpha \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right). \tag{6}$$

The board's prior beliefs match the distribution of skill α in the CEO talent pool. Therefore, parameter σ_0 plays two roles. It is both the initial uncertainty about a newly hired CEO's skill, and also the dispersion in true skill in the population of potential CEO replacements. Each period, the board updates its beliefs about ability α according to Bayes' Rule, using information contained in firm-specific profitability y_t, and z_t, which is an additional signal of CEO ability. The additional signal represents all information held privately in the firm (e.g. the CEO's specific actions and choices, the performance of individual projects, the CEO's strategic plan, turnover in other senior management), as well as public information (e.g. stock returns, sales growth, market share, discretionary earnings accruals, media coverage). The signal z_t contains all information arriving in period t which is not already contained in the firm's profitability. Without loss of generality, I treat this additional information as independent of profitability and centered at the CEO's skill, α. I also assume the signal is normally distributed with constant volatility, and is iid over time:

$$z_t \sim \mathcal{N}\left(\alpha, \sigma_z^2,\right). \tag{7}$$

The signal z is more precise when its volatility σ_z is lower.

Like all models, this model presents a simplified view of the world. The simplifications allow me to

[7]The board's firing decisions are optimal ex post, because the board optimizes each period and cannot commit up front to a different long-run policy. Later I discuss whether personal turnover costs may be optimal for shareholders ex ante in a more general model that allows long-run commitments.

obtain predictions from the model and identify parameter values from the data. In Section 4 I discuss several elements which are missing from my model, including firm fixed effects, time-varying personal turnover costs, CEO learning on the job, fluctuating CEO skill, board risk aversion, and earnings manipulation.

B. Solving the Model

First I solve the board's learning problem, which is a Kalman filtering problem. I introduce notation to distinguish between μ_t^{inc}, the posterior mean of the incumbent CEO's skill α going into period t, and μ_t, the prior mean of the CEO chosen to serve in period t. If the firm decides not to fire the incumbent, then $\mu_t = \mu_t^{inc}$, otherwise $\mu_t = \mu_0$. I also use the notation $\kappa_\epsilon \equiv \sigma_\epsilon^2 / \left(\phi^2 \sigma_0^2\right)$, and $\kappa_z \equiv \sigma_z^2 / \sigma_0^2$. The surprises in persistence-adjusted profitability and the additional signal equal

$$\delta_{z,t} \equiv z_t - \mu_t \tag{8}$$

$$\delta_{y,t} \equiv \frac{1}{\phi}\left(y_t - y_{t-1}\right) + y_{t-1} - \mu_t = \alpha + \frac{1}{\phi}\epsilon_t - \mu_t. \tag{9}$$

I show in Appendix A[8] that the posterior mean equals the prior mean plus two mean-zero shocks, one from the profitability surprise and one from the z_t surprise:

$$\mu_{t+1}^{inc} = \mu_t + \theta_y\left(\tau_t\right)\delta_{y,t} + \theta_z\left(\tau_t\right)\delta_{z,t}, \tag{10}$$

$$\theta_y(\tau) \equiv \kappa_\epsilon^{-1}\left(1 + (\tau+1)\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right)^{-1} \tag{11}$$

$$\theta_z(\tau) \equiv \kappa_z^{-1}\left(1 + (\tau+1)\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right)^{-1}. \tag{12}$$

The posterior mean follows a random walk with no drift. The board rationally ignores the industry component of profitability, v_t, which contains no information about the CEO's skill. Also, the board adjusts for persistence in profitability (Equation 9). Appendix A shows that uncertainty about a CEO's ability drops deterministically with tenure as the board learns.

The following proposition characterizes the board's optimization problem.

Proposition 1 *(Bellman equation): The board's objective function can be simplified as*

$$\frac{U_t}{\kappa B_t} = E_t\left[\sum_{s=0}^{\infty}\beta^s v_{t+s}\right] + \kappa\left(\frac{1-\phi}{1-\beta(1-\phi)}\right)y_{t-1} + V\left(\mu_t^{inc}, \tau_t, b_t\right), \tag{13}$$

where the value function $V(\mu, \tau, 0)$ solves the Bellman equation

$$V(\mu, \tau, 0) = \max\{V_{fire}, V_{keep}(\mu, \tau)\}, \tag{14}$$

$$V_{fire} = V(\mu_0, 0, 0) - c \tag{15}$$

$$c = c^{(firm)} + c^{(pers)}/\kappa \tag{16}$$

$$V_{keep}(\mu,\tau) = \left(\frac{\phi}{1-\beta(1-\phi)}\right)\mu + \beta f(\tau)\, V(\mu,\tau,1) + \tag{17}$$

$$\beta\left(1 - f(\tau)\right) E\left[V\left(\mu + \theta_X(\tau)\,\delta_X + \theta_z(\tau)\,\delta_z, \tau+1, 0\right)\right] \tag{18}$$

$$\begin{pmatrix}\delta_X \\ \delta_z\end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_\epsilon^2/\phi^2 + \sigma^2(\tau) & 0 \\ 0 & \sigma_z^2 + \sigma^2(\tau)\end{bmatrix}\right),$$

[8] All technical appendices are available in a separate document available on the author's website: Insert website address here.

subject to a boundary condition if the CEO has just retired:

$$V(\mu, \tau, 1) = V(\mu_0, 0, 0) - c. \tag{19}$$

Proof in Appendix B.

Equation (13) shows that the board's objective function is the sum of an industry-specific component, a component due to persistence in profitability, and a component V which depends on the CEO's posterior mean skill and tenure in office. Each period the board makes a firing decision by comparing its utility from firing the CEO (V_{fire}) and not firing him (V_{keep}) (Equation 14). Expression (15) shows that after firing the CEO, the board hires a new one and incurs the firing cost; the firing utility V_{fire} is constant over time. The board's decision depends on the total κ-adjusted turnover cost, defined in equation (16), not on the firm and personal costs separately. I call $c^{(pers)}/\kappa$ the effective personal cost. In equation (17), the utility V_{keep} from keeping the CEO depends on his expected contribution this period (the μ term) and the expected utility V next period, which in turn depends on whether the CEO quits (with probability $f(\tau)$) at the end of the period. If the CEO does not quit, he enters next period with posterior mean $\mu' = \mu + \theta_X(\tau)\delta_X + \theta_z(\tau)\delta_z$ (from the learning rule), and one more year of tenure (hence $\tau + 1$). The boundary condition in equation (19) shows that following a voluntary succession, the board hires a new CEO and pays cost c.

I obtain an approximate solution for $V(\mu, \tau, 0)$ by discretizing the state space and iterating on the Bellman equation, as described in Appendix C. I obtain additional predictions by simulating CEO spells from the model. I define a CEO spell as all the periods a CEO serves in office. To simulate a single spell, I draw the CEO's true skill α from the prior distribution, I generate firm-specific profitability y_t and additional signals z_t using the CEO's true skill α, and I update the board's beliefs according to the learning rule in equation (10). Simulated CEOs are fired according to the optimal rule (next section), and they leave office voluntarily with probability $f(\tau)$.

C. Model Predictions

In this subsection I discuss the model's predictions about the board's firing policy, the frequency and timing of firings, and turnover's relation to profitability and stock returns. These predictions hold for a wide range of plausible parameter values. However, since I solve the model numerically, I cannot prove the predictions hold for all parameter values, and so I do not present these predictions as formal propositions.

The Board's Firing Policy

The board fires the CEO as soon as its assessment of the CEO's skill, i.e., the posterior mean of α, drops below an endogenous threshold. The threshold depends on all model parameters, as well as the number of periods the CEO has been in office. Raising the total turnover cost c shifts the firing threshold down, making firings less likely. Intuitively, when firing the CEO is more costly, the CEO must have lower perceived skill to make firing him worth it. This result does not depend on whether cost c is larger due to higher firm turnover costs $c^{(firm)}$ or higher effective personal turnover costs $c^{(pers)}/\kappa$ (recall Equation 16).

The firing threshold increases with tenure, meaning the board becomes more willing to fire a CEO the longer he has been in office, all else equal. The explanation relates to uncertainty and the CEO's option value. When the firm hires a CEO, it acquires an option to fire him. All else equal, firms prefer higher uncertainty about CEO skill, because higher uncertainty raises the option's value, and the board is risk neutral. Intuitively, firms enjoy the upside of uncertainty by keeping CEOs who end up being highly skilled, but firms avoid the downside of uncertainty by firing CEOs who end up being unskilled. A CEO's uncertainty drops with tenure as the board learns about his skill, so his option value declines and the board becomes more willing to fire him. Hermalin and Weisbach (1998) make a similar prediction.

The Frequency and Timing of CEO Turnover

I illustrate the model's predictions using the following parameter values: $\beta = 0.9$, $\mu_0 = 1\%$, $\sigma_0 = 2\%$, $\sigma_\epsilon = 3\%$, $c = 3\%$, $\phi = 0.12$, and $\sigma_z = 7\%$. These parameter values are close to the empirical estimates in Section 3. I assume CEOs retire if they complete 15 periods, but not before then.

The top panel of Figure 1 plots firing hazard hazard rates vs. tenure for three values of total turnover cost c. When firing the CEO is costless ($c = 0$), slightly more than half of CEOs are fired after just one period. If the CEO disappoints expectations in his first period (which happens 50% of the time), then his posterior mean ability is below his replacement's, and the board is better off hiring the replacement at no cost. The gains in option value from a hiring a new, uncertain CEO push the firing rate slightly above 50%.

INSERT FIGURE 1 NEAR HERE

Not surprisingly, firing rates are lower and more CEOs survive to retirement when the total turnover cost c is higher. Consistent with this prediction, Parrino (1997) finds that forced CEO turnover is more likely in homogenous industries. As Parrino notes, the real costs of firing a CEO are probably lower when the firm can find a replacement in a similar firm.

Turnover costs also affect the timing of firings. Firings typically happen at later tenures when turnover costs are higher. Hazard rates decline monotonically when $c = 0$, are hump-shaped when $c = 3\%$, and increase monotonically when $c = 5\%$. Intuitively, the board is cautious and waits for more information when firing the CEO is more costly.

The bottom panel of Figure 1 plots firing hazard rates for three values of prior uncertainty about CEO skill, σ_0. Forced turnover is more likely when σ_0 is higher, meaning there is more variation in skill across

CEOs. To my knowledge, this reduced-form prediction has never been tested empirically. Intuitively, if CEOs are roughly alike in terms of their ability (i.e., low σ_0), then there is not much incentive to replace one CEO with another, especially when doing so is costly.

Prior uncertainty also affects the timing of firings. Firings occur later when prior uncertainty σ_0 is lower. The hazard function is downward sloping when $\sigma_0 = 3\%$, hump shaped when $\sigma_0 = 2\%$, and upward sloping when $\sigma_0 = 1\%$. The board learns more slowly when prior beliefs are stronger (i.e. lower σ_0), so it takes longer for the board to decide whether to fire the CEO. Supporting this intuition, the model also predicts that CEOs are fired later when the additional signal's volatility σ_z is higher, profit volatility σ_ϵ is higher, and persistence parameter ϕ is lower, all of which cause the board to learn more slowly. The board learns more slowly when profitability is more persistent (lower ϕ), because profitability spends less time closer to its long-run mean, the CEO's skill level α. In other words, when profitability is more persistent, it reacts more slowly to CEOs' actions and is therefore less informative about CEO skill. To my knowledge, no one has tested the reduced-form prediction that firings occur later in office when boards learn more slowly about CEO skill.

CEO Dismissals and Profitability

Figure 2 illustrates the turnover-profitability relationship by plotting average firm-specific profitability in event time around CEO dismissals. The figure also shows the average across CEOs of $\mu_t = E_t[\alpha]$, the board's posterior mean of the CEO's ability α.

INSERT FIGURE 2 NEAR HERE

First I discuss the pattern in beliefs about CEO skill, μ_t. The posterior mean drops gradually leading up to forced CEO turnover at time zero. In other words, the board's opinion of the CEO deteriorates leading up to his firing. We know this must occur in order for the posterior mean to drop below the firing threshold, so that the CEO is fired at time zero. The posterior mean jumps up to the prior mean following turnover, because the firm hires a new CEO and starts with new prior beliefs about his skill. The average posterior mean creeps up after period zero, because some CEOs are fired after year zero, and survivors are perceived to be more skilled than the average new hire.

Next I discuss the pattern in average realized excess profitability. As noted above, the posterior mean skill μ_t must drop before the CEO is fired. To cause downward revisions in posterior mean skill μ_t, either excess profitability y_t or the signal z must repeatedly be lower than expected. On average, excess profitability y_t drops leading up to forced CEO turnover at time zero. Consistent with this prediction, Weisbach (1988), Murphy and Zimmerman (1993), Huson, Parrino, and Starks (2001), and others find that CEO turnover is more likely following low profitability; Huson, Malatesta, and Parrino (2004) document a drop in profitability before forced CEO turnover.

Average excess profitability rises after the CEO is fired. This prediction is due to a replacement effect and a learning effect[9]. First, the replacement effect: compared to the fired CEO, the new CEO is more skilled on average and hence can generate higher profits. The learning effect is more subtle. Posterior mean skill must drop for the CEO to be fired. In order to "pull down" the posterior mean via Bayesian updating, realized excess profitability typically drops below the posterior mean by the time the CEO is fired. In Figure 2, this shows up as a gap between realized profitability and the CEO's posterior mean skill at time zero. Even if the firm hires a new CEO with the same low perceived skill, there is still a gap between realized and expected profitability going forward. Therefore, realized profitability is expected to rise, closing the gap. This learning effect is similar to the one in Pástor, Taylor, and Veronesi (2007). For this particular calibration, the learning and replacement effects are roughly equal in magnitude.

Profitability is not a perfect predictor of CEO dismissal in the model, because the board also uses the signal z_t to evaluate CEO ability. Even if a CEO achieves very low profitability, he may avoid being fired if the z_t signal is high enough. For instance, the board may learn that the low earnings were due to bad luck. The opposite is also possible. Even if profits are not low, the board may nevertheless fire the CEO if the additional signal is low enough. For example, the board may observe the CEO making poor choices and may fire the CEO before his choices affect profits.

The top panel of Figure 3 illustrates this effect. It shows average changes in excess profitability in event time around forced turnover for three values of σ_z, the volatility of the board's additional signal. The V-shaped pattern in profitability is less pronounced when the signal z is more precise, i.e., its volatility σ_z is lower. The reason is that the board rationally relies more on z_t and less on profitability to evaluate CEO skill, so the board is less likely to fire the CEO when profitability is low, which produces a smaller average drop in profitability around CEO dismissals.

INSERT FIGURE 3 NEAR HERE

Firm turnover costs also affect the turnover-profitability relation. The bottom panel of Figure 3 plots average profitability in event time for three values of the firm turnover cost $c^{(firm)}$, holding constant the total turnover cost $c = c^{(firm)} + c^{(pers)}/\kappa$. Profitability drops more around forced turnover when firm turnover costs $c^{(firm)}$ are larger. This prediction follows mechanically from the assumption that firm turnover costs reduce profitability by $c^{(firm)}$ in the turnover period.

Figure 3 highlights two potential reasons why profitability is low around CEO dismissals: (1) Low profitability causes a forced turnover (top panel), and (2) forced turnover causes low profitability (bottom panel). Reason (1) is due to learning, and reason (2) is due to firm turnover costs. Empirically disentangling these two effects is a challenge. However, Figure 3 suggests a solution. The change in profitability from

[9] Huson, Malatesta, and Parrino (2004) provide empirical evidence of a post-dismissal increase in firm profitability. They attribute this result to an improvement in CEO ability (which I call the replacement effect) as well as "bad luck" for the fired CEO, which is similar to my learning effect.

event year -2 to -1 depends on σ_z (the volatility of signal z) but not $c^{(firm)}$ (the firm turnover costs), so this change in profitability helps measure σ_z. Intuitively, from years -2 to -1, the board has not yet decided to fire the CEO, so the firm turnover costs do not yet show up in profitability. The change in profitability from years -1 to 0 depends on both z's volatility and firm costs. Since we have already pinned down z's volatility, we can back out firm costs. I will rely on this result in the empirical section to identify firm costs $c^{(firm)}$ and signal z's volatility σ_z.

To my knowledge, no one has tested Figure 3's reduced-form prediction that the turnover-profitability relation is stronger when the board's additional signals of CEO ability are less precise. The model also predicts that the turnover-profitability relation is stronger when profitability is less volatile (low σ_ϵ) and less persistent (high ϕ), both of which make profitability more informative of CEO ability. Consistent with these predictions, Engel et al. (2003) find a stronger empirical turnover-profitability relation when profitability is less volatile and more sensitive to CEOs' actions. Also, Wu and Zhang (2008) find a stronger turnover-profitability relation after firms adopt international accounting standards that improve transparency.

Several empirical papers estimate logit or probit models that forecast CEO dismissals using lagged profitability (e.g. Kaplan and Minton (2006)). My model predicts that multiple lags of firm-specific profitability help forecast forced turnover, because a CEO's posterior mean skill depends on profitability signals over all his periods in office, not just the most recent one or two periods. Consistent with this prediction, Kim (1996) finds that up to 14 years of lagged performance help explain CEO turnover. Also, the model predicts that probit slopes vary with the CEO's tenure in office, because the firing threshold, dynamics of beliefs, and distribution of surviving CEOs' skill all change with tenure in office. Kim (1996) shows that, consistent with this prediction, recent years' performance has less explanatory power when the CEO has been in office longer. My model suggests two possible reasons why. First, new profitability signals have less effect on beliefs after several periods of learning have already occurred. Also, CEOs who survive several years in office are perceived to be highly skilled, so new signals are less likely to pull the CEO below the firing threshold.

Stock Returns around CEO Dismissals

The board's objective function U_t equals the firm's market value in the special case where investors and the board have common beliefs, there are no personal costs of CEO turnover, and the board and investors use the same discount rate β. To build intuition, I obtain predictions about stock returns under these additional assumptions. However, I do not use stock returns to estimate the model, for two main reasons. First, Weisbach (1988) and Murphy and Zimmerman (1993) find empirically that profitability outperforms stock returns as a predictor of CEO turnover. Hermalin and Weisbach (1998) and Bond, Goldstein, and Prescott (2008) provide a theoretical explanation for this result, namely, that earnings reflect the quality of current management, whereas stock prices reflect both current management and the expectation of future management. Second, the additional assumptions I use to obtain stock returns are quite strong— especially the assumption of symmetric information—, so using stock returns as well as profitability to estimate the model may increase rather than reduce estimation error. Obtaining predictions for stock returns in the more general case with asymmetric information is left for future work.

Details on obtaining stock returns are in Appendix D. Figure 4 plots the average cumulative abnormal return (CAR) in event time around forced CEO turnover. I also plot the empirical pattern, which I discuss in Section 3.B. The model predicts that the CAR drops gradually during the five years before the CEO is fired. One reason is that profitability drops gradually (recall Figure 2), which results in lower dividends and lower expected future profitability due to persistence. Another reason is that beliefs about CEO skill decline (also Figure 2), which also results in lower expected future profitability. Also, investors begin to anticipate forced CEO turnover, which would bring real turnover costs (pushing the stock price down) and a CEO who is expected to be more talented (pulling stock prices up). The model predicts no more abnormal returns after the CEO dismissal, because there is no more conditioning on a future event, and investors rationally anticipate future changes profitability.

INSERT FIGURE 4 NEAR HERE

II. Estimation

A. Data

Data come from Compustat and a CEO turnover database constructed using the method of Huson, Parrino, and Starks (2001). The sample consists of CEOs in the *Forbes* annual compensation surveys who left office between 1971-2006. To my knowledge, this sample is larger than any found in the existing CEO turnover literature. I am grateful to Robert Parrino for providing CEO turnover data from 1971-1994. CEO successions are classified as forced or unforced, as described in Huson, Parrino, and Starks (2001), p.2273:

> First, if the *Wall Street Journal* reports that the CEO is fired, forced from the position, or departs due to unspecified policy differences, the succession is classified as forced. For the remaining cases, the succession is classified as forced if the departing CEO is under the age of 60 and the *Wall Street Journal* announcement of the succession (1) does not report the reason for the 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. The circumstances surrounding the departures of the second group are further investigated by searching the business and trade press for relevant articles in order to reduce the likelihood that a turnover is incorrectly classified as forced. 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.

I interpret one model period as a year, and I assign successions to the closest fiscal year end. I use data from all years a CEO spent in office, including years before 1971. To avoid estimation bias from this sampling method, I use the same sampling method in my simulation estimator (Section 2.C). The sample does not include CEOs who left office due to a takeover, whose succession was not announced in the *Wall Street Journal*, or who have missing data in Compustat. I set firm profitability Y_t equal to the firm's return on assets (ROA) in year t.[10] Industry average profitability v_t is an equal-weighted average of ROA across firms in each of 12 industries defined on Kenneth French's website. When computing v_t I use each year's 1,000 largest firms (by lagged assets) in Compustat to avoid bias from changes in Compustat's coverage.

Table 1 contains summary statistics. The full sample contains 981 CEOs and 7,325 firm/year observations. Out of the 981 successions in my sample, 168 (17.1%) are forced. On average, 2.29% of sample firms fire their CEOs in a given year. Forced CEO turnover appears to be more common in some industries than others. For example, CEOs in the business equipment industry face a 3.80% average chance of being forced out each year, compared to 0.69% for CEOs in the chemicals industry. Forced turnover increases steadily over the time period, increasing from 7.8% of successions from 1970-1974 to 24.7% of successions in 2005-2006. Kaplan and Minton (2006) also document an increase in forced turnover. Panel B indicates that the mean industry-adjusted profitability is 2.00%. This average is not mechanically zero, because I compute mean industry profitability using 1,000 firms each year, and my full sample contains fewer (and apparently more profitable) firms. Subtracting off industry profitability reduces the volatility in profitability from 9.07% per year (for ROA) to 7.37% (for ROA$-v_t$). The median CEO spell lasts 6 years. Median spell length is shorter for CEOs forced out of office (4 years) compared to CEOs who leave voluntarily (7 years).

INSERT TABLE 1 NEAR HERE

B. Identification and Additional Restrictions

This subsection discusses additional assumptions which deliver an identified model and improve precision. I also provide intuition for how the model is identified.

The model assumes parameters are constant over a firm's lifetime. Solving the board's optimization problem depends on this assumption, which I maintain here. With a long enough time series for each firm, we could exploit variation across CEOs to identify firm-specific parameters. Allowing heterogeneity across firms is challenging, because there are only 168 forced successions in my sample. Estimating the model in

[10] Annual ROA equals operating income before depreciation (Compustat annual item 13) divided by the midpoint of assets (item 6) from the current and previous fiscal years. I eliminate ROA observations outside [-100%,100%].

industry subsamples, for example, would result in samples that are too small. In order to obtain more precise estimates, I assume parameters are constant across firms. This same assumption is implicit in much of the reduced-form empirical work on CEO turnover. For instance, when Kaplan and Minton (2006) estimate a probit model of CEO turnover on lagged firm performance measures, they implicitly assume the probit slopes are constant across all firms and years in their pooled sample. In a later subsection I estimate the model in subsamples.

I do not attempt to estimate the board's discount factor β. Instead, I fix the value of β at 0.9, a plausible value given firms' annual cost of capital. For robustness, in Section 4 I estimate the model using other values of β. To explain the intuition for how the remaining parameters are identified, I assume we have data on many CEO spells governed by the same parameters— either a long time series for a single firm, or a large cross section of firms sharing the same parameters.

I start with the parameters governing firm-specific profitability and prior beliefs. Time-series autocorrelation in firm-specific profitability helps identify ϕ, the profitability persistence parameter. After removing the persistent component in profitability, profitability volatility within CEO spells will help identify σ_ϵ, the volatility of profitability shocks. The average level of profitability helps identify the prior mean skill μ_0, and dispersion in average profitability across CEOs helps identify dispersion in CEO skill, σ_0. The endogenous firing decisions complicate identification, but the following trick helps explain the intuition. In the subsample containing only CEOs' first year in office, the board has made no turnover decision yet, so there are no endogeneity concerns in this subsample. For each CEO i in the subsample we can compute persistence-adjusted profitability $X_{i,0}$ from his first year in office:

$$X_{i,0} \equiv \left(y_{i,0} - y_{i,-1}\right)/\phi + y_{i,-1} = \alpha_i + \epsilon_{i,0}/\phi. \tag{20}$$

The mean of $X_{i,0}$ across CEOs equals the prior mean CEO skill μ_0, since shocks $\epsilon_{i,0}$ have mean zero. The variance of $X_{i,0}$ across CEOs equals $\sigma_0^2 + \sigma_\epsilon^2/\phi^2$. The first term in the sum comes from dispersion in skill α across CEOs, and the second term comes from the iid shocks firms receive. Knowing ϕ and σ_ϵ, we can back out σ_0, the dispersion in skill across CEOs.

The frequency of forced turnover at different tenures helps identify total turnover costs $c = c^{(firm)} + c^{(pers)}/\kappa$. As we saw in Figure 1, forced turnover is less frequent when the cost c is higher.

Changes in average firm-specific profitability around CEO dismissals help disentangle the firm turnover cost $c^{(firm)}$ and effective personal cost $c^{(pers)}/\kappa$, and also help identify signal z's volatility, σ_z. As we saw in

Figure 3, there are two reasons why average profitability drops leading up to CEO dismissals. First, learning contributes to a gradual drop in average profitability leading up to forced turnovers, and then a gradual recovery. The larger σ_z is, the stronger the profitability-turnover relationship, and the larger the gradual dip in profitability around forced successions. Second, firm turnover costs make profitability drop and then rise abruptly around forced turnovers. We can therefore infer z's volatility σ_z and firm costs $c^{(firm)}$ by measuring how much profitability drops around CEO dismissals, and decomposing the drop into a gradual component due to learning and an abrupt component due to firm turnover costs. This identification strategy exploits two assumptions embedded in equation (1): first, firm turnover costs affect profits, but personal turnover costs do not; and second, the firm turnover costs are realized in the period when the CEO is fired. Since CEO successions do not perfectly line up with fiscal year ends in my data, I relax (2) slightly and assume that half the firm costs are realized in the CEO's last year in office, and half are realized in the new CEO's first year in office.

Once we know the firm cost $c^{(firm)}$ and the total cost c, we can back out the effective personal cost: $c^{(pers)}/\kappa = c - c^{(firm)}$. The ratio $c^{(pers)}/\kappa$ is identified, but $c^{(pers)}$ and κ are not. In other words, we cannot distinguish between a board with a strong distaste for firing the CEO (large $c^{(pers)}$) and a board that does not care much about shareholder value (low κ). I only report estimates of the ratio $c^{(pers)}/\kappa$.

C. Estimation Method

I estimate the parameters $\theta \equiv \left\{\mu_0, \sigma_0, \sigma_z, \sigma_\epsilon, \phi, c^{(firm)}, c^{(pers)}/\kappa\right\}$ using the simulated method of moments (SMM) [11]. Like GMM, SMM estimates parameter values by matching certain data moments and model-implied moments as closely as possible. Whereas GMM uses closed-form expressions for the model-implied moments, SMM estimates the model-implied moments using simulations. This project lends itself to SMM estimation, because few closed-form expressions are available. The SMM estimator $\widehat{\theta}$ is

$$\widehat{\theta} \equiv \arg\min_{\theta} \left(\widehat{M} - \frac{1}{S}\sum_{s=1}^{S} \widehat{m}^s(\theta)\right)' W \left(\widehat{M} - \frac{1}{S}\sum_{s=1}^{S} \widehat{m}^s(\theta)\right). \tag{21}$$

$\widehat{M}$ is a vector of estimated moments from the empirical data, and $\widehat{m}^s(\theta)$ is a vector of estimated moments from the sth sample simulated using parameters θ. I use 14 moments, defined below. Since my empirical sample contains 981 CEO spells, each simulated sample contains 981 CEO spells as well. Michaelides and Ng (2000) find that using a simulated sample 10 times as large as the empirical sample generates good small-sample performance. I use $S = 20$ simulated samples to be conservative.

[11] For example, McFadden (1989), Pakes and Pollard (1989), Rust (1994), Hennessy and Whited (2005, 2007).

I sample the real and simulated data the same way to avoid sampling bias. I collect all simulated CEOs who left office between simulation years 1971-2006, and I bring in profitability data from all years a CEO spent in office, including years before 1971. I begin simulations in year 1900 to ensure the model has reached a steady state by 1971.

W can be any positive definite weighting matrix. I use the efficient weighting matrix, which is the inverse of the estimated covariance of moments M. Following Rust (1994) and Hennessy and Whited (2005, 2007), I use a simulated annealing optimization algorithm to avoid local minima of (21). Additional details on the procedure are in Appendix E.

I estimate the seven parameters using 14 moments. The extra moments will provide a test of the model's over-identifying restrictions. All moments use data on excess profitability y_t^*, which equals firm profitability Y_t minus industry profitability v_t. At this point I add i subscripts to index the firms in the sample.

The first seven moments are the coefficients from the pooled regression

$$y_{it}^* = \lambda_0 + \lambda_1 y_{i,t-1}^* + \Delta^{(-2)} + \Delta^{(-1)} + \Delta^{(0)} + \Delta^{(1)} + \Delta^{(2)} + \delta_{it}. \tag{22}$$

The coefficient $\Delta^{(k)}$ is a fixed effect for whether firm i experienced forced CEO in period $t-k$. The intercept λ_0 will help pin down the prior mean skill μ_0, and the slope λ_1 will help pin down the persistence parameter ϕ. The fixed effects $\Delta^{(k)}$ measure the changes in average profitability around forced turnovers. As discussed in Section 2.B, these changes help measure z's volatility σ_z and firm turnover costs $c^{(firm)}$. The eighth moment is $Var(\delta_{it})$, the variance of the residual from equation (22), which is most informative and the time series volatility of profitability, σ_ϵ.

The next four moments are forced turnover hazard rates. I define $h^{(k)}$ to be the percent of CEOs fired per year in tenure category (k) years, conditional on the CEO reaching (k). I use hazard rates $h^{(1-2)}$, $h^{(3-4)}$, $h^{(5-7)}$, and $h^{(8+)}$. These four rates will help pin down c, the total costs of forced CEO turnover. Using total costs c and firm costs $c^{(firm)}$, the model can infer effective personal costs according to $c^{(pers)}/\kappa = c - c^{(firm)}$.

The last two moments help tease apart σ_ϵ and σ_0, both of which affect variation in profitability. Both moments use data on persistence-adjusted profitability $\widehat{X}_{it} \equiv \left(y_{it}^* - \widehat{\lambda}_1 y_{it-1}^*\right) / \left(1 - \widehat{\lambda}_1\right)$, where $\widehat{\lambda}_1$ is estimated in regression (22). First, for each CEO j I compute $E_j[X_{it}]$ and $Var_j(X_{it})$, respectively, the mean and variance of $\widehat{X}_{it}$ across all the years CEO j spent in office. The 13th moment is $E\left[Var_j\left(X_{it}\right)\right]$, the mean of CEOs' variances. Since this moment removes the effect of each CEO's ability, it is most informative about σ_ϵ, the time-series volatility of profitability. The 14th moment is $Var(E_j[X_{it}])$, the variance of CEOs' means.

This moment is most informative about σ_0, the dispersion in ability across CEOs, because it measures the cross-CEO dispersion in a proxy for a CEO's ability, i.e., his average realized profitability.

Finally, the hazard function for voluntary turnover, $f(\tau)$, is an input to the model. I estimate $f(\tau)$ directly from the CEO turnover database, calculating the frequency of voluntary turnover after τ years conditional on the CEO surviving $\tau - 1$ years, pooling all CEO spells. The hazard rate is low when the CEO first starts in office, and then rises gradually (results available on request).

III. Empirical Results

A. Parameter Estimates

Parameter estimates are in Table 2. The estimated turnover cost to the firm, $c^{(firm)}$, is 1.33% of the firm's assets. This cost is significantly positive, with 95% confidence interval [0.1%, 2.5%]. The firm turnover cost is \$53 million in 2007 dollars for the median sample firm. The estimated firm cost is larger than known CEO succession costs. For instance, Yermack (2006) reports average separation payments to CEOs of \$18 million and \$2 million for forced and voluntary successions, respectively. Fees to executive search firms are on the order of \$1 million[12]. Since we lack direct measurements of other, less tangible CEO succession costs to shareholders, it is unclear whether my estimated firm cost is unreasonably large.

INSERT TABLE 2 NEAR HERE

The estimated effective personal turnover cost, $c^{(pers)}/\kappa$, is even larger, at 4.61% of the firm's assets (95% confidence interval: [3.5%, 5.7%]), or \$183 million for the median firm. Adding together the firm and personal costs, the board behaves as if firing the CEO costs shareholders 5.9% of the firm's assets. However, firing the CEO really only costs shareholders $c^{(firm)} = 1.3\%$ of assets. This gap makes the board retain some CEOs whom shareholders would rather see fired. In this sense, the large effective personal cost implies a high degree of CEO entrenchment. Consistent with high entrenchment, CEOs had considerable influence on the choice of directors during much of the period I investigate[13]. Since I can only estimate the ratio $c^{(pers)}/\kappa$, I cannot determine whether this entrenchment is due to a strong distaste for firing CEOs (large $c^{(pers)}$) or a board which does not care much about shareholder value (low κ). If boards internalize 100%

[12]The industry standard CEO search fee is one-third of the CEO's total cash compensation in his first year in office, so the average search fee in my sample is roughly \$1 million.

[13]Under corporate law, shareholders choose the board of directors. However, DeAngelo and DeAngelo (1989) show that shareholders almost always approve the slate which management proposes. The CEO approves and often proposes the slate (e.g. Mace (1971), Lorsch and MacIver (1989), Demb and Neubauer (1992)). Shivdasani and Yermack (1999) provide additional empirical evidence. In Section 4.B I examine changes over time in entrenchment.

of shareholder value ($\kappa = 100\%$) then the personal cost is the full \$183 million. However, if boards only internalize 1% of shareholder value ($\kappa = 1\%$) then the personal cost is only \$1.83 million.

To summarize the main results so far, the model needs huge turnover costs to fit the data, and these costs mainly reflect CEO entrenchment rather than a real cost to shareholders. Entrenchment does not necessarily imply bad governance. Some degree of entrenchment may be optimal for shareholders ex ante. I discuss this issue and present additional evidence in Section 3.C.

The estimated prior mean skill μ_0 is 0.88% per year, slightly less than than the 2.0% average industry-adjusted profitability in the sample (Table 1). This result is expected. Since the average skill across surviving CEOs mechanically equals average industry-adjusted profitability (2.0%), and since the average newly hired CEO is less skilled than the average surviving CEO, the prior mean skill is less than 2.0%.

Parameter σ_0 is both the standard deviation of ability across new CEOs, and also the uncertainty about a newly hired CEO's ability. The estimate of σ_0 is 2.42% of assets per year. For comparison, Bertrand and Schoar (2003) estimate manager-specific fixed effects in annual ROA. They find a 7% standard deviation in fixed effects across managers, implying even greater dispersion in ability than reported here. The estimate of σ_0 implies that the difference in average industry-adjusted ROA generated by a newly hired CEO at the 95th ability percentile and one at the 5th percentile is $2 \times 1.65 \times \sigma_0 = 8.0\%$ per year. The interval's width implies that CEO skill matters greatly, in the sense that the difference in expected profitability between a talented and untalented CEO is several percent per year.

The estimated persistence parameter ϕ is 0.125, indicating that firm-specific profitability nearly follows a random walk. In contrast, Fama and French (2000) estimate persistence parameters for ROA roughly equal to 0.6, suggesting that profitability is closer to iid.[14] Our results are different because we estimate fundamentally different economic rates. Fama and French measure mean reversion around firm-specific average profitability, whereas my persistence parameter measures mean reversion around industry average profitability. To illustrate the difference, when I estimate a panel regression of excess profitability on its lag, the estimated slope is 0.89, which is close to my estimate of one minus persistence parameter ϕ. When I estimate the same regression firm by firm, the firms' average slope drops to 0.59, because I have largely removed the effects of firm-specific average profitability. My model's notion of persistence and CEO skill seems plausible: CEOs are considered skilled not only if they can beat the firm's long-run average profitability, but also if they can increase the firm's long-run average profitability relative to the industry. For robustness, in Section 4.A I introduce firm fixed effects in profitability, which results in a higher estimate of ϕ.

[14] Fama and French (2000) find a rate of mean reversion of 37% per year. Their rate of mean reversion is analogous to one minus my persistence parameter.

To interpret the estimate of σ_z, the volatility of the board's additional signal, I compare the influence of the profitability signal and additional z signal on the board's beliefs about CEO skill. Specifically, I compare the change in posterior beliefs resulting from a one standard deviation z shock and a one standard deviation profitability signal shock. The model predicts that the response to the z shock is $P \equiv \sigma_\epsilon/(\phi\sigma_z)$ times larger than the response to the profitability signal shock[15]. The P ratio indicates that the additional z signal is more influential when it is more precise (σ_z lower), and when profits are noisier (σ_ϵ higher) and more persistent (ϕ lower). Applying the delta method, I obtain an estimate of P equal to 5.3, with a standard error of 0.3. In other words, the additional signal z has a 5.3 times larger influence on the board's beliefs, compared to the profitability signal. This result implies boards rely heavily on non-earnings information when evaluating the CEO. Consistent with this result, Bushman, Indjejikian, and Smith (1996) find that boards give considerable weight to information besides earnings and stock performance when determining a CEO's bonus. It is plausible that boards also use this additional information in firing decisions.

B. Model Fit

In this subsection I assess how well the estimated model fits empirical patterns in forced CEO turnover, firm profitability, and stock returns. The first test is a formal test of the overall model. Since I estimate 7 parameters using 14 moments, the SMM procedure delivers a χ^2 test of over-identifying restrictions (bottom of Table 3). The p-value rejects at the 1% confidence level the hypothesis that all 14 simulated moments equal the empirical moments. In other words, the data reject the model. I do not consider this result particularly damning, since any model will be rejected with enough data.

INSERT TABLE 3 NEAR HERE

Next I examine the 14 moments individually to gauge where the model fails. Each row in Table 3 shows a moment's empirical estimate, simulated value, standard error, and a p-value that tests whether the empirical and simulated moments are equal. For 4 out of 14 moments we can reject equality at the 5% confidence level; only two of these are reliably different at the 1% level. The model matches the intercept and AR slope from the profitability regression very closely, indicating the model can fit the long-run mean and persistence of firm-specific profitability. The fixed effects Δ measure average changes in profitability around CEO dismissals. The model matches all the Δ's except $\Delta^{(1)}$ fairly closely. I examine this pattern in

[15] This result follows from equations (10)-(12). A one standard deviation z shock corresponds to $\delta_z = \sigma_z$, which moves beliefs by $\theta_z(\tau)\sigma_z$. A one standard deviation X shock corresponds to $\delta_X = \sigma_\epsilon/\phi$, which moves beliefs by $\theta_X(\tau)\sigma_\epsilon/\phi$. Taking ratios,

$$\frac{\theta_z(\tau)\sigma_z}{\theta_X(\tau)\sigma_\epsilon/\phi} = \frac{\kappa_z^{-1}\sigma_z}{\kappa_X^{-1}\sigma_\epsilon/\phi} = \frac{\sigma_\epsilon}{\sigma_z\phi} \equiv P.$$

detail later. Turning to the forced turnover hazard rates $h^{(k)}$, the model produces too few firings in the first two years, too many firings in years 3 and 4, and not enough firings after year 7. However, the gap between simulated and empirical hazard rates is less than 1% per year for all four moments, and the model successfully produces the hump-shaped empirical relation between tenure and firings, which Allgood and Farrell (2003) also document. There are no significant differences between the empirical and simulated values for any of the second moments of profitability, $Var(\delta)$, $E[Var(X)]$ and $Var(E[X])$. The model therefore appears to closely match time-series volatility in profitability for a given CEO ($Var(\delta)$ and $E[Var(X)]$), and also the variation in realized profitability across CEOs, $Var(E[X])$.

The top panel of Figure 5 plots the percent of CEOs fired at different tenures, comparing the empirical and predicted patterns. The empirical points are within the model's grey 95% confidence region at almost all tenure levels, indicating a good fit. Almost 2% of CEOs are fired after their first year in office in the empirical sample. The fraction rises to almost 3% fired after CEOs' 2th year in office, and then it decays gradually with tenure. The model fits this hump-shaped pattern fairly closely. Because of its dynamic setup, the model captures an important reason why forced turnover is rare, namely, it takes time to learn about CEO ability, so CEOs often survive several years in office before being fired.

INSERT FIGURE 5 NEAR HERE

More measures of model fit are in Table 4. The model can exactly match the median tenure of CEOs forced out of office (4 years) and CEOs who leave voluntarily (7 years). In the real data, 17.1% of successions are forced, compared to 16.2% produced by the model. The model matches the low overall rate of forced turnovers quite closely, predicting that 2.16% of CEOs are fired each year on overage, compared to 2.29% in the real data. On these dimensions, the model fits the data quite well.

INSERT TABLE 4 NEAR HERE

The next diagnostics address the relation between profitability and forced CEO turnover. First, I estimate a probit model which use lagged firm-specific profitability to forecast whether a CEO is fired. I use one year of lagged profitability, although results are similar using three lags. The last columns in Table 4 compare results when I estimate the probit model using real and simulated data. Profitability has a statistically significant negative slope in both the real and simulated data, indicating that CEOs are more likely to be fired after a year of low profitability. The estimated slope from the real data is within two standard errors of the slope estimated from simulated data, indicating the model helps explain not just the negative correlation in the data, but also the magnitude. Using the real data, the pseudo-R^2 value is very low (3%), which is typical for such models (Brickley (2003)). In other words, profitability poorly predicts forced CEO turnover.

The model can fit this feature of the data quite closely, predicting a pseudo-R^2 of 2%. The model generates a weak profitability-firing relation because profitability has a small influence on board's beliefs about CEO skill. As discussed in Section 3.A, parameter estimates imply that the additional signal z has a 5.3 times larger influence on the board than the profitability signal has. Essentially, boards do not rely much on profits when evaluating a CEO's skill, because other, unrelated information is more reliable.

The bottom panel of Figure 5 shows average firm-specific profitability in event time around CEO dismissals, comparing real and simulated data. As expected, there is a V-shaped pattern in profitability around dismissals, both in the real and simulated data. The empirical pattern is within the model's grey 95% confidence region in all event years except year –5. Overall, the model can closely match the level of profitability, the magnitude of the changes in profitability, and the timing of the changes. The model has the most difficulty matching the change in profitability in the year after the CEO is fired (event years 0 to 1). In that year the model predicts a rise in profitability, whereas profitability drops in the data. The model would fit the data better if I assumed all firm turnover costs were realized in the year after the firing (year 1); I currently assume half are realized in each of years 0 and 1. The model would then attribute the empirical drop in profitability from years 0 to 1 to a larger firm turnover cost.

Finally, I assess how well the model matches average stock returns around CEO dismissals. As discussed in Section 1, I derive stock prices from the model by assuming boards and investors have common beliefs about CEO skill. Figure 4 compares the empirical and predicted cumulative abnormal return (CAR) in the five years around CEO dismissals. The model misses in terms of magnitudes. Leading up to CEO dismissal, the average empirical five-year CAR is –35%, compared to –18% from the model. The model's CAR is outside the 95% empirical confidence region. However, the model succeeds in matching the general shape. In both the model and the data, the CAR drops gradually leading up to forced turnover, and is essentially flat after.

C. Does the Personal Turnover Cost Reflect Bad Governance?

A central result of this study is that the model needs a large effective personal turnover cost to fit the data. I offer two extreme interpretations of the large personal cost, and then present evidence on which interpretation is closer to the truth.

The bad governance interpretation is that personal cost prevents boards from acting in shareholders' interests, even ex ante. Shareholders would prefer a board with no personal costs, i.e., a board that is more willing to fire the CEO. Shareholders cannot elect such a board in the first place because of problems with the governance system.

The good governance interpretation is that the personal cost and resulting CEO entrenchment are optimal for shareholders ex ante. By electing a board with large personal turnover cost, shareholders commit up front to a low probability of firing the CEO. This commitment may benefit shareholders by allowing them to pay the CEO less (Almazan and Suarez (2003), Hermalin and Weisbach (2007)), hire from a better CEO talent pool, or provide the CEO with incentives to make risky innovations (Manso (2007)). According to this view, the estimated 4.6% effective personal cost is simply the level of entrenchment that is optimal for shareholders ex ante.

If the estimated personal costs truly reflect bad governance, then personal costs should be smaller in firms or years with better governance. To test this hypothesis, I examine whether personal turnover costs are related to measures of governance quality. I split the sample using a measure of governance quality, estimate the model independently in each subsample, and then test whether personal costs are equal across the sub-samples.[16]

I use four measures of governance quality[17]. First, I create early and late subsamples based on whether the CEO left office between 1971-1989 or 1990-2006. Pointing to time trends in board composition and size, director compensation, and institutional stock ownership and activism, Huson, Parrino, and Starks (2001) argue that monitoring improved from 1970-2000. Monitoring continued to intensify due to the Sarbanes-Oxley Act in 2002. If governance quality has improved over time, then the bad governance story predicts lower personal turnover costs in the late subsample. Next, I form two subsamples based on whether the percent of the firm's shares owned by non-CEO officers and directors is above or below the median value, 1.31%. Boards owning more shares may care more about shareholder value, which in the model means they have a higher value of κ and hence a lower effective personal cost, $c^{(pers)}/\kappa$. Therefore, the bad governance story predicts lower effective personal costs in the high ownership subsample. Third, I form two subsamples based on whether the percent of directors who are not officers of the firm is above or below the median value, 72.7%. Fama and Jensen (1983) and Weisbach (1988) argue that outsider-dominated boards monitor management more effectively, because outsiders have fewer ties to the CEO and are more concerned with their labor market reputations. The bad governance story therefore predicts lower personal costs in the

[16] I perform simple split sample tests for two reasons. Including several control variables at once, as in a multiple regression, would be too expensive computationally. Splitting the sample along several dimensions (e.g. by industry and size) is problematic because the full sample contains only 168 forced successions, and the resulting subsamples would be extremely small.

[17] I thank Robert Parrino for providing data on board share ownership data (originally from proxy statements) and board composition (originally from the Million Dollar Directory). Both measures are available only up to 1994. I exclude CEO spells with missing governance measures.

subsample with more outsiders. Finally, I create a large-firm and small-firm subsample by comparing firms' inflation-adjusted assets to the sample median, \$6.6B. If shareholders face a fixed cost of monitoring a board, then they have a larger incentive to monitor boards of larger firms, which tend to make up more of their portfolio. Therefore, the bad-governance story predicts smaller personal costs in larger firms.

Parameter estimates for the sub-samples are in Panel A of Table 5. Panel B compares CEO tenures, firing rates, and firing sensitivities across subsamples and also between the real and simulated data.

INSERT TABLE 5 NEAR HERE

First I discuss subsamples split by year. Consistent with the bad governance interpretation, the estimated effective personal cost drops from 8.32% in the 1971-1989 subsample to 2.28% in the 1990-2006 subsample. This difference in personal costs between subsamples is economically large and statistically significant at the 1% level[18]. Although the personal cost drops, it remains significantly positive in the late subsample. The total turnover cost drops from 8.4% to 4.0% of assets. Lower turnover costs should result in more forced successions, all else equal. Indeed, the percent of successions that are forced rises from 12% to 23% between subsamples in the real data. The model generates a rise from 10% to 23%, a close match. Lower turnover costs are not the only reason firing rates rose over time, according to parameter estimates. Dispersion in ability across CEOs (σ_0) is significantly higher in the later subsample (difference has t-statistic=2.4), which raises the benefits of replacing the CEO and increases the speed of learning, both of which contribute to higher firing rates (Section 1).

Next I examine subsamples split by board stock ownership. Estimated personal costs are higher in the subsample with higher stock ownership (7.99% vs. 6.41% of assets), but this difference is not statistically significant (t statistic = 0.83). This result is not consistent with the bad governance story. Forced successions are more common in the high ownership subsample (17.6% compared to 12.6%, Panel B). Interestingly, the model attributes the difference not to lower turnover costs, but to more dispersion in ability across CEOs in the high-ownership subsample: σ_0 increases from 3.10% to 3.99% (difference has t-stat = 4.2). The model needs higher ability dispersion in order to fit the higher empirical dispersion in realized profitability across CEOs in the high ownership subsample: $\sigma(E[X])$ increases from 12.3% to 24.6% per year in Panel B.

Consistent with the bad governance story, personal turnover costs are significantly lower in the subsample with more outsiders on the board (3.00% compared to 8.25%; difference has t-stat 2.7). The lower turnover costs explain why more successions are forced in the subsample with more outsiders (16.0% vs. 11.6%).

[18] I conduct inference by assuming estimators from the two sub-samples are uncorrelated with each other. This assumption is plausible under the model's assumption that draws from the CEO talent pool, profitability shocks, and realizations of signal z are all independently distributed across both firms and time.

The difference in turnover costs outweighs an effect going in the opposite direction: there is more dispersion in ability across CEOs in the subsample with fewer outsiders (σ_0 of 2.93% vs. 1.98%, difference has t-stat 7.5), which pushes firing rates up in the subsample with few outsiders. Ability dispersion is higher because the model needs to fit the higher dispersion in realized profitability across CEOs in the subsample with few outsiders ($\sigma(E[X])$ increases from 10.4% to 17.9% in Panel B).

Also consistent with the bad governance story, personal costs are significantly lower in large firms than small firms (0.00% compared to 8.53%, difference has t-stat 7.1). The lower turnover costs explain why more successions are forced in large firms (18.6% vs 15.7%). The difference in turnover costs outweighs an effect going in the opposite direction, namely, there is more dispersion in ability across CEOs in small firms (σ_0 of 3.26% vs. 1.28%, difference has t-stat 35), which tends to push firing rates up in small firms. Ability dispersion is higher because the model needs to fit the higher dispersion in realized profitability across CEOs in small firms ($\sigma(E[X])$ increases from 10.3% to 19.1% in Panel B).

To summarize, three out of four split-sample tests (year, outsiders, size) are consistent with the bad-governance interpretation of personal costs. The fourth test (stock ownership) does not support the bad-governance view. One could always object that the proxies for governance quality are imperfect. For instance, it may be optimal for shareholders to elect an insider-dominated board if doing so allows shareholders to pay the CEO less or hire a more talented CEO. Such arguments seem like a stretch to this author. I conclude that, on balance, results support the bad-governance interpretation, although not all evidence supports this view. Another message from these tests is that a higher firing rate alone does not constitute evidence of lower CEO entrenchment, because the higher rate may be due instead to higher dispersion in CEO ability or faster learning about CEO ability. My approach teases apart these different drivers of the firing rate.

IV. Robustness

This section describes robustness exercises regarding firm fixed effects in profitability, a flat firing threshold, a different assumption about voluntary turnover costs, alternate discount rates, and a more aggressive classification of CEO successions into forced and voluntary. I also discuss how earnings manipulation relates to my results.

A. Firm Fixed Effects

The model attributes all intra-industry variation in average profitability to variation in CEO skill. If there are other reasons why profitability varies, for instance if some industry sectors are more profitable than others, then I over-estimate the variation in CEO skill, σ_0. Since higher values of σ_0 require higher turnover costs to fit turnover rates, my estimated turnover costs are also biased upwards.

To address this concern, I introduce firm fixed effects in profitability, which allows some firms to be more profitable than others for reasons unrelated to CEO ability. If the fixed effects are independent of CEO ability, then we can measure each firm's fixed effect by averaging profitability across multiple CEOs in the firm. After subtracting this average from the firm's yearly profitability, the remaining variation in average profitability across CEOs is due only to variation in CEO ability and not the firm fixed effects. Following this logic, I demean excess profitability y^*_{it} at the firm level using Compustat data from 1970-2006, and I estimate the model using the demeaned data.

Parameter estimates are in the "fixed effects" rows of Table 6. The table also repeats the main results from Table 2 to make comparisons easier. The χ^2 statistic indicates the model fits the data slightly worse when I introduce fixed effects. As expected, the estimated dispersion in ability across CEOs (σ_0) is lower with fixed effects (down from 2.42% to 1.61%). Interestingly, there is no significant change in turnover costs. The model still needs huge turnover costs to fit the low firing rates, even when we allow firm fixed effects in profitability.

INSERT TABLE 6 NEAR HERE

B. Dynamic Entrenchment, Board Risk Aversion, Learning on the Job, Fluctuating Skill, and the Slope of the Firing Threshold

The model assumes the board is risk neutral, and that turnover costs and a CEO's ability are both constant over time. The model predicts that uncertainty about the CEO's skill drops over time due to learning, so the CEO's option value declines with tenure, and hence the firing threshold rises with tenure. In other words, the board is more willing to fire CEOs the longer they have been in office, because the board prefers CEOs with more uncertain skill. Next I discuss four model extensions which could change this prediction.

First, the board's effective personal cost of CEO turnover, $c^{(pers)}/\kappa$, may increase with tenure as the CEO appoints more of his allies to the board or gains bargaining power. Rising personal costs will lower the slope of the firing threshold, since the board becomes less willing to fire the CEO as tenure increases. The model of Hermalin and Weisbach (1998) makes a similar prediction.

Second, if the board is risk averse then it will prefer CEOs with lower uncertainty. Since uncertainty is lower for CEOs who have been in office longer, adding risk aversion to the model will make the board less willing to fire long-tenured CEOs, i.e., will lower the firing threshold's slope. Intuitively, the board may prefer a mediocre CEO who is a known quantity compared to a new CEO whose skill is potentially better but more uncertain.

Third, if CEOs gain human capital from learning on the job, and if shareholders receive at least part of the surplus, then boards should be less willing to fire CEOs the longer they have spent in office. In other words, the firing threshold should rise less with tenure.

Fourth, CEOs' skill level may fluctuate randomly over time as their human capital gains or loses productivity, e.g., due to changing industry conditions. Random fluctuations in skill cause uncertainty to drop less with tenure, because old signals lose relevance. As a result, the gains in option value from replacing a long-tenured CEO with a new one are smaller, and the threshold's slope is lower. Dangl, Wu, and Zechner (2007) show that when skill fluctuates, the firing threshold can even be perfectly flat.

All four extensions suggest the firing threshold increases less with tenure than my main model predicts. Interestingly, the main model can fit the empirical tenure-firing relationship quite well (Figure 5), implying we do not need these extensions to explain the data. Nevertheless, to examine the extensions' effects on my results, I estimate the model forcing the firing threshold to be perfectly flat and equal to its value at tenure zero. A flat firing threshold is admittedly *ad hoc*. The four factors above may result in a threshold with a slightly positive slope, or even a negative slope. I impose a flat threshold and use the same learning dynamics as before to keep the robustness exercise simple.

Results are in the "flat threshold" rows of Table 6. The χ^2 statistic indicates the model with a flat threshold fits the data roughly as well as the main model. The main result does not change when I impose a flat threshold: the total turnover cost is huge (5.99%, up from 5.94%) and mainly reflects a personal cost (5.28%, up from 4.61%). Dispersion in ability (σ_0) is higher with a flat threshold (up from 2.42% to 2.94%). Flattening the threshold makes the firing region smaller and forced turnover less frequent. Since the data have not changed, the model compensates by increasing ability dispersion σ_0, which raises forced turnover rates (Figure 1).

C. Costless Voluntary Turnover

The model assumes forced and voluntary CEO turnover are equally costly. For robustness, I re-solve and re-estimate the model assuming voluntary turnover is costless to the board and to shareholders. The board's optimal firing policy changes, because the board now has an incentive not to fire CEOs who are close to retirement, but instead to wait until they retire at no cost. Overall, there is less of an incentive to fire the CEO when voluntary turnover is costless. The model compensates by raising ability dispersion (σ_0) from 2.42% to 3.34% (Table 6, rows "$c_{retire} = 0$"). The personal cost is now even higher, up from 4.61% to 5.94%. The χ^2 statistic indicates this version of the model fits the data slightly worse than the main model.

D. Alternate Discount Rates

Next, I estimate the model using different assumed values of β, the board's discount factor. The main results use $\beta = 0.9$, and in these robustness tests I use β=0.85 and 0.95. Estimates of model parameters for these two cases are in Table 6. The estimated total turnover costs are even higher using the new values of β. I find a negative relation between the assumed value of β and the estimate of σ_0, the dispersion in CEO skill. All else equal, raising β shifts the firing threshold up and hence makes boards more willing to fire the CEO; intuitively, the benefits of firing an unskilled CEO have higher net present value when β is higher. Since the underlying data do not change when we raise β, the model compensates by lowering the board's incentive to fire the CEO by lowering σ_0, the dispersion in skill.

E. Alternate Forced/Voluntary Classification

Kaplan and Minton (2006) suggest that many successions classified as voluntary in the data are in fact forced. To address this concern, I try a more aggressive classification. I assume any CEO who left office at age 64 or younger was forced out. For the 884 successions where CEO age is available, 69% (609) are forced according to this definition, compared to 17% using Parrino's classification. I then estimate the model using data from the new classification.

The model tries to match the higher rate of forced turnover by lowering the total costs of turnover from 5.94% to 1.34% (Table 6). Although smaller, the estimated personal turnover cost is still significantly positive. The χ^2 statistic increases from 33.2 to 398.6, indicating the model has a harder time fitting these new data. In particular, the model has trouble simultaneously matching the higher firing rate (9.0% per year in the real data, compared to 5.3% from the model) and the median tenure of a fired CEOs (6 years in the real data, 3 years from the model). These failures are not surprising, since the forced/voluntary classification used in this robustness exercise is crude compared to Parrino's classification, which I use everywhere else.

F. Earnings Management

The model assumes reported earnings equal true earnings. In reality, CEOs have incentives to manipulate earnings in at least three ways.

First, if a CEO believes he is close to being fired, then he may try to inflate reported earnings. However, Murphy and Zimmerman (1993) find no empirical evidence of such manipulation.

CEOs have incentives to take an earnings bath when they first enter office, in order to unravel any previous manipulation and boost future compensation and chances of staying in office. Weisbach (1995) and Murphy and Zimmerman (1993) report evidence of earnings baths after successions. Since my model will mistakenly attribute these earnings baths to firm turnover costs, my estimated firm turnover cost $c^{(firm)}$ may be biased upwards. Earnings baths may also explain why the model fails to explain the drop in profitability in the year after forced successions (Figure 5).

Finally, CEOs may engage in "signal jamming," injecting noise into earnings to make it harder for the board to learn the CEO's ability (e.g. Fudenberg and Tirole (1986), Hermalin and Weisbach (2007)). While signal jamming may help explain my estimates of profitability volatility and persistence– for instance, why I find volatility σ_ϵ is 3.4% instead of some lower number—, signal jamming does not imply any obvious bias in these estimates. Signal jamming may also help explain my finding that boards rely heavily on non-earnings signals when evaluating the CEO.

V. Conclusion

Previous empirical work has established that CEOs are rarely fired, and that profitability poorly predicts CEO dismissals. Attributing these stylized facts to bad governance would be premature, as the literature provides few quantitative benchmarks for how a rational, well governing board would behave. This study provides one such benchmark. I develop and solve a dynamic model which features a rational board, costly turnover, and learning about CEO ability. To gauge magnitudes and overcome endogeneity problems, this study takes a structural estimation approach. I estimate the model's fundamental parameters by applying the simulated method of moments to data on CEO turnover and firm profitability. I find three main results. First, to rationalize the observed rate of forced turnover, boards must behave as if firing the CEO costs shareholders 5.9% of the firm's assets, or $236 million for the median firm. Second, this cost mainly reflects CEO entrenchment and bad governance rather than a real cost for shareholders, although not all evidence supports this view. Third, to rationalize the weak relation between CEO dismissals and profitability, boards must rely very heavily on non-earnings signals of CEO ability. The model can fit several empirical patterns, including the overall rate of forced CEO turnover, the relation between turnover and tenure, the average changes in profitability and stock prices around CEO dismissals, and the forecasting relation between profitability and forced turnover. In almost all cases, the model matches these empirical patterns both in terms of direction and magnitude.

One interpretation of these results is that the turnover costs the model needs to fit the data are implausibly large, so the model must be wrong. According to this interpretation, my results present a quantitative CEO turnover puzzle. A second interpretation is that the parameter values are not implausible, and the model is a good description of reality. For instance, high CEO entrenchment is consistent with the CEOs' considerable influence on board selection during the period I study.

More work is needed to evaluate these two interpretations. While I have explored a few alternate models, it would be worthwhile to consider models with endogenous board composition, contracting and bargaining between the CEO and board, costly monitoring, and asymmetric information. Like my model, these alternate models should be judged on their ability to explain magnitudes and not just qualitative features of the data. Directly measuring directors' personal CEO turnover costs would help to evaluate whether my $183 million estimated cost is reasonable. Also, identifying boards' non-earnings signals of CEO ability, incorporating them into the model, and measuring their influence on CEO firing decisions is interesting avenue for future work.

REFERENCES

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

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

Bertrand, Marianne, and Antoinette Schoar, 2003, Managing with style: The effects of managers on firm policies, *Quarterly Journal of Economics* 118(4), 1169–1208.

Bond, Philip, Itay Goldstein, and Edward S. Prescott, 2008, Market-Based Corrective Actions, Working paper, University of Pennsylvania.

Brickley, James A., 2003, Empirical research on CEO turnover and firm-performance: A discussion, *Journal of Accounting and Economics* 36, 227–233.

Bushman, Robert M., Raffi J. Indjejikian, and Abbie Smith, 1996, CEO compensation: The role of individual performance evaluation, *Journal of Accounting and Economics* 21, 161–193.

Coughlan, Anne T., and Ronald M. Schmidt, 1985, Executive compensation, management turnover, and firm performance, *Journal of Accounting and Economics* 7, 43–66.

Dangl, Thomas, Youchang Wu, and Josef Zechner, 2006, Market discipline and internal governance in the mutual fund industry, Forthcoming, *Review of Financial Studies.*

DeAngelo, Harry and Linda DeAngelo, 1989, Proxy contests and the governance of publicly held corporations, *Journal of Financial Economics* 23(1), 29–60.

Demb, Ada and F. Friedrich Neubauer, 1992. *The Corporate Board* (Oxford University Press, Oxford).

Eisfeldt, Andrea L. and Adriano A. Rampini, 2008, Managerial incentives, capital reallocation, and the business cycle, *Journal of Financial Economics* 87, 177–199.

Engel, Ellen, Rachel M. Hayes, and Xue Wang, 2003, CEO turnover and properties of accounting information, *Journal of Accounting and Economics* 36, 197–226.

Fama, Eugene F., and Kenneth R. French, 2000, Forecasting profitability and earnings, *Journal of Business* 73, 161–175.

Fama, Eugene F., and Michael C. Jensen, 1983, Separation of ownership and control, *Journal of Law and Economics* 26, 301–325.

Fisman, Ray, Rakesh Khurana, and Matthew Rhodes-Kropf, 2005, Governance and CEO turnover: Do something or do the right thing?, Working paper, Columbia University.

Fudenberg, Drew and Jean Tirole, 1986, A 'Signal-Jamming' Theory of Predation, *Rand Journal of Economics* 17, 366–376.

Gabaix, Xavier, and Augustin Landier, 2008, Why Has CEO Pay Increased So Much?, Forthcoming, *Quarterly Journal of Economics.*

Hennessy, Christopher A., and Toni M. Whited, 2005, Debt dynamics, *Journal of Finance* 60(3), 1129–1165.

Hennessy, Christopher A., and Toni M. Whited, 2007, How costly is external financing? Evidence from a structural estimation, Working paper, University of California Berkeley.

Hermalin, Benjamin E., and Michael S. Weisbach, 2007, Transparency and corporate governance, Working paper, University of California Berkeley.

Hermalin, Benjamin E., and Michael S. Weisbach, 1998, Endogenously chosen boards of directors and their monitoring of the CEO, *American Economic Review* 88(1), 96–118.

Herschleifer, David, and Anjan V. Thakor, 1994, Managerial performance, boards of directors, and takeover bidding, *Journal of Corporate Finance* 1, 63–90.

Herschleifer, David, and Anjan V. Thakor, 1998, Corporate control through board dismissals and takeovers, *Journal of Economics and Management Strategy* 7(4), 489–520.

Huson, Mark R., Paul H. Malatesta, and Robert Parrino, 2004, Managerial succession and firm performance, *Journal of Financial Economics,* 237–275.

Huson, Mark R., Robert Parrino, and Laura T. Starks, 2001, Internal monitoring mechanisms and CEO turnover: A long-term perspective, *Journal of Finance* 56, 2265–2297.

Jenter, Dirk, and Fadi Kanaan, 2006, CEO turnover and relative performance evaluation, Working Paper, MIT.

Kaplan, Steven N. and Bernadette A. Minton, 2006, How has CEO turnover changed? Increasingly performance sensitive boards and increasingly uneasy CEOs, Working paper, University of Chicago.

Kim, Yungsan, 1996, Long-term firm performance and chief executive turnover: An empirical study of the dynamics, *Journal of Law, Economics, and Organization* 12(2), 480–496.

Lehn, Kenneth M., and Mengxin Zhao, 2006, CEO turnover after acquisitions: Are bad bidders fired?, *Journal of Finance* 61(4), 1759–1811.

Lorsch, Jay W., and Elizabeth MacIver, 1989. *Pawns or potentates: The reality of America's corporate boards* (Harvard Business School Press, Boston, MA.).

Mace, Myles L., 1971. *Directors: Myth and reality* (Harvard University Press, Boston, MA.).

Manso, Gustavo, 2007, Motivating innovation, Working paper, MIT Sloan School of Management.

McFadden, Daniel, 1989, A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* 57(5), 995–1026.

Michaelides, Alexander, and Serena Ng, 2000, Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators, *Journal of Econometrics* 96, 231–266.

Miller, Robert A., 1984, Job matching and occupational choice, *Journal of Political Economy* 92(6), 1086–1120.

Murphy, Kevin J., and Jerold L. Zimmerman, 1993, Financial performance surrounding CEO turnover, *Journal of Accounting and Economics* 16, 273–315.

Pakes, Ariel, and David Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica* 57(5), 1027–1057.

Parrino, Robert, 1997, CEO turnover and outside succession: A cross-sectional analysis, *Journal of Financial Economics* 46, 165–197.

Parrino, Robert, Richard W. Sias, and Laura T. Starks, 2003, Voting with their feet: Institutional ownership changes around forced CEO turnover, *Journal of Financial Economics* 68, 3–46.

Pástor, Ľuboš, Lucian A. Taylor, and Pietro Veronesi, 2007, Entrepreneurial learning, the IPO decision, and the post-IPO drop in firm profitability, Forthcoming, *Review of Financial Studies*.

Rust, John, 1994, Structural estimation of Markov decision processes," in R.F. Engle and D.L. McFadden, ed.: *Handbook of Econometrics, Volume IV* (Elsevier Science B.V.).

Shivdasani, A., and D. Yermack, 1999, CEO involvement in the selection of new board members: An empirical analysis, *Journal of Finance* 54, 1829–1854.

Warner, Jerold B., Ross L. Watts, and Karen H. Wruck, 1988, Stock prices and top management changes, *Journal of Financial Economics* 20, 461–492.

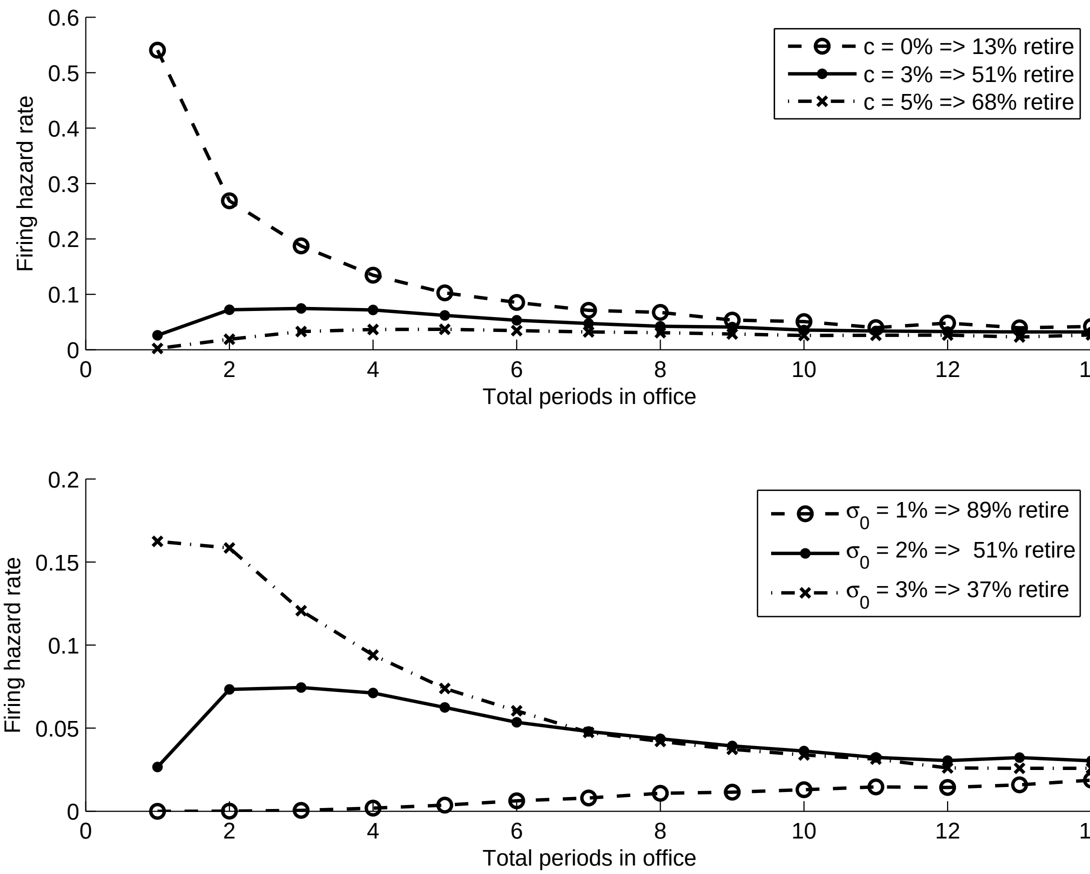
Weisbach, Michael S., 1995, CEO turnover and the firm's investment decisions, *Journal of Financial Economics* 37, 159–188.

Weisbach, Michael S., 1988, Outside directors and CEO turnover, *Journal of Financial Economics* 20, 431–460.

Wu, Joanna S., and Ivy X. Zhang, 2008, The adoption of international accounting standards and firm internal performance evaluation, Working paper, University of Rochester.

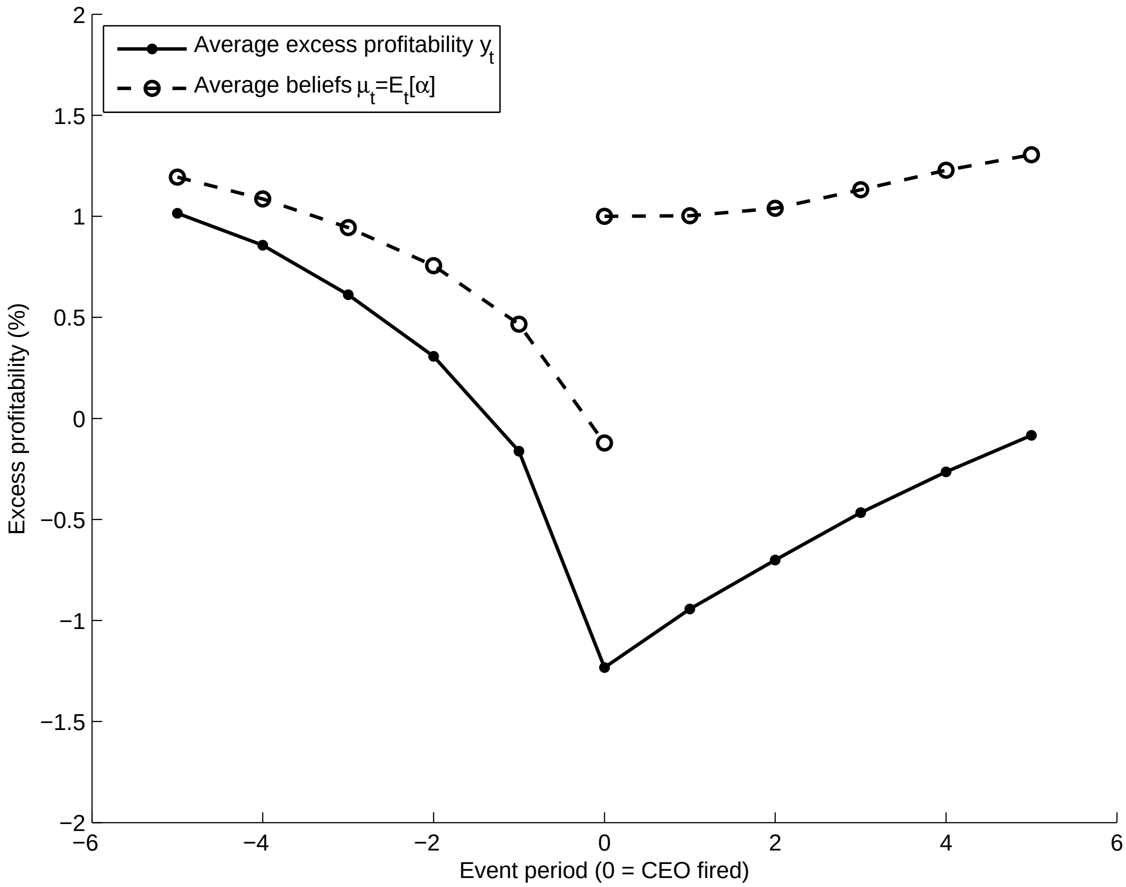
Yermack, David, 2006, Golden handshakes: Separation pay for retired and dismissed CEOs, *Journal of Accounting and Economics* 41, 237–256.

Figure 1: **Predicted CEO Dismissal Hazard Rates**



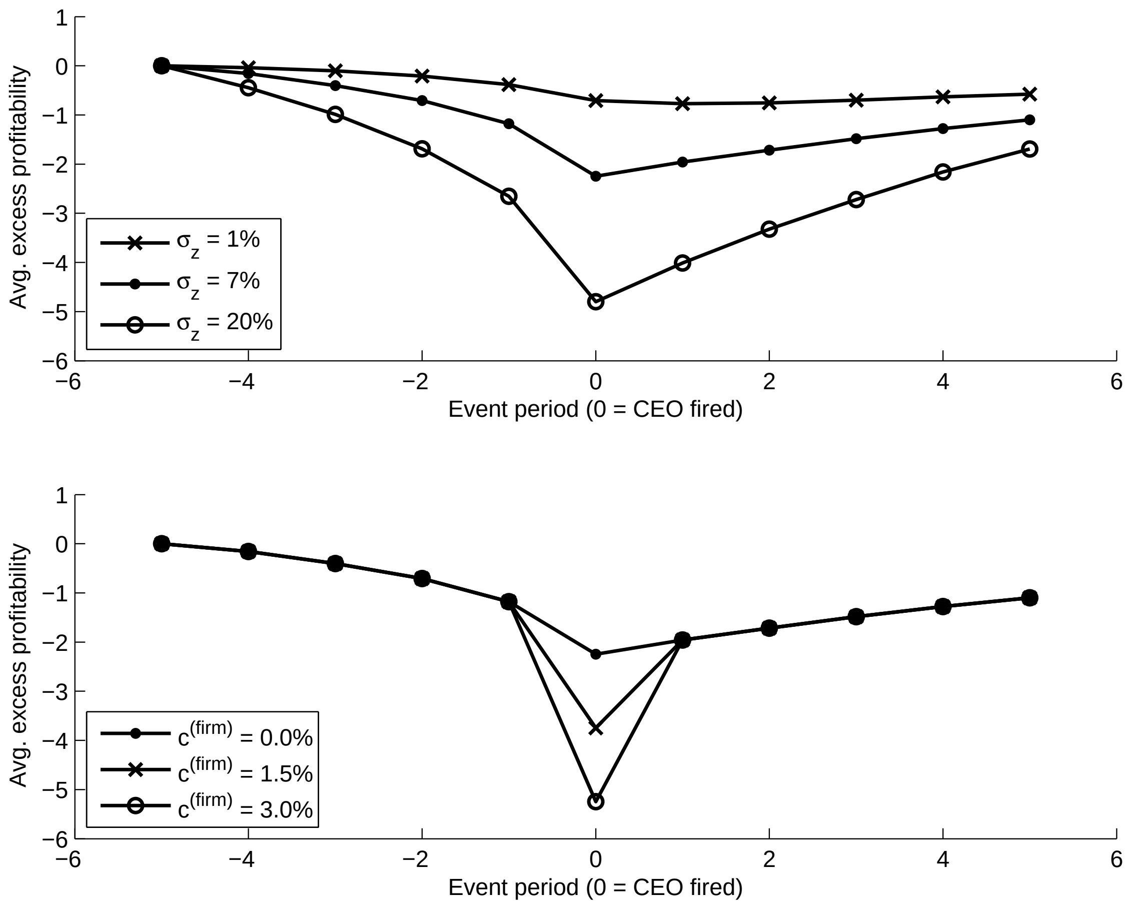
This figure shows the hazard rates for CEO dismissals at different tenure levels τ. The hazard rate for tenure τ is the probability that the CEO will be fired after his τth period in office, conditional on him surviving to period τ. The legends indicate the percent of CEOs who reach retirement after 15 periods without being fired. The top panel shows hazard rates for three different values of the total turnover cost, $c = c^{(firm)} + c^{(pers)}/\kappa$. The bottom panel shows hazard rates for three values of σ_0, the standard deviation of CEO ability α in the pool of replacement CEOs. These results are from simulations of the model using parameter values $\beta = 0.9$, $\mu_0 = 1\%$, $\sigma_0 = 2\%$, $\sigma_\epsilon = 3\%$, $c = 3\%$, $\phi = 0.12$, and $\sigma_z = 7\%$; voluntary turnover occurs after (and only after) completing 15 periods in office, so $f(\tau) = 0$ for $\tau = 0, 1, ..., 13$, $f(14) = 1$.

Figure 2: **Profitability around CEO Dismissals**



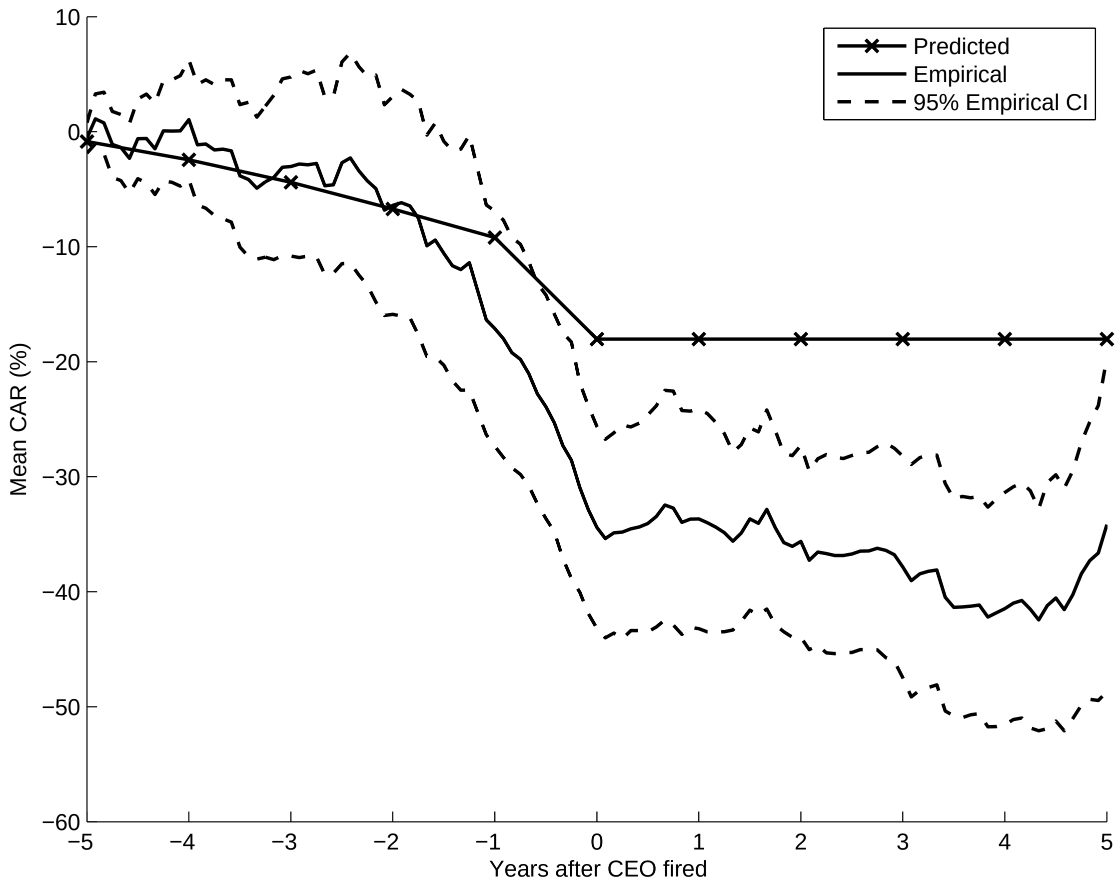
This figure plots average excess profitability in event time around CEO dismissals. Excess profitability y_t equals firm profitability Y_t minus industry profitability v_t. μ_t is the average across CEOs of the board's posterior mean beliefs about the CEO's ability α. Results are from 100,000 simulated CEO spells, selecting only those spells where the CEO was forced out of office at the end of period 0. Simulations use the following parameter values: $\beta = 0.9$, $\mu_0 = 1\%$, $\sigma_0 = 2\%$, $\sigma_\epsilon = 3\%$, $c = 3\%$, $\phi = 0.12$, and $\sigma_z = 7\%$; voluntary turnover occurs after (and only after) completing 15 periods in office, so $f(\tau) = 0$ for $\tau = 0, 1, ..., 13$, $f(14) = 1$. Also, there are no firm costs of turnover, so $c^{(firm)} = 0$ and hence $c = c^{(pers)}/\kappa$.

Figure 3: **Profitability Around CEO Dismissals: Comparative Statics**



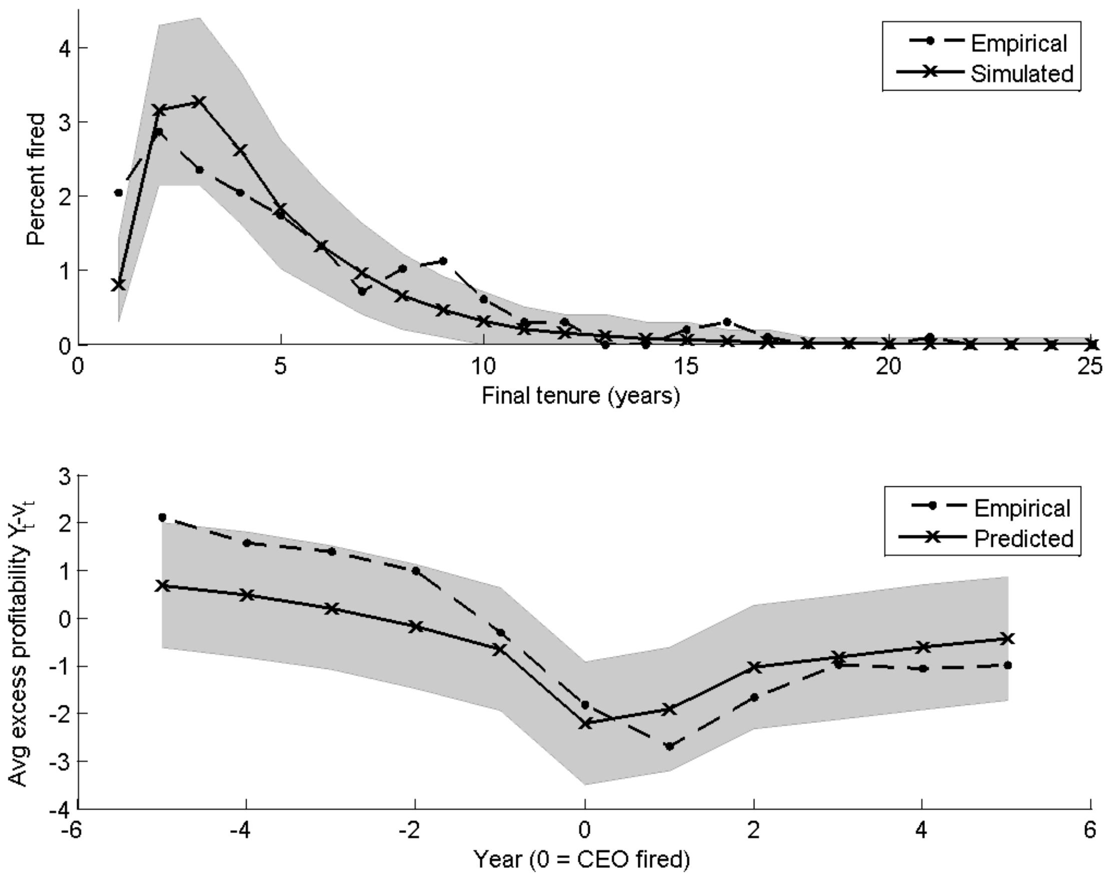
This figure shows average excess profitability in event time around CEO dismissals. Excess profitability y_t equals firm profitability Y_t minus industry profitability v_t. All results are from 100,000 simulated CEO spells, selecting only those spells where the CEO was forced out of office at the end of period 0. Unless otherwise noted in the legend, simulations use parameter values $\beta = 0.9$, $\mu_0 = 1\%$, $\sigma_0 = 2\%$, $\sigma_\epsilon = 3\%$, $c = 3\%$, $\phi = 0.12$, and $\sigma_z = 7\%$; voluntary turnover occurs after (and only after) completing 15 periods in office, so $f(\tau) = 0$ for $\tau = 0, 1, ..., 13$, $f(14) = 1$. In the top panel there are no firm costs of turnover, so $c^{(firm)} = 0$, and $c = c^{(pers)}/\kappa$. In the bottom panel, I hold constant total costs $c = c^{(firm)} + c^{(pers)}/\kappa$ constant, and I vary the firm cost $c^{(firm)}$. To allow easier comparisons, I shift all lines vertically so that they match perfectly at period -5, where I set excess profitability equal to 0%.

Figure 4: **Abnormal Stock Returns Around CEO Dismissals**



The figure shows the average cumulative abnormal stock return (CAR) in event time around CEO dismissals. The predicted returns are computed from 100,000 simulated CEO spells, as described in Appendix D. Simulations use the parameter estimates from Table 2. The empirical returns use data around the 168 forced successions from the sample defined in Section 2.A. Abnormal empirical returns equal the stock return minus the equal-weighted industry stock return. I use the 12-industry definition from Kenneth French's website. The dashed lines indicate a 95% confidence interval for the empirical mean CAR. I compute the confidence interval treating abnormal returns as statistically independent across time and firms.

Figure 5: **Comparison of Empirical and Predicted Patterns in CEO Dismissals and Profitability**



The top panel shows the unconditional percent of CEOs fired at different tenure levels. The bottom panel shows average excess profitability in event time around forced successions. In both panels, the empirical pattern is computed from the sample of 981 CEO spells described in Section 2.A. The predicted pattern is the average across 10,000 simulated samples, each of which contains 981 artificial CEOs. Simulations use parameter values from Table 2. The grey 95% confidence region covers the area between the 2.5% and 97.5% percentiles from the 10,000 simulated samples. Average excess profitability $Y_t - v_t$ equals firm profitability Y_t minus industry profitability v_t. In the empirical sample, profitability is annual ROA.

Table 1: **Summary Statistics**

Panel A contains summary statistics on CEO spells in various sub-samples. The full sample consists of CEO spells for firms in the 1970-2007 *Forbes* annual compensation surveys. I include complete CEO spells that ended between 1971-2006. Additional details are in Section 2.A. The 12 industries are defined on Kenneth French's website. "1970-1974" is the subset of full CEO spells ending in 1970-1974, and so on. Firm/years is the number of firm/year observations in the given subsample. Total spells is the number of CEO spells. "Percent forced" is the percent of CEO spells that ended in forced succession. "Pct. forced per year" is the percent of sample firm/years that ended in a CEO dismissal. Panel B contains additional statistics for the full sample. ROA is the percent firm-level annual return on assets. $y_t^* = ROA - v_t$ equals firm ROA minus average industry ROA. Assets is Compustat item 6. Spell length is the number of years the CEO completed in office before leaving his position. Statistics for ROA, v_t, and assets are computed pooling all firm/years, and statistics for spell length are computed across CEO spells.

Panel A: CEO Spells

Subsample	Firm/years	Total spells	Forced successions	Unforced successions	Percent forced	Pct. forced per year
Full sample	7,325	981	168	813	17.1	2.29
Consumer nondurables	757	87	13	74	14.9	1.72
Consumer durables	325	42	8	34	19.0	2.46
Manufacturing	1,543	204	28	176	13.7	1.81
Energy	315	40	7	33	17.5	2.22
Chemicals	436	63	3	60	4.8	0.69
Business equipment	605	90	23	67	25.6	3.80
Telecom	153	23	5	18	21.7	3.27
Utilities	662	82	7	75	8.5	1.06
Wholesale and retail	517	88	19	69	21.6	3.68
Health	481	57	8	49	14.0	1.66
Finance	920	127	27	100	21.3	2.93
Other	611	78	20	58	25.6	3.27
1970-1974	580	102	8	94	7.8	1.38
1975-1979	886	132	15	117	11.4	1.69
1980-1984	1,182	146	21	125	14.4	1.78
1985-1989	1,340	163	23	140	14.1	1.72
1990-1994	1,113	126	23	103	18.3	2.07
1995-1999	713	90	20	70	22.2	2.81
2000-2004	989	149	40	109	26.8	4.04
2005-2006	522	73	18	55	24.7	3.45

Panel B: Additional Statistics

Variable	Observations	Mean	Std. dev.	Median	Min	Max
ROA	7325	16.0	9.07	15.5	-23.8	85.6
y_t^* =ROA-v_t	7325	2.00	7.37	0.75	-35.7	68.0
Assets ($billion)	7325	12.5	55.0	2.38	0.015	1264
Spell length:						
All	981	7.5	4.9	6	1	29
Forced	168	5.1	3.8	4	1	21
Unforced	813	8.0	4.9	7	1	29

Table 2: **Parameter Estimates**

This table contains estimates of the parameters from the model in Section 1. Estimation uses data on a sample of 981 CEOs, described in Section 2.A. Parameters are estimated using the simulated method of moments, as described in Section 3. $c^{(firm)}$ is the CEO turnover cost to the firm, and $c^{(pers)}$ is the personal CEO turnover cost to the board. κ controls the degree to which the board internalizes shareholder value. μ_0 and σ_0 are the mean and standard deviation, respectively, of boards' prior beliefs about a newly hired CEO's ability. ϕ controls the persistence in profitability. σ_ϵ is the conditional time-series volatility of profitability. σ_z is the standard deviation of the board's additional signal about CEO ability. All parameters are in units of percent of assets per year, except ϕ (unitless), and the costs c (percent of assets). Standard errors are in parentheses.

Firm cost $c^{(firm)}$	Personal cost $c^{(pers)}/\kappa$	Prior mean μ_0	Prior stdev. σ_0	Persis-tence ϕ	Profit. stdev. σ_ϵ	z signal stdev. σ_z
1.33	4.61	0.88	2.42	0.125	3.43	5.15
(0.61)	(0.58)	(0.34)	(0.06)	(0.004)	(0.09)	(0.33)

Table 3: **Moments used in SMM Estimation**

This table shows the 14 moments used in the SMM estimation described in Section 2. Empirical moments are computed from the sample of 981 CEOs described in Section 2.A. Simulated moments are computed from data simulated from the model using parameter values in Table 2. Moments' standard errors and p-values are computed by Monte Carlo, as follows. I create 10,000 sets of 14 moments, each from a simulation of 981 CEOs (to match the empirical sample size) using parameter values in Table 2. The standard error is the standard deviation of the 10,000 simulated moments, and the p-value is fraction of simulated moments which are as or more extreme than the empirical moment. The profitability regression is $y^*_{it} = \lambda_0 + \lambda_1 y^*_{i,t-1} + \Delta^{(-2)} + \Delta^{(-1)} + \Delta^{(0)} + \Delta^{(1)} + \Delta^{(2)} + \delta_{it},$, estimated pooling all years and firms. $y^*_{i,t}$ is firm i's annual percent ROA in year t minus industry average profitability. $\Delta^{(k)}$ is a fixed effect for whether forced CEO turnover occurred at the end of period $t-k$. $E[Var(X)]$ is the mean across CEO spells of the within-spell variance of X_{it}, which is persistence-adjusted firm profitability. $Var(E[X])$ is the variance across CEO spells of within-spell average X_{it}. The hazard rates $h^{(j)}$ equal the percent of CEOs forced out of office per year during tenure period (j), conditional on the CEO reaching (j). The bottom panel shows the χ^2 statistic and corresponding p-value for SMM's test of over-identifying restrictions, which jointly tests whether the empirical and simulated moments are equal. This test is defined in Appendix D.

	Empirical moment	Simulated moment	Standard error	p-value
Moments 1-8 (Profitability regression)				
λ_0	0.22	0.20	0.04	0.66
λ_1	0.87	0.88	0.01	0.38
$\Delta^{(-2)}$	-0.37	-0.62	0.27	0.29
$\Delta^{(-1)}$	-1.36	-1.06	0.27	0.14
$\Delta^{(0)}$	-1.73	-1.99	0.26	0.20
$\Delta^{(1)}$	-1.15	-0.02	0.27	0.00
$\Delta^{(2)}$	0.60	0.42	0.27	0.38
$Var(\delta)$	11.99	11.92	0.20	0.60
Moments 9-12 (Forced turnover hazard rates)				
$h^{(1-2)}$	2.52	1.76	0.29	0.00
$h^{(3-4)}$	2.66	3.46	0.46	0.03
$h^{(5-7)}$	2.10	2.71	0.39	0.10
$h^{(8+)}$	1.96	1.47	0.26	0.04
Moments 13-14 (Other 2nd moments for profitability)				
$E[Var(X)]$	799.3	782.1	73.5	0.77
$Var[E(X)]$	221.5	183.5	20.9	0.11
Test of over-identifying restrictions				
$\chi^2 =$	33.2		p-value=	0.000

Table 4: **Additional Statistics on Model Fit**

This table compares statistics computed from the empirical sample (containing 981 CEOs, described in Section 2.A) and a sample of 100,000 CEOs simulated from the model using parameter estimates in Table 2. "Median tenure" is the number of years the CEO completed in office before leaving, computed separately for CEOs whose successions were forced and unforced. "Percent forced" is the percent of CEOs who were forced out of office. "Percent forced per year" is the percent of firm/years which end in forced CEO turnover. In the probit model, the dependent variable is an indicator for whether year t ended in a CEO dismissal, and the independent variable is y_t^*, the firm's annual ROA minus the industry's ROA in year t. "Slope" is the estimated slope on y_t^*, and "Stderr." is the associated robust standard error.

	Median tenure		Percent	Percent forced	Probit model		
	Unforced	Forced	forced	per year	Slope	Stderr.	Pseudo-R^2
Empirical	7	4	17.1	2.29	-0.168	[0.026]	0.03
Simulated	7	4	16.2	2.16	-0.125	[0.002]	0.02

Table 5: **Estimation in Subsamples**

Each pair of rows contains estimation results using a subset of the 981 CEOs from the full empirical sample. Subsample 1971-1989 and 1990-2006 contain all CEO spells that ended between the given years. "Low (high) ownership" firms are those where the median percent of shares owned by officers and directors, not including the CEO, is below (above) 1.31%. "Less (more) outsiders" are firms where the median percentage of directors who are not also officers in the firm is below (above) 72.7%. "Small (large) firms" have assets below (above) \$6.6B, in CPI-adjusted 2007 dollars. Panel A contains parameter estimates with standard errors in parentheses, and also the number of CEO spells used and the χ^2 test of over-identifying restrictions with associated p-value. See Table 2 for brief definitions of parameters, and see model in Section 1 for full definitions. Panel B compares statistics computed from the empirical subsample and a sample of 100,000 CEOs simulated from the model using corresponding parameters from Panel A. Table 4 defines the statistics in Panel B. $\sigma(E[X])$ is the standard deviation across CEOs of each CEO's average persistence-adjusted profitability X_{it} during his time in office, in units percent per year.

Panel A: Parameter Estimates in Subsamples

Subsample	Firm cost $c^{(firm)}$	Personal cost $c^{(pers)}/\kappa$	Prior mean μ_0	Prior stdev. σ_0	Persis-tence ϕ	Profit. stdev. σ_ϵ	z signal stdev. σ_z	CEOs	χ^2 (p-val.)
1971-1989	0.05	8.32	0.49	2.52	0.125	3.30	4.37	543	21.3
	(0.91)	(1.02)	(0.44)	(0.05)	(0.003)	(0.10)	(0.20)		(0.00)
1990-2006	1.67	2.28	1.24	2.72	0.123	3.62	9.51	438	27.5
	(0.70)	(0.76)	(0.67)	(0.07)	(0.002)	(0.14)	(0.46)		(0.00)
Low ownership	0.04	6.41	-0.76	3.10	0.126	3.10	12.92	222	16.1
	(1.07)	(1.09)	(0.60)	(0.11)	(0.003)	(0.11)	(0.93)		(0.02)
High ownership	0.00	7.99	0.14	3.99	0.088	3.35	8.62	222	8.6
	(1.19)	(1.54)	(0.94)	(0.18)	(0.004)	(0.15)	(1.03)		(0.29)
Few outsiders	0.00	8.25	1.57	2.93	0.117	3.69	5.84	327	9.7
	(1.26)	(1.69)	(0.67)	(0.11)	(0.004)	(0.18)	(0.94)		(0.21)
More outsiders	0.00	3.00	-0.56	1.98	0.116	2.87	10.80	325	14.7
	(0.85)	(0.99)	(0.47)	(0.07)	(0.002)	(0.10)	(0.65)		(0.04)
Small firms	0.01	8.53	1.38	3.26	0.120	4.01	5.34	491	18.7
	(1.05)	(1.06)	(0.69)	(0.05)	(0.004)	(0.14)	(0.24)		(0.01)
Large firms	1.51	0.00	0.32	1.28	0.133	2.91	7.59	490	31.1
	(0.59)	(0.58)	(0.38)	(0.03)	(0.002)	(0.09)	(0.26)		(0.00)

Panel B: Model Fit in Subsamples

Subsample		Median tenure		Percent	Percent		Probit model		
		Unforced	Forced	forced	forced/yr	$\sigma(E[X])$	Slope	Stderr.	Pseudo-R^2
1971-1989	Data	7	5	12.3	1.68	14.6	-0.121	0.031	0.03
	Model	7	4	10.2	1.30	12.7	-0.072	0.002	0.02
1990-2006	Data	7	4	23.1	3.03	16.1	-0.242	0.045	0.04
	Model	7	4	23.2	3.27	15.4	-0.225	0.002	0.04
Low ownership	Data	8	5	12.6	1.51	12.3	-0.161	0.040	0.06
	Model	7	5	12.1	1.54	12.1	-0.161	0.002	0.08
High ownership	Data	8	4	17.6	2.14	24.6	-0.121	0.031	0.03
	Model	7	4	16.6	2.22	19.8	-0.072	0.002	0.02
Less outsiders	Data	7	4.5	11.6	1.58	17.9	-0.080	0.038	0.02
	Model	7	4	12.0	1.54	14.8	-0.074	0.002	0.02
More outsiders	Data	8	5	16.0	2.05	10.4	-0.200	0.043	0.05
	Model	7	5	13.7	1.77	10.7	-0.171	0.002	0.05
Small firms	Data	7	4	15.7	2.14	19.1	-0.143	0.029	0.04
	Model	7	3	16.6	2.23	17.1	-0.089	0.002	0.01
Large firms	Data	7	4	18.6	2.45	10.3	-0.207	0.050	0.03
	Model	7	5	19.2	2.60	10.0	-0.249	0.003	0.04

Table 6: **Alternate Specifications**

Each pair of rows contains estimation results (parameter estimates, standard errors in parentheses, number of CEO spells used, and statistics for the test of over-identifying restrictions) from a different specification of the model. "Main results" are the same as in Tables 2 and 3. "Fixed effects" uses profitability data de-meaned at the firm level (Section 4.A). "Flat threshold" forces the firing threshold to stay constant with tenure (Section 4.B). "$c_{(retire)} = 0$" sets the turnover costs of voluntary succession to zero (Section 4.C). The next rows use different values for β, the board's discount factor (Section 4.D). "Alt. forced def'n" assumes CEO successions are forced if and only if the CEO is aged 64 or less at the time of succession (Section 4.E).

Specification	Firm cost $c^{(firm)}$	Personal cost $c^{(pers)}/\kappa$	Prior mean μ_0	Prior stdev. σ_0	Persis-tence ϕ	Profit. stdev. σ_ϵ	z signal stdev. σ_z	CEOs	χ^2 (p-val.)
Main results	1.33	4.61	0.88	2.42	0.12	3.43	5.15	981	33.22
	(0.61)	(0.58)	(0.34)	(0.06)	(0.00)	(0.09)	(0.33)		(0.00)
Fixed effects	1.00	5.13	-0.10	1.61	0.26	3.31	2.90	981	39.1
	(0.53)	(0.55)	(0.19)	(0.03)	(0.00)	(0.08)	(0.15)		(0.00)
Flat threshold	0.71	5.28	0.87	2.94	0.13	3.32	6.74	981	33.1
	(0.67)	(0.52)	(0.34)	(0.05)	(0.00)	(0.11)	(0.47)		(0.00)
$c_{(retire)}$=0	0.01	5.94	0.67	3.34	0.13	3.42	9.71	981	36.0
	(0.57)	(0.53)	(0.32)	(0.04)	(0.00)	(0.08)	(0.35)		(0.00)
$\beta = 0.85$	0.00	7.01	0.61	3.57	0.12	3.26	7.66	981	21.6
	(0.68)	(0.63)	(0.40)	(0.06)	(0.00)	(0.10)	(0.65)		(0.00)
$\beta = 0.95$	0.00	6.23	0.73	1.91	0.11	3.31	6.52	981	21.9
	(0.60)	(0.56)	(0.41)	(0.05)	(0.00)	(0.10)	(0.42)		(0.003)
Alt. fired def'n	0.00	1.34	1.46	1.24	0.12	3.46	6.08	884	398.6
	(0.41)	(0.34)	(0.38)	(0.03)	(0.00)	(0.08)	(0.33)		(0.00)

Technical Appendix for "Why Are CEOs Rarely Fired? Evidence from Structural Estimation"

[This Appendix is intended to exist as a separate document on the author's website.]

August 5, 2008

Appendix A: The Board's Learning Problem

This Appendix derives the rule boards use to learn about CEO skill. I start with firm-specific profitability:

$$y_t = y_{t-1} + \phi\left(\alpha - y_{t-1}\right) + \epsilon_t.$$

Rearranging,

$$\begin{aligned}
\frac{y_t - y_{t-1}}{\phi} &= \alpha - y_{t-1} + \frac{1}{\phi}\epsilon_t \\
X_t &\equiv \frac{y_t - y_{t-1}}{\phi} + y_{t-1} = \alpha + \frac{1}{\phi}\epsilon_t.
\end{aligned}$$

We then have

$$\left(\begin{array}{c} X_t \\ z_t \end{array}\right) |\alpha \sim \mathcal{N}\left(\left[\begin{array}{c} \alpha \\ \alpha \end{array}\right], \left[\begin{array}{cc} \sigma_\epsilon^2/\phi^2 & 0 \\ 0 & \sigma_z^2 \end{array}\right]\right).$$

Using the notation $\kappa_\epsilon \equiv \sigma_\epsilon^2/\left(\phi^2\sigma_0^2\right)$ and $\kappa_z \equiv \sigma_z^2/\sigma_0^2$, standard results on Bayesian learning (e.g. Zellner 1971) imply,

$$\mu_{t+1} = \left(\mu_t\sigma^{-2}\left(\tau\right) + \frac{\phi^2}{\sigma_\epsilon^2}X_t + \frac{1}{\sigma_z^2}z_t\right)\left(\sigma^{-2}\left(\tau\right) + \phi^2/\sigma_\epsilon^2 + 1/\sigma_z^2\right)^{-1}$$

where $\sigma^2\left(\tau\right)$ is prior variance at the beginning of the period, which decays monotonically over time according to

$$\sigma^2\left(\tau\right) = \sigma_0^2\left[1 + \tau\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right]^{-1}.$$

Then we have

$$\begin{aligned}
\mu_{t+1} &= \left(\mu_t + \frac{\sigma^2\left(\tau_t\right)\phi^2}{\sigma_\epsilon^2}X_t + \frac{\sigma^2\left(\tau_t\right)}{\sigma_z^2}z_t\right)\left(1 + \frac{\sigma^2\left(\tau_t\right)\phi^2}{\sigma_\epsilon^2} + \frac{\sigma^2\left(\tau_t\right)}{\sigma_z^2}\right)^{-1} \\
&= \mu_t + \delta_{y,t}\theta_y\left(\tau_t\right) + \delta_{z,t}\theta_z\left(\tau_t\right) \\
\delta_{z,t} &\equiv z_t - \mu_t \\
\delta_{y,t} &\equiv \frac{1}{\phi}\left(y_t - y_{t-1}\right) + y_{t-1} - \mu_t = \alpha + \frac{1}{\phi}\epsilon_t - \mu_t \\
\theta_y\left(\tau\right) &\equiv \frac{\sigma^2\left(\tau\right)\phi^2}{\sigma_\epsilon^2}\left(1 + \sigma^2\left(\tau\right)\phi^2/\sigma_\epsilon^2 + \sigma^2\left(\tau\right)/\sigma_z^2\right)^{-1} \\
&= \kappa_\epsilon^{-1}\left(1 + \left(\tau + 1\right)\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right)^{-1} \\
\theta_z\left(\tau\right) &= \kappa_z^{-1}\left(1 + \left(\tau + 1\right)\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right).^{-1}
\end{aligned}$$

Appendix B: Proof of Proposition 1 (Bellman equation)

This Appendix derives the board's Bellman equation, which characterizes its optimal firing decisions. I distinguish between total turnover costs from forced turnover (c_{fire}) and total turnover costs from voluntary turnover (c_{retire}). In my main model results and estimation, I set $c_{fire} = c_{retire} = c$. In the robustness section, I allow $c_{fire} \neq c_{retire}$, so separating the two here is useful. Substituting equation (16) into into (4), and then substituting the result into (3), the board's optimization problem is

$$\max_{\{d_{t+s}\}_{s=0}^{\infty}} U_t = \max_{\{d_{t+s}\}_{s=0}^{\infty}} \kappa E_t \left[\sum_{s=0}^{\infty} \beta^s B_{t+s} \left(v_{t+s} + y_{t+s} - d_{t+s} c_{fire} - b_{t+s} c_{retire} \right) \right],$$

where d_t and b_t are indicator variables equal to 1 if the CEO is fired or retired, respectively, in period t. Since the firm pays out profits immediately as dividends, the firm's book value is constant over time, so $B_{t+s} = B_t$ and

$$\begin{aligned}
\max_{\{d_{t+s}\}_{s=0}^{\infty}} \frac{U_t}{\kappa B_t} &= \max_{\{d_{t+s}\}_{s=0}^{\infty}} E_t \left[\sum_{s=0}^{\infty} \beta^s \left(v_{t+s} + y_{t+s} - d_{t+s} c_{fire} - b_{t+s} c_{retire} \right) \right] \\
&= E_t \left[\sum_{s=0}^{\infty} \beta^s v_{t+s} \right] + VF_t,
\end{aligned}$$

$$VF_t = \max_{\{d_{t+s}\}_{s=0}^{\infty}} E_t \left[\sum_{s=0}^{\infty} \beta^s \left(y_{t+s} - d_{t+s} c_{fire} - b_{t+s} c_{retire} \right) \right].$$

Next I write y_{t+s} as a function of y_{t-1}, shocks, and future posterior means:

$$\begin{aligned}
y_t &= y_{t-1} (1-\phi) + \phi \mu_t + \phi \delta_{X,t} \\
y_{t+1} &= \left[y_{t-1} (1-\phi) + \phi \mu_t + \phi \delta_{X,t} \right] (1-\phi) + \phi \mu_{t+1} + \phi \delta_{X,t+1} \\
&\vdots \\
y_{t+s} &= y_{t-1} (1-\phi)^{s+1} + \phi \sum_{\tau=0}^{s} \mu_{t+\tau} (1-\phi)^{s-\tau} + \phi \sum_{\tau=0}^{s} \delta_{X,t+\tau} (1-\phi)^{s-\tau} \\
E_t \left[y_{t+s} \right] &= y_{t-1} (1-\phi)^{s+1} + E_t \left[\phi \sum_{\tau=0}^{s} \mu_{t+\tau} (1-\phi)^{s-\tau} \right],
\end{aligned}$$

since $E_t \left[\delta_{Xt+\tau} \right] = E_t \left[E_{t+\tau} \left[\delta_{Xt+\tau} \right] \right]$ and $E_{t+\tau} \left[\delta_{Xt+\tau} \right] = 0$. Next, we have

$$\begin{aligned}
E_t \left[\sum_{s=0}^{\infty} \beta^s y_{t+s} \right] &= \sum_{s=0}^{\infty} \beta^s E_t \left[y_{t+s} \right] \\
&= \sum_{s=0}^{\infty} \beta^s \left[y_{t-1} (1-\phi)^{s+1} + E_t \left[\phi \sum_{\tau=0}^{s} \mu_{t+\tau} (1-\phi)^{s-\tau} \right] \right] \\
&= y_{t-1} (1-\phi) \sum_{s=0}^{\infty} \beta^s (1-\phi)^s + \phi \sum_{s=0}^{\infty} \sum_{\tau=0}^{s} \beta^s (1-\phi)^{s-\tau} E_t \left[\mu_{t+\tau} \right] \\
&= \left(\frac{1-\phi}{1-\beta(1-\phi)} \right) y_{t-1} + \left(\frac{\phi}{1-\beta(1-\phi)} \right) \sum_{s=0}^{\infty} \beta^s E_t \left[\mu_{t+s} \right].
\end{aligned}$$

Plugging this into the expression for VF,

$$VF_t \equiv \left(\frac{1-\phi}{1-\beta(1-\phi)} \right) y_{t-1} + V_t^*,$$

$$V_t^* = \max_{d_t} \left\{ \left(\frac{\phi}{1-\beta(1-\phi)} \right) \mu_t - d_t c_{fire} - b_t c_{retire} + \beta E_t \left[V_{t+1}^* \right] \right\},$$

so

$$V\left(\mu_t^{inc},\tau_t,b_t\right)=\max_{d_t}\left\{\left(\frac{\phi}{1-\beta\left(1-\phi\right)}\right)\mu_t-d_tc_{fire}-b_tc_{retire}+\beta E_t\left[V\left(\mu_{t+1}^{inc},\tau_{t+1},b_{t+1}\right)\right]\right\}.$$

If the incumbent CEO has just retired, the firm hires a new CEO and pays the retirement cost:

$$V_{retire}=V\left(\mu_t^{inc},\tau_t,1\right)=V\left(\mu_0,0,0\right)-c_{retire}.$$

Otherwise, if $b_t=0$ and $d_t=1$ (the firm fires its CEO), then the firm hires a new CEO and pays the firing cost:

$$V_{fire}\left(\mu_t^{inc},\tau_t,0\right)=V\left(\mu_0,0,0\right)-c_{fire}.$$

If $b_t=0$ and $d_t=0$ (the firm keeps its CEO), then

$$\begin{aligned}V_{keep}\left(\mu_t^{inc},\tau_t,0\right)&=\left(\frac{\phi}{1-\beta\left(1-\phi\right)}\right)\mu_t^{inc}+\beta E_t\left[V\left(\mu_{t+1}^{inc},\tau_{t+1},b_{t+1}\right)\right]\\&=\left(\frac{\phi}{1-\beta\left(1-\phi\right)}\right)\mu_t^{inc}+\beta f\left(\tau_t\right)V^{retire}+\beta\left(1-f\left(\tau_t\right)\right)E_t\left[V\left(\mu_{t+1}^{inc},\tau_{t+1},0\right)\right].\end{aligned}$$

The firm chooses d_t (fire or keep CEO) according to

$$V\left(\mu_t^{inc},\tau_t,0\right)=\max\left\{V^{fire}\left(\mu_t^{inc},\tau_t,0\right),V^{keep}\left(\mu_t^{inc},\tau_t,0\right)\right\}.$$

Recalling from equation (10) that

$$\mu_{t+1}^{inc}=\mu_t^{inc}+\theta_X\left(\tau_t\right)\delta_{X,t}+\theta_z\left(\tau_t\right)\delta_{z,t},$$

I write the Bellman in its final form by dropping time and incumbent subscripts and substituting in for V^{retire}.

Appendix C: Numerical Solution of Dynamic Program

This Appendix describes how I numerically solve for the board's optimal CEO firing rule. I approximate the value function using the Jacobi Iteration method. I start by discretizing the state space. State variable τ_t takes values in set $\varsigma=\{0,1,...,\overline{\tau}-1\}$, where $\overline{\tau}=\sup\tau$ is the maximum possible number of terms in office. I let μ takes values in finite set M, which contains 1,001 equally spaced points in the interval $[\mu_0-c_{fire}-2\sigma_0,\mu_0+c_{fire}+2\sigma_0]$; the length of the interval does not need to be extremely large, because the extrapolation used below ends up being quite accurate. To speed up the iteration, I start with a guess of V^0 over the grid $\varsigma\times M$:

$$V^0(\mu,\tau,0)=\left(\frac{\phi}{1-\beta\left(1-\phi\right)}\right)\left[\frac{\mu_0}{(1-\beta)}+\max(\mu-\mu_0,0)\frac{1-\beta^{\overline{\tau}-\tau}}{1-\beta}\right].$$

Then I update the value function according to

$$\begin{aligned}V^{t+1}\left(\mu,\tau,0\right)&=\max\{V^t\left(\mu_0,0,0\right)-c_{fire},\left(\frac{\phi}{1-\beta\left(1-\phi\right)}\right)\mu+\beta f\left(\tau\right)\left[V^t\left(\mu_0,0,0\right)-c_{retire}\right]+\\&\qquad\beta\left(1-f\left(\tau\right)\right)E\left[V^t\left(\mu+\theta_X\left(\tau\right)\delta_X+\theta_z\left(\tau\right)\delta_z,\tau+1,0\right)\right]\}.\end{aligned}$$

$$\left(\begin{array}{c}\delta_X\\ \delta_z\end{array}\right)\sim N\left(\left[\begin{array}{c}0\\0\end{array}\right],\left[\begin{array}{cc}\sigma_\epsilon^2/\phi^2+\sigma^2\left(\tau\right)&0\\0&\sigma_z^2+\sigma^2\left(\tau\right)\end{array}\right]\right)$$

I approximate the expectation above using Gauss-Hermite quadrature, as follows. Recall $V^t\left(\mu,\tau\right)$ is defined only for μ in the finite set M. First, I create a function $\widehat{V}^t\left(\mu,\tau\right)$ which is defined for all $\mu\in\mathbb{R}$ by performing piecewise cubic spline interpolation and extrapolation of the function $V^t(\mu,\tau)$. Second, I apply two-dimensional Gauss-Hermite quadrature with 7 nodes as follows: For each $\mu \in M$ and $\tau = 0, 1, ..., \overline{\tau}-1$,

$$\begin{aligned} & E\left[V^t\left(\mu+\theta_X(\tau)\,\delta_X+\theta_z(\tau)\,\delta_z,\tau+1,0\right)\right] \\ \approx\quad & \pi^{-1}\sum_{i=1}^{7}\sum_{j=1}^{7}\omega_i\omega_j\widehat{V}^t\left(\mu+\theta_X(\tau)\left[\sqrt{2\left(\sigma_\epsilon^2/\phi^2+\sigma^2(\tau)\right)}x_i\right]+\theta_z(\tau)\left[\sqrt{2\left(\sigma_z^2+\sigma^2(\tau)\right)}x_j\right],\tau+1,0\right) \end{aligned}$$

where $\{x_i\}$ and $\{\omega_i\}$ are the Gauss-Hermite quadrature nodes and weights, respectively.

I stop iterating as soon as

$$\max_{(\tau,\mu)\in\varsigma\times M}\left|V^{t+s}-V^t\right|<10^{-5}.$$

Appendix D: Computing Stock Prices

This Appendix describes how I compute stock prices by simulating the model. Since profits are realized at the end of periods and turnover costs are realized at the beginning of periods, the net dividend paid at the end of period t (as a fraction of assets) equals

$$\frac{D_t}{B_t}=v_t+y_t-c_{t+1}^{(firm)}. \tag{1}$$

I assume industry profits v_t are constant over time at 14% per period, which is the average annual industry ROA from the empirical sample defined in Section 3. The ex-dividend market-to-book ratio at the end of period t equals

$$\frac{M_t}{B_t}=\beta\left[\frac{U_{t+1}}{B_{t+1}}+c_{t+1}^{(firm)}\right]. \tag{2}$$

Costs $c_{t+1}^{(firm)}$ must be added back to U_{t+1}/B_{t+1} to avoid double-counting. I compute U_t using equation (13) from Proposition 1. Since assets B_t are constant over time, the return during period t equals

$$R_t=\frac{D_t/B_t+M_t/B_t}{M_{t-1}/B_{t-1}}. \tag{3}$$

I simulate the model using parameter estimates from Table 2.

The average simulated stock return is 11% per period, consistent with a discount rate $\beta = 0.9$.[1] The standard deviation of stock returns is 13% per year. Return volatility is low in part because I have assumed there is no industry volatility. I compute abnormal returns by subtracting the 11% average return from simulated returns.

Appendix E: Additional Details on SMM Estimation

I use the optimal weighting matrix

$$W=\left[N\mathrm{var}\left(\widehat{M}_N\right)\right]^{-1}.$$

[1] Since β is the cost of capital, we have $\beta=\frac{1}{1+E[R]}=\frac{1}{1.111}=0.9$.

I compute the 14x14 covariance matrix $\widehat{M_N}$ using the seemingly unrelated regressions approach. The moments can be expressed as the coefficients from the following system of regression equations:

$$\begin{aligned}
y_{it}^* &= \lambda_0 + \lambda_1 y_{it-1}^* + \Delta^{(-2)} + \Delta^{(-1)} + \Delta^{(0)} + \Delta^{(1)} + \Delta^{(2)} + \delta_{it} \\
\delta_{it}^2 &= Var(\delta) + w_{it} \\
d_{it} &= h^{(1-2)} + h^{(2-3)} + h^{(4-6)} + h^{(7+)} + \eta_{it} \\
Var_i(X_{it}) &= E[Var(X)] + e_i \\
(E_i[X_{it}] - E[E_i[X_{it}]])^2 &= Var(E[X]) + \iota_i
\end{aligned}$$

The coefficients $h^{(j)}$ are fixed effects for tenure (j). Var_i denotes variance within CEO spell i, and E_i denotes average within CEO spell i. I estimate each regression separately using ordinary least squares, which provides consistent estimates for each moment as well as regression disturbances. Each regression above has the form

$$Y_i = X_i \beta_i + \varepsilon_i,$$

where Y_i is $N_i \times 1$ and β_i is $k_i \times 1$. The covariance between moments estimators β_i and β_j is the $k_i \times k_j$ matrix

$$Cov\left(\widehat{\beta}_i, \widehat{\beta}_j\right) = \left(X_i' X_i\right)^{-1} X_i' \Omega_{ij} X_j \left(X_j' X_j\right)^{-1},$$

where $\Omega_{ij} = Cov(\varepsilon_i, \varepsilon_j)$ is the $N_i \times N_j$ matrix whose element t, s is $Cov(\varepsilon_{it}, \varepsilon_{js})$. I estimated the covariance matrix Ω_{ij} for each pair of moments ij, allowing for time series autocorrelation and also correlation across regressions.

I define

$$G_N = M_N - \frac{1}{S}\sum_{s=1}^{S} m_n^s(\theta).$$

Applying the result of Pakes and Pollard (1989) with the efficient weighting matrix, we obtain

$$\begin{aligned}
\sqrt{N}\left(\widehat{\theta} - \theta_0\right) &\rightarrow {}^d\mathcal{N}(0, \Omega) \\
\Omega &= \left(1 + \frac{1}{S}\right)\left(\Gamma' \Lambda^{-1} \Gamma\right)^{-1},
\end{aligned}$$

where S is the number of simulated data sets (I choose $S = 10$), $\Gamma = \text{plim}_{N\to\infty} \partial \widehat{G}(\theta_0) / \partial\theta'$ and $\Lambda = N\text{avar}(\widehat{M}(\theta_0)) = N\text{avar}(\widehat{m}(\theta_0))$. I estimate Γ by numerically differentiating $\widehat{G}\left(\widehat{\theta}\right)$ with respect to θ, and using $\widehat{\Lambda} = N\widehat{\text{var}}\left(\widehat{M}\right)$, as described above.

We have

$$\sqrt{N}\widehat{G}(\theta_0) \rightarrow^d \mathcal{N}\left(0, \left(1 + \frac{1}{S}\right)\Lambda\right),$$

so SMM provides the following test of the model's over-identifying restrictions:

$$\frac{NS}{1+S}\widehat{G}(\theta_0)' \Lambda^{-1} \widehat{G}(\theta_0)' \rightarrow^d \chi^2 \left(\#\text{moments - }\#\text{parameters}\right).$$