CEO Wage Dynamics: Estimates from a Learning Model

Lucian A. Taylor*

February 10, 2012

Abstract: The level of CEO pay responds asymmetrically to good and bad news about the CEO’s ability. The average CEO captures approximately half of the surpluses from good news, implying CEOs and shareholders have roughly equal bargaining power. In contrast, the average CEO bears none of the negative surplus from bad news, implying CEOs have downward rigid pay. These estimates are consistent with the optimal contracting benchmark of Harris and Holmstrom (1982) and do not appear to be driven by weak governance. Risk-averse CEOs accept significantly lower compensation in return for the insurance provided by downward rigid pay.

JEL codes: D83, G34, M12, J31, J33, J41

Keywords: CEO, compensation, learning, dynamics, bargaining, SMM

Note: The Internet Appendix is at the end of this document but can eventually live elsewhere.

* The Wharton School, University of Pennsylvania. Email: luket@wharton.upenn.edu. I am grateful for comments from Chris Armstrong, Philip Bond, Alex Edmans, Carola Frydman (discussant), Itay Goldstein, Mariassunta Giannetti (discussant), Wayne Guay, Andre Kurmann, Michael Lemmon, Evgeny Lyandres (discussant), Bang Nguyen, Kasper Nielsen, Gordon Phillips (discussant), Michael Roberts, Alexi Savov (discussant), Motohiro Yogo, and seminar participants at Carnegie Mellon University, Duke University, Laval University, University of Amsterdam, University of Pennsylvania, University of Utah, Yale University, the European Finance Association Meeting (2011), the FIRS Conference (2010), the Jackson Hole Finance Conference (2011), the Olin Conference on Corporate Finance (2011), the Rothschild Caesarea Center Conference (2010), the Society for Economic Dynamics Conference (2010), and the Western Finance Association Meeting (2011). I am grateful for financial support from the Rodney L. White Center for Financial Research.
I. Introduction

There is considerable debate over the level of executive pay. One on side, Bebchuk and Fried (2004) and others argue that weak governance allows executives to effectively set their own pay while disregarding market forces and shareholder value. On the other side, Gabaix and Landier (2008) and others argue that executive pay is determined in a competitive labor market, so executives have limited influence on their own pay.

This paper uses CEO wage dynamics as a laboratory for exploring this debate. Specifically, I examine how learning about the CEO’s ability affects the level (i.e. expectation) of the CEO’s pay. For example, a CEO’s perceived ability may increase after his firm delivers high profits. This in turn increases the firm’s expected future profits, since high-ability CEOs generate higher profits on average. The increase in expected profits is a positive surplus. How the CEO and shareholders split this surplus depends on their relative bargaining positions, which in turn depend on governance strength, contractual constraints, and outside options in the labor market. For instance, the CEO may be able to capture the positive surplus by bargaining for higher future compensation, as long as the CEO can make a credible threat to leave the firm, the firm cannot make a credible threat to replace the CEO, and renegotiating the CEO’s contract is not too costly. This paper’s goal is to measure the surpluses that result from learning, and also to measure how the CEO and shareholders split them. The estimates allow us to gauge CEOs’ influence over their own compensation.

Measuring the surpluses from learning presents a challenge, since we cannot directly observe perceived CEO ability, and since compensation and stock prices depend endogenously on perceived ability. These challenges lend themselves to a structural estimation approach, which infers unobservable quantities directly from endogenous patterns in the data. I estimate a model in which the CEO and shareholders gradually learn the CEO’s ability by observing the firm’s profits and an additional, latent signal. Stock prices, return volatility, and changes in the level of CEO pay respond endogenously to news about CEO ability.
I estimate the model’s five parameters using the simulated method of moments (SMM). Separate parameters control how the CEO and shareholders split positive surpluses (from good news) and negative surpluses (from bad news). Estimation uses data on excess stock returns and total CEO compensation (including stock and option grants) for 4,545 Execu-comp CEOs. I estimate the five parameters by matching 12 moments. The first 10 moments measure how excess stock return volatility varies with CEO tenure, and the last two measure the sensitivity of changes in the level of CEO pay to positive and negative lagged excess stock returns. The model fits these moments well, including the observed decline in return volatility with CEO tenure, first documented by Clayton, Hartzell and Rosenberg (2005).

The paper’s main result is that the level of pay responds asymmetrically to good and bad news about CEO ability: the average CEO bears none of the negative surplus resulting from bad news about ability, whereas he or she captures roughly half of the surplus from good news. One reason I find this result is that observed changes in the level of pay are more sensitive to positive than negative lagged excess stock returns.

Since the level of CEO pay does not fall after bad news about ability, CEOs have downward rigid pay. Shareholders, not the CEO, bear the negative surpluses from bad news about the CEO’s ability. Downward rigid pay is consistent with the optimal contracting benchmark of Harris and Holmstrom (1982). A long-term contract promising that the level of pay will never drop is optimal in their model, because workers are risk-averse, whereas the firm is risk neutral and hence can cheaply provide insurance against bad news. Downward rigid pay therefore does not necessarily imply weak governance. Indeed, I find that pay is downward rigid even in the subsample with high institutional ownership, a proxy for strong governance. Pay is slightly more downward rigid, although not significantly so, when the CEO has an explicit contract, also suggesting that the result is due to contracting rather than governance. Downward rigid pay is pervasive across subsamples formed on firm size.

\footnote{This result is about the level of pay, which is the CEO’s expected compensation each year. A CEO’s realized pay can decrease over time due to the random incentive component of compensation.}
industry, calendar year, and ten other characteristics. Downward rigid pay is also pervasive in non-CEO occupations (e.g. Baker, Gibbs, and Holmstrom (1994), Dickens et al. (2007)).

According to Harris and Holmstrom (1982), CEOs accept lower average pay in return for the insurance provided by downward rigidity, which saves the firm money. To quantify these savings, I ask how much higher CEO pay would have to be to keep CEOs indifferent between their actual, downward rigid compensation paths and counterfactual paths that are not downward rigid. If CEOs’ relative risk aversion is 0.5 (4.0), first-year pay must increase by a multiple of 1.3 (11.8), and the net present cost to the firm of five years of CEO pay would increase by a multiple of 1.003 (8.1). If CEOs are sufficiently risk averse, the firm’s savings from offering downward rigid pay are considerable.

Since CEOs capture roughly half of the positive surplus from good news, CEOs and firms appear to have roughly equal bargaining power on average. In contrast, the models of Jovanovic (1979), Harris and Holmstrom (1982), and Gibbons and Murphy (1992) all predict that workers capture 100% of the surplus from good news, because they can always threaten to take their ability to another firm at no cost. The estimates here suggest that the average CEO’s actual outside employment options are not as strong as these models assume. CEOs’ share of positive surpluses is significantly higher in the subsample with more institutional ownership, implying that strong CEO bargaining power is not inconsistent with strong governance. CEOs’ share of positive surpluses is also higher in subsamples with insider CEOs and heterogeneous industries, potentially because their firms have fewer potential replacement CEOs and hence less bargaining power. After controlling for multiple characteristics, CEOs’ share of positive surpluses is positively related to the number of similarly sized firms in the industry, a proxy for CEOs’ outside employment options.

Across industries, CEOs’ estimated share of positive surpluses ranges from 18% in Business Equipment to 77% in Shops. CEOs capture just 19% of positive surpluses in the finance industry, where the debate on executive pay has been especially contentious. This result im-
plies that financial firms’ shareholders, not their CEOs, are the main beneficiaries of good news about their CEOs’ ability.

For robustness, I show that extended models with endogenous CEO firings, gradual vesting of CEO pay, learning about firm quality, and persistent shocks to firm profitability produce slightly different parameter estimates. The average CEO’s estimated share of positive surpluses ranges from 44 to 68%. Pay is downward rigid in all specifications. An important caveat is that CEOs may be fired after enough bad news, so CEOs do bear personal costs from bad news about their ability. While important and interesting in their own right, these personal costs are unrelated to how CEOs and their current shareholders split the CEO’s surplus, which is this paper’s focus.

My model is a simplified version of the learning models by Jovanovic (1979), Harris and Hölmstrom (1982), Murphy (1986), Gibbons and Murphy (1992), and Hölmstrom (1999).2 This paper’s contribution is not to provide new predictions or further tests of existing predictions. Instead, the contribution is to quantify the surpluses from learning and measure how they are shared, which sheds light on CEOs’ bargaining power. The paper therefore complements existing empirical work that tests these models’ directional, reduced-form predictions but reveals less about underlying economic magnitudes.3 For example, Boschen and Smith (1995) find a positive correlation between pay and lagged stock returns. Without a structural model it is difficult to judge from their correlation whether CEOs capture 1%, 10%, or even 100% of positive surpluses. (I find they capture roughly 50%.) Whether CEOs capture 1% or 100% has very different implications for CEOs’ bargaining power. Also, the structural approach allows me to address an interesting counterfactual question: How much more would firms have to pay CEOs if their pay were not downward rigid?

Like this paper, Gabaix and Landier (2008) and Törvio (2008) also measure how CEOs

---

2Milbourn (2003) also models CEO pay and learning about ability, but his goal is to explain cross-sectional variation in stock-based compensation.

and shareholders split the CEO’s surplus. Their evidence comes from the cross section, whereas my evidence comes from the time series. Gabaix and Landier find that CEOs capture only 2% of the value they create. Alder (2009) applies different functional forms and finds a capture rate higher than Gabaix and Landier’s. Tervio (2008) finds that CEOs capture roughly 20% of the value they add to their firms. Examining stock returns around CEO deaths, Nguyen and Nielsen (2010) conclude that executives capture 80% of their surplus. Whereas the previous papers measure CEOs’ total surplus, this paper measures the surpluses created by news arriving each year, which allows me to distinguish between positive and negative surpluses.

The paper makes one modest methodological contribution. Allowing parameter values to vary with observable characteristics is rare in the structural corporate finance literature, probably due to computational costs. Taylor (2010) and Nikolov and Whited (2009) circumvent the problem by estimating in subsamples, which makes it difficult to control for several characteristics at once. I develop a method that solves this problem with minimal computational costs. The method can apply to any project using GMM or SMM estimation.

The paper is structured as follows. Section 2 presents the learning model’s assumptions and discusses identification. Section 3 describes the data and estimator. Section 4 presents parameter estimates and results on model fit. Section 5 quantifies the value of downward rigid pay. Section 6 describes how the parameters vary with firm, CEO, and industry characteristics. Section 7 discusses robustness, and Section 8 concludes.

---

4Notable exceptions include Morellec, Nikolov, and Schuerhoff (2008), Korteweg and Polson (2009), Albuquerque and Schroth (2010), and Dimopoulos and Sacchetto (2011). Hennessy and Whited (2005, 2007) use firm and year fixed effects to remove heterogeneity from the data before estimating.
II. The Dynamic Model of CEO Pay

The model features CEOs with different ability levels, meaning they can produce different average firm-specific profitability. Neither CEOs nor shareholders can observe a CEO’s true ability. Instead, they learn about ability over time by observing the firm’s realized profits and a shared, additional signal. When a CEO’s perceived ability changes, so does his or her perceived contribution to future firm profits. This change is a surplus, which the CEO and shareholders split according to parameters $\theta^{up}$ and $\theta^{down}$.

Despite its simplicity, the model still allows me to empirically identify the magnitude of surpluses from learning and how they are split. In Section 7 I extend the model to include endogenous CEO firings, learning about firm quality, and persistent profitability shocks. I show that these extended models produce similar estimates. I focus on the simpler model here, because the intuition, solution, and estimation results are more transparent.

A. Assumptions

The model features firms $i$ that live an infinite number of years $t$.

**Assumption 1:** The gross profitability (profits before CEO pay, divided by assets) of firm $i$ realized at the end of year $t$ equals

$$Y_{it} = a_i + \eta_i + v_t + \varepsilon_{it}. \quad (1)$$

The unobservable ability of the CEO in firm $i$ at time $t$ is $\eta_i$, which is constant over the CEO’s tenure. To be precise, $\eta_i$ should have a CEO-specific subscript, since different CEOs within the firm can have different abilities. Parameter $a_i$, which will drop out of the analysis, reflects the contribution of non-CEO factors in firm $i$. For now I assume $a_i$ is known and constant. Shock $v_t$ has conditional mean zero; this shock is common to all firms in the
industry. Shock $\varepsilon_{it}$ is an unobservable i.i.d. firm-specific shock distributed as $\mathcal{N}(0, \sigma^2)$. There are very many firms in the same industry as firm $i$, which implies that the industry shock $v_t$ is observable even though $\eta_t$ and $\varepsilon_{it}$ are not.

**Assumption 2:** Investors use exogenous discount factor $\beta$ to discount future dividends. The firm immediately pays out any cash flows, including negative cash flows, as dividends.

This assumption allows me to solve for the firm’s market value. It improves tractability by making the firm’s book assets constant over time.\(^5\)

CEO $j$ spends a total of $T_j$ years in office. $T_j$ is exogenous and known when the CEO is hired.

**Assumption 3:** Agents have common, normally distributed prior beliefs about the ability of a newly hired CEO in firm $i$: $\eta_t \sim \mathcal{N}(m_{0i}, \sigma^2_0)$. Different firms $i$ may hire from different CEO talent pools, so the prior mean ability of CEOs, $m_{0i}$, is firm specific. The prior mean $m_{0i}$ will drop out of the analysis.

**Assumption 4:** Investors use Bayes’ Rule to update beliefs about CEO ability $\eta_t$ after each year. They update their beliefs by observing the firm’s profitability $Y_{it}$ and an additional, latent, orthogonal signal $z_{it}$ that is distributed as $z_{it} \sim \mathcal{N}(\eta_t, \sigma^2_z)$.

The additional signal $z$ represents information unrelated to current profitability, possibly from the CEO’s specific actions and choices, the performance of individual projects, the CEO’s strategic plan, the firm’s growth prospects, discretionary earnings accruals, and media coverage. I include this additional signal for two reasons. First, there is evidence that investors use signals besides profitability when learning about CEO ability (Cornelli, Kominek, Ljungqvist (2010)). Second, the additional signal helps the model fit certain features of the data, as I explain later. The additional signal $z$ is more precise when its volatility

\(^5\)This assumption has little effect on the estimation results, since identification does not rely on changes in firm size, and since I use data on stock returns rather than earnings or cash flows.
Assumption 5: Realized total compensation for the CEO in firm $i$ and year $t$ equals

$$ w_{it} = E_t[w_{it}] + b_{it}r_{it}. $$

Realized pay is the sum of its level ($E_t[w_{it}]$, known at the beginning of $t$) and a random component that depends on the firm’s firm’s endogenous industry-adjusted stock return ($r_{it}$) and the CEO’s contemporaneous pay-performance sensitivity ($b_{it}$).

The model makes predictions about changes in the level of CEO pay, and I use these predictions to estimate the model. The contemporaneous pay-performance sensitivity $b_{it}$ represents the incentive component of pay and depends on the CEO’s bonus and holdings of stock and options. I treat this sensitivity $b_{it}$ as exogenous, for four reasons. First, I do not need predictions about $b_{it}$ to estimate the model. Second, making $b_{it}$ endogenous does not materially change the model’s predictions about the changes over time in expected pay. Gibbons and Murphy (1992) make $b_{it}$ endogenous by incorporating moral hazard, effort choice, and optimal contracts into a model of learning about an executive’s ability. They show that the optimal contract sets the contemporaneous pay-performance sensitivity so that the CEO exerts optimal effort. The contract sets expected pay so that the CEO agrees to stay in the firm rather than leave to some outside option. More importantly, they show that making $b_{it}$ endogenous does not significantly change the model’s predictions about the change over time in a CEO’s expected pay. Third, modeling $b_{it}$ in a reduced-form manner allows it to depend flexibly on firm and CEO characteristics. Of course, making $b_{it}$ exogenous significantly simplifies the model solution and estimation.
Assumption 6: The change in the level of pay is

$$\Delta E_t \[w_{it}\] \equiv E_t \[w_{it}\] - E_{t-1} \[w_{it-1}\] \tag{3}$$

$$= \theta_t B_i (E_t \[\eta_i\] - E_{t-1} \[\eta_i\]) \tag{4}$$

$$\theta_t = \theta^{up} \quad \text{if} \quad E_t \[\eta_i\] \geq E_{t-1} \[\eta_i\] \quad \text{(beliefs increase)} \tag{5}$$

$$\theta_t = \theta^{down} \quad \text{if} \quad E_t \[\eta_i\] < E_{t-1} \[\eta_i\] \quad \text{(beliefs decrease)} \tag{6}$$

By equation (1), the CEO’s expected contribution to firm profits (in dollars) in year $t$ is $B_i E_t \[\eta_i\]$, where $B_i$ is the firm’s assets. The change in this expected contribution is $B_i (E_t \[\eta_i\] - E_{t-1} \[\eta_i\])$. This quantity is the surplus created by news in year $t - 1$ about the CEO’s ability. Assumption 6 therefore states that the CEO captures a fraction $\theta_t$ of the change in his expected contribution to firm profits. When beliefs increase (decrease), the CEO captures a fraction $\theta^{up}$ ($\theta^{down}$) of this surplus. The parameters $\theta^{up}$ and $\theta^{down}$ measure the CEO’s bargaining power over changes in the level of pay. This assumption follows the literature’s practice of assuming that a surplus is split according to a constant, exogenous parameter (e.g. Morellec, Nikolov, and Schurhoff (2010)).

Assumption 6 nests the reduced-form predictions from several existing theories with stronger micro foundations, which I summarize below. Adding these theories’ micro foundations would constrain $\theta^{up}$ and $\theta^{down}$ but not otherwise change Assumption 6. Estimating $\theta^{up}$ and $\theta^{down}$ allows me to compare the data to these theories’ predictions.

The theories below illustrate that there are several economic factors that affect how CEOs and shareholders split surpluses. This paper’s main goal is not to measure the relative importance of these and other factors, but to measure their total effect on CEO bargaining power. However, Section 6 takes initial steps toward comparing the various factors by measuring how my estimates vary cross-sectionally.

A special case of the model is when the CEO’s expected pay exactly equals the CEO’s
expected contribution to firm profits \( (B_t E_t[\eta_t]) \) every year. This special case matches the predictions from the equilibrium learning models of Jovanovic (1979), Gibbons and Murphy (1992), and Hölmstrom (1999). I summarize the assumptions of Gibbons and Murphy (1992) to provide micro foundations for this special case. They assume that every period there are multiple identical firms competing for the CEO. The firms offer the CEO single-period contracts, and the CEO chooses his or her preferred contract. The CEO’s outside option is to work at one of these firms, and the firm’s outside option is to hire a new CEO whose ability is a random draw from the talent pool. The equilibrium level of pay therefore equals the CEO’s perceived contribution to the firm every year. As a result, the CEO captures \( \theta^{up} = \theta^{down} = 100\% \) of the surplus from learning. Later, I show that the data are not consistent with this benchmark.

As discussed earlier, Harris and Hölmstrom (1982) provide a second micro foundation with optimal contracts. Their model predicts that the CEO’s expected pay never drops \( (\theta^{down} = 0) \), but the CEO has strong enough outside options to capture 100% of positive surpluses \( (\theta^{up} = 1) \). I show that the data are closer to this benchmark.

Other models omit learning but focus on labor market frictions that affect bargaining power. If the CEO’s human capital is specific to the firm, then the CEO possibly cannot make a strong threat to leave firm, hence the CEO captures less of a positive surplus (Murphy and Zabojnik, 2007). If the CEO’s outside option is to work in a smaller firm, as in the matching models of Gabaix and Landier (2008) and Tèrvio (2008), then the CEO’s outside option and bargaining power are weaker. If the CEO would lose unvested shares and options by leaving the firm, the CEO’s bargaining position is weaker.

B. Model Solution and Identification

Next, I summarize the model’s predictions and provide intuition for how the estimation procedure identifies parameter values from the data. Closed-form solutions and proofs are
in the Internet Appendix.\textsuperscript{6}

The first predictions are about the volatility of excess stock returns. The model predicts that return volatility decreases with CEO tenure. The reason is that uncertainty about CEO ability contributes to uncertainty about dividends, and this uncertainty gradually decays to zero as investors learn. Return volatility eventually reaches a level that depends only on $\sigma_\epsilon$, the volatility of shocks to profitability. Data on return volatility for long-tenured CEO therefore identify parameter $\sigma_\epsilon$.

Parameter $\sigma_z$, the additional signal’s noise, is mainly identified off how quickly return volatility drops with CEO tenure. Return volatility drops faster when $\sigma_z$ is lower, meaning the additional signal $z_{it}$ is more precise. The reason is that a more precise signal allows agents to learn the CEO’s ability more quickly. A more precise signal also increases return volatility in the CEO’s first year, because beliefs about the CEO and hence stock prices move more during that year.

The amount by which stock return volatility drops is (i) increasing in the amount of prior uncertainty about CEO ability ($\sigma_0$), but (ii) decreasing in CEOs’ share of the surplus ($\theta^{up}$ and $\theta^{down}$). The intuition for (i) is that if uncertainty about the CEO has farther to drop, so does return volatility. Return volatility does not drop at all if there is no uncertainty about the CEO ($\sigma_0 = 0$). The intuition for (ii) is that there is more uncertainty about dividends if CEOs capture a smaller share (and hence shareholders capture a larger share) of the surpluses from news about ability.

The next predictions are about the relation between CEO pay and stock returns. The model predicts a positive relation between changes in expected pay and the firm’s lagged excess stock return, consistent with the empirical evidence of Boschen and Smith (1995). To see why, suppose the firm experiences higher than expected profits in year $t - 1$. This has two effects: a positive excess stock return in year $t - 1$, and an increase in the CEO’s pay.\textsuperscript{6}
perceived ability, which in turn makes expected CEO pay higher in year $t$ than $t-1$.

The sensitivity of expected pay to lagged returns depends on $\theta_{up}$ when surpluses are positive and on $\theta_{down}$ when surpluses are negative. Positive surpluses typically (but not always\(^7\)) coincide with positive excess stock returns. As a result, the sensitivity of changes in expected CEO pay to positive lagged excess returns is most informative about $\theta_{up}$. Conversely, the sensitivity of pay to negative returns is most informative about $\theta_{down}$. The sensitivities of CEO pay to positive and negative lagged returns will help to disentangle $\theta_{up}$ and $\theta_{down}$ in the estimation procedure.

The predicted sensitivity of pay to lagged returns increases in both prior uncertainty ($\sigma_0$) and the CEO’s share of the surplus ($\theta_t$). When there is more prior uncertainty, beliefs move more in response to any given signal. The surpluses are therefore larger in magnitude, so the changes in CEO pay are also larger. Expected CEO pay moves more with lagged stock returns when $\theta_t$ is higher, because the CEO captures a larger fraction of the surpluses.

**INSERT FIGURE 1 HERE**

Figure 1 illustrates how the predictions above allow us to separately identify prior uncertainty ($\sigma_0$) and the CEO’s share of the surplus ($\theta_t$) from the data. The key is that these two parameters have different predicted effects on (a) the drop in return volatility and (b) the sensitivity of pay to lagged returns. To simplify the explanation, suppose $\theta_{up} = \theta_{down} = \theta$. The solid line shows the infinitely many combinations of $\sigma_0$ and $\theta$ that allow the model to match a given, observed drop in return volatility. The line slopes up, because the predicted drop in return volatility is increasing in $\sigma_0$ but decreasing in $\theta$, as explained above. The dashed line shows the infinitely many combinations of $\sigma_0$ and $\theta$ that allow the model to fit a given, observed sensitivity of CEO pay to lagged returns. The line slopes down, because the predicted sensitivity is increasing in both $\sigma_0$ and $\theta$ (also above). The lines’ opposite slopes make them cross at a unique point $\{\sigma_0, \theta\}$. The estimation procedure will find this unique

\(^7\)If the additional signal $z_{it}$ is high enough, the CEO’s perceived ability and pay level can increase even if the profitability shock and excess return are zero or negative.
point that lets the model simultaneously match both moments.

III. Estimation

A. Data

Data come from Execucomp; CRSP; Compustat; Kenneth French’s website; Thomson Financial; Risk Metrics; Gillan, Hartzell, and Parrino (2009); Peters and Wagner (2009); and Jenter and Kanaan (2011).\textsuperscript{8} The sample includes CEOs in the Execucomp database from 1992 to 2007. Execucomp includes S&P 1500 firms, firms removed from the S&P 1500 that are still trading, and some client requests. The Internet Appendix provides details on how I construct the sample.

Excess return \( r_{it} \) is the firm’s annual stock return minus the equal-weighted Fama French 49 industry return. I account for firms’ fiscal calendars when computing fiscal year returns.

The measure of CEO pay is Execucomp’s TDC1, which includes each year’s salary, bonus, total value of restricted stock granted, total value of stock options granted (using Black-Scholes), long-term incentive payouts, “other annual,” and “all other total.” A plausible interpretation of the model is that the firm and CEO renegotiate the labor contract at the beginning of each year. The contract sets expected pay in the coming year to the level that induces the CEO to remain at the firm and work throughout year \( t \). The benefit of using TDC1 is that it excludes stock and options that were granted and vested in past years, which are sunk and therefore should not affect the CEO’s decision to stay in the firm in the current year. For robustness, in Section 7 I use an alternate measure that includes stock and options in the year they vest. Estimation uses data on the change in realized CEO pay, divided by lagged market cap. I winsorize this variable at the 1st and 99th percentiles.

\textsuperscript{8}I thank these authors for generously sharing their data.
I measure the annual variance of excess stock returns by taking the variance of weekly industry-adjusted stock returns during each firm’s fiscal year, then multiplying by 52 to annualize. I winsorize this variable at the 1st and 99th percentiles.

Another estimation input is $T_j$, the total years CEO $j$ spends in office. If $T_j$ is known (i.e. CEO’s last year in office is in Execucomp) then I use the actual value. If $T_j$ is not known (i.e. CEO’s last year is not in Execucomp), then I forecast it using the CEO’s age and tenure from his last observation in the database; details are in the Internet Appendix.

**INSERT TABLE 1 HERE**

Summary statistics are in Table 1. The database contains 20,700 firm/year observations and 4,545 CEOs. Mean realized pay is $4.13 million, with a standard deviation of $5.60 million. A CEO’s realized pay fluctuates considerably over time: the change in realized pay, as a fraction of lagged market cap, has a standard deviation of 0.40%. The median firm/year observation is for a CEO in his 6th year in office. The median firm has $1.27 billion in assets and a market capitalization of $1.16 billion. Section 6 discusses the remaining variables in Table 1.

**B. Estimator**

I estimate the five model parameters in $\Theta = \left[ \sigma^2_{\varepsilon}, \sigma^2_z, \sigma^2_{0}, \theta^{up}, \theta^{down} \right]$ using SMM. The estimator is

$$\hat{\Theta} \equiv \arg \min_\Theta \left( \hat{\mathbf{M}} - \hat{\mathbf{m}}(\Theta) \right)' \mathbf{W} \left( \hat{\mathbf{M}} - \hat{\mathbf{m}}(\Theta) \right).$$  \hspace{1cm} (7)

$\hat{\mathbf{M}}$ is a vector of moments estimated from the actual data, and $\hat{\mathbf{m}}(\Theta)$ is the corresponding vector of model-implied moments. The hat on $\hat{\mathbf{m}}$ indicates that some model-implied moments are estimated by simulation. For these simulations, I use parameter values $\Theta$ to simulate a sample many times larger than the empirical sample, then I compute the moment from

---

9See, for instance, Strebulaev and Whited (2012).
simulated data in the same way I compute the empirical moment. I set $\mathbf{W}$ equal to the efficient weighting matrix, which is the inverse of the estimated covariance of moments $\mathbf{M}$.

I estimate the five parameters in $\Theta$ using 12 moments in vectors $\mathbf{M}$ and $\mathbf{m}$. The first moment is the average variance of excess returns for CEOs in their first year in office. Moments 2–10 measure how return volatility varies with CEO tenure. Specifically, using firm/year data I regress the variance of excess stock returns on nine dummy variables for CEO tenure equal to 2, ..., 9, and 10+ years; the log of the firm’s lag assets; the log of firm age; and industry × year fixed effects.\(^{10}\) The slopes on the nine CEO tenure dummies make up moments 2–10.

The 11th and 12th moments are the slopes $M^{(11)}$ and $M^{(12)}$ from a regression of scaled changes in expected pay on positive and negative lagged excess returns as well as an indicator for the lagged return’s sign:

$$\frac{\Delta \mathbf{E}_{t \mid w_{it}}}{M_{it-1}} = a_0 + a_1 \mathbf{1} (r_{it-1} > 0) + M^{(11)} r_{it-1}^{(+)} + M^{(12)} r_{it-1}^{(-)} + \epsilon_{it};$$

(8)

where $r_{it-1}^{(+)} = r_{it-1}$ if $r_{it-1} \geq 0$ and equals zero otherwise; vice-versa for $r_{it}^{(-)}$. Estimating (8) is straightforward using simulated excess returns and changes in expected pay. To estimate (8) using actual data on realized pay, first I parameterize the contemporaneous pay-performance sensitivity $b_{it}$ from Assumption 5 as

$$b_{it} = \begin{cases} M_{it} b^{(+)} & \text{if } r_{it} \geq 0, \\ M_{it} b^{(-)} & \text{if } r_{it} < 0, \end{cases}$$

(9)

(10)

where $b^{(+)}$ and $b^{(-)}$ are nuisance parameters to be estimated. This parametrization takes into account that dollar changes in CEO pay are larger in larger firms, and that firms may set

\(^{10}\)I include the control variables and fixed effects so that the slopes on the CEO tenure dummies do not simply capture correlated changes in firm size, firm age, or unobserved heterogeneity across industry/years. Hennessy and Whited (2005, 2007) also used fixed effects to remove heterogeneity from the data before estimating.
different sensitivities to positive and negative contemporaneous returns. I substitute these equations into (8) to derive a regression model for changes in realized pay:\footnote{Assumption 5 implies \( E_t[w_{it}] = w_{it} - b_{it} r_{it} \), so equation (8) implies}

\[
\frac{\Delta w_{it}}{M_{it-1}} = a_0 + a_1 1 (r_{it-1} > 0) + \lambda_{11} r_{it-1}^{(+)1} + \lambda_{12} r_{it-1}^{(-)} + b^{(+)1} \left[ \frac{M_{it}}{M_{it-1}} r_{it}^{(+)} \right] + b^{(-1)} \left[ \frac{M_{it}}{M_{it-1}} r_{it}^{(-)} \right] + e_{it} \tag{11}
\]

\[
\lambda_{11} = M^{(11)} - b^{(+)1} \tag{12}
\]

\[
\lambda_{12} = M^{(12)} - b^{(-1)}. \tag{13}
\]

I estimate this regression by OLS and then estimate the last two moments as

\[
\hat{M}_{11} = \hat{\lambda}_{11} + \hat{b}^{(+)1} \tag{14}
\]

\[
\hat{M}_{12} = \hat{\lambda}_{12} + \hat{b}^{(-1)}. \tag{15}
\]

This procedure accounts for the estimation error in nuisance parameters \( b^{(+)1} \) and \( b^{(-1)} \) when measuring the error in moments \( M^{(11)} \) and \( M^{(12)} \).

\section*{IV. Estimation Results}

I begin by describing how the model fits the data. I then present the main parameter estimates, which characterize the average firm and CEO. In section 6 I describe how the parameter estimates vary across firms and CEOs.
A. Model Fit

Panel A of Table 2 contains actual and simulated values of the 12 moments used in the SMM estimation along with t-statistics testing their difference. The model closely matches the level of return volatility in CEOs’ first year in office and also the tenure fixed effects in return volatility. Figure 2 plots these fixed effects to make interpretation easier. In both the actual data (dashed line) and simulated data (solid line), return volatility drops rapidly in CEOs’ first three years in office, indicating that agents learn the CEO’s ability quite fast. The model fits the timing and magnitude of these changes remarkably well.

INSERT TABLE 2 HERE

INSERT FIGURE 2 HERE

Return volatility decreases with tenure in the model due to learning. Outside the model, an alternate explanation for this decrease is that earnings volatility declines with tenure. In the Internet Appendix I show there is no significant relation between CEO tenure and the volatility of firm profitability. Also, I show that the decline in return volatility is robust to controlling for the magnitude of the shock to profitability in the same firm and year.

The last two moments in Table 2 measure the sensitivity of changes in the pay level to both positive and negative lagged excess returns. Both sensitivities are significantly positive, and the model fits both almost exactly. The sensitivity to positive returns is more than 3 times larger than the sensitivity to negative returns. The result is not due to the convexity of stock option payoffs, because the measure of CEO pay values options at their grant date, not their payout or vesting dates.

Table 2 also presents the J-test of the model’s overidentifying restrictions. We cannot reject the hypothesis that the model matches all 12 empirical moments (p=0.965). While this result is good news for the model, we know that with enough additional data we will surely reject any model, including this one.
Panel B of Table 2 illustrates this point by showing some features of the data that the model fits less well. Specifically, I measure the inflation-adjusted change in a CEO’s realized pay from year 1 to year 5 in office, scaled by the firm’s assets at the beginning of the 1st year. In both actual and simulated data, the median CEO’s pay level increases with tenure, consistent with the empirical findings of Murphy (1986) and Cremers and Palia (2011). Pay ratchets up over time in the model because, as explained below, CEOs avoid negative surpluses but capture a large portion of positive surpluses. Pay increases even for the simulated 25th percentile CEO, who receives mostly bad news about ability. This last prediction is at odds with the data: the actual 25th percentile CEO sees pay decrease, albeit by only 0.019% of firm assets. Also, the model generates larger changes in pay than we see in the data: pay for the 75th percentile CEO increases by 1.65% of firm assets in the model but only by 0.10% in the actual data. One potential explanation, which additional regressions reject, is that the changes in expected pay correlated with lagged returns are transitory, not permanent as the model assumes. Extending the model to accommodate these additional features of the data is an interesting avenue for future work.

B. Main Parameter Estimates

Table 3 presents the main parameter estimates along with results from robustness specifications discussed in Section 7. The estimated standard deviation of profitability shocks ($\sigma_\varepsilon$) is 36% per year. The model needs this high value to match the high level of stock return volatility. DeAngelo, DeAngelo, and Whited (2010) directly estimate the annual volatility of innovations to profitability, and they find a much lower value, 7% per year. This paper’s high estimate of $\sigma_\varepsilon$ reflects the well known excess volatility puzzle (Shiller, 1981): it is difficult to reconcile the high level of return volatility with the relatively low level of earnings.

---

12 Following equation (2), realized pay equals expected pay plus the pay-performance sensitivity $b_{it}$ (estimated in regression (11)) times the contemporaneous excess return.

13 Specifically, I include $r_{t-2}$ in regression (11) and find that its slope is indistinguishable from zero. If the shocks were transitory then we should find a negative slope.
volatility. In Section 7 I show that we can interpret $\sigma_{\epsilon}$ as the volatility of shocks to current plus discounted future profitability, in which case an estimate of 36% makes more sense.

INSERT TABLE 3 HERE

The high estimate of $\sigma_{\epsilon}$ implies that profitability is a very noisy signal of CEO ability. The additional signal $z$ is much less noisy, with an estimated volatility of 3.3% per year. The precise $z$ signal allows agents to learn quickly, which allows the model to fit the sharp drop in return volatility during CEOs’ first three years in office. Consistent with these results, Cornelli, Kominek, and Ljungqvist (2010) show that boards rely on soft information in addition to firm performance when learning a CEO’s ability. Taylor (2010) finds that a precise additional signal of CEO ability is needed to rationalize data on firm profitability and CEO firings.

The estimated standard deviation of prior beliefs about CEO ability ($\sigma_0$) is 4.1%. Using this estimate, the difference in average profitability between CEOs at the 5th and 95th ability percentiles is $2 \times 1.65 \times \hat{\sigma}_0 = 13.6\%$ of assets per year, which is quite large. For comparison, using a different data set and identification strategy, Taylor (2010) estimates a 2.4% standard deviation in prior beliefs about shareholders’ share of the surplus from CEO ability. Not surprisingly, the prior uncertainty about the total surplus, which I measure in this paper, is higher. Bertrand and Schoar (2003) estimate manager-specific fixed effects in annual profitability. They find a 7% standard deviation in fixed effects across managers, implying even greater dispersion in ability than reported here.

CEOs’ estimated share of a negative surplus ($\theta^{\text{down}}$) is -5.1%, indicating the level of pay actually increases slightly following bad news. However, the estimate is not statistically different from zero, so I cannot reject the hypothesis that CEO pay is perfectly downward rigid. This result implies that shareholders, not the CEO, bear the entire negative surplus resulting from bad news about a CEO’s ability. Since a negative value of $\theta^{\text{down}}$ is somewhat implausible ex ante, I re-estimate the model with the constraint $\theta^{\text{down}} \geq 0$ (Table 3, “Con-
strained $\theta^*$. This constraint has little effect on the other parameter estimates or model fit. I impose this constraint throughout the rest of the paper.

CEOs’ estimated share of a positive surplus ($\theta^{up}$) is 48.9%. In other words, the level of CEO pay changes 0.489 for one with increases in the CEO’s perceived contribution to firm profits. CEOs and shareholders almost equally share the benefits from an improvement in the CEO’s perceived ability.

In sum, I find that CEO pay responds asymmetrically to good and bad news. I can reject the hypothesis that $\theta^{up} = \theta^{down}$ with a $t$-statistic of 3.9. The main reason I find $\theta^{up} > \theta^{down}$ is that changes in CEO pay are more sensitive to positive lagged returns than to negative lagged returns, as shown in Table 2.

V. How Valuable Is Downward Rigid Pay?

Downward rigid pay insures CEOs against bad news about their ability. According to Harris and Holmstrom (1982), CEOs are willing to pay for this insurance by accepting lower average compensation, which adds value to the firm. In this section I quantify the value of this insurance to the CEO and firm by comparing the estimated model to a counterfactual model without downward rigid wages.

I start by simulating compensation paths for the median sample firm using Table 3’s parameter estimates. The main model only makes predictions about changes in pay. To obtain predictions about the level of pay, I set CEOs’ average first-year pay to $2.84M, the sample median in 2011 dollars. Panel A of Table 4 shows that simulated pay rarely decreases. The simulated net present cost to the firm of five years of CEO pay is $55.3M.

To gauge how CEO pay would change if it were not downward rigid, I compare the
compensation paths above to paths simulated from a counterfactual model that is identical to the main model but assumes \( \hat{\theta}_{down} = \hat{\theta}_{up} = 0.489 \). I solve for the level of first-year pay that makes the CEO’s expected utility the same as in the base-case simulations above. In other words, I ask how the starting level of pay must change to make the CEO indifferent between having and not having downward rigid pay. I assume the CEO has constant relative risk aversion preferences over realized pay in years one to five. I repeat the exercise using risk aversion coefficients between 0.5 to 4.\textsuperscript{14}

Simulated counterfactual results are in Panel B. Pay now decreases as often as it increases, which by itself makes average future pay lower than in the base-case model. First-year pay must increase to compensate the CEO for the lower average future pay. The larger spread between the 25th and 75th percentile CEOs indicates that pay has also become riskier, so first-year pay must increase even more to compensate CEOs for this added risk. With relative risk aversion set to 0.5, first-year pay increases from $2.84M to $13.1M, and the net present cost of pay to the firm increases from $55.3M to $62.6M (a factor of 1.13). When relative risk aversion increases from 0.5 to 4, the CEO requires more compensation for his riskier pay, so first-year pay increases from $13.1M to $113.8M, and the net present cost to the firm increases from $62.6M to $419.0M, 7.6 times higher than with downward rigid pay. These costs of removing downward rigid pay are not trivial compared to the median sample firm’s assets, $1.6 billion in 2011 dollars.

A potential concern is that pay is more volatile in the model than in the actual data (recall Table 2), so the results above likely overestimate the value of downward rigid pay. To address this concern, I repeat the exercise by bootstrapping actual data instead of simulating. Specifically, I compare compensation paths sampled from their actual distribution (which features downward rigid pay) and a counterfactual distribution constructed so that pay decreases exactly as often as it increases. The counterfactual distribution includes all actual

\textsuperscript{14}CRRA preferences cannot accommodate negative wages, which the model sometimes produces. I bound pay from below at $100K per year.
pay paths where 5th year pay exceeds 1st year pay, and also the mirror image of those same paths.\textsuperscript{15} As before, I ask how much first-year pay must increase to make the CEO indifferent between the actual and counterfactual compensations paths, the latter being both riskier and lower, all else equal.

Panel C describes the empirically sampled paths. As expected, pay increases more often than it decreases. The net present cost of 5 years of pay is $15.46M. Panel D describes the counterfactual paths. With a risk aversion coefficient of 0.5, first-year pay must increase from $2.84M to just $3.58M, and the net present cost of pay increases by just a multiple of 1.003. However, when the risk aversion coefficient increases to 4, first-year pay must increase to $33.6M and the net present cost increases by $109M (a multiple of 8.1).

To summarize, the costs of eliminating downward rigid pay are not small if CEOs are sufficiently risk averse. This exercise may push the model beyond its limitations. Holding other parameters constant while varying $\theta_{\text{down}}$ is especially aggressive. For instance, eliminating downward rigid pay would likely change the risk aversion and prior uncertainty ($\sigma_0$) of those who choose to become CEOs, which in turn could affect firm value. This exercise takes a first step toward valuing the insurance provided by CEOs’ downward rigid wages. Hopefully future research will provide more refined estimates.

VI. CEO Wage Dynamics and the Cross Section

The results so far quantify the surpluses from learning and how they are shared for the average CEO and firm. Now I begin exploring why surpluses are shared the way they are, and why the surpluses have their observed magnitude. I do so by measuring how parameters estimates vary in the cross section with proxies for governance strength, CEOs’ outside

\textsuperscript{15}The mirror image paths start from the same first-year pay, but the subsequent changes in pay have the opposite sign as the actual data. As a result, 5th year pay is less than 1st year pay in the mirror image paths. As before, I bound pay from below by $\$100K$, and I normalize all paths’ first-year pay to the sample median, $\$2.84M$. 
options, contractual constraints, and prior uncertainty. An important caveat is that all the proxies are endogenous and I lack instruments, so the correlations below do not have a causal interpretation.

I use five proxies for CEOs’ outside employment opportunities: the number of years the CEO spends in the firm before becoming CEO (“insider” status, a proxy for firm-specific human capital), the fraction of industry CEOs promoted from within the firm, the homogeneity of firms in the industry,\(^ {16}\) the number of similarly sized firms in the same industry, and the number of outside directorships the CEO holds. I also include two contracting variables that may affect bargaining outcomes: the amount of unvested shares and options the CEO holds, and an indicator for whether the CEO has an explicit employment agreement.\(^ {17}\) The proxies for prior uncertainty are the CEO’s age and insider status. I also include as control variables the log of the firm’s lagged assets, the fraction of shares held by institutional investors (a proxy for governance strength), and the log of the firm’s age. Detailed definitions of these variables are in the Appendix, and summary statistics are in Table 1.

Next I describe the method for measuring how the model’s five parameters vary with the characteristics above. The main idea is that the structural parameters vary with a characteristic like firm size (for instance) only if the 12 moments \(M\) used in SMM estimation vary with firm size. I use the following formula to measure the change in parameter estimates \(\hat{\Theta}\) associated with a small change in characteristic \(Z_j\) (e.g., firm size), holding constant other characteristics \(Z_{\sim j}\):

\[
\frac{\partial \hat{\Theta}}{\partial Z_j} = \frac{\partial \hat{\Theta}}{\partial M} \frac{\partial M}{\partial Z_j}.
\] (16)

The right-hand side equals the parameter estimates’ sensitivity to moments’ values (computed by perturbing each moment and re-estimating), times the moments’ sensitivity to

\(^{16}\)This variable, which is from Parrino (1997) and Gillan, Hartzell, and Parrino (2009), equals the median across Execucomp firms in the same industry of the \(R^2\) from time-series regressions of monthly stock returns on equal-weighted industry portfolio returns.

\(^{17}\)Data on employment agreements are from Gillan, Hartzell, and Parrino (2009). CEOs have an explicit employment agreement in 42% of firm/year observations.
characteristics (computed in OLS regressions). The Appendix provides implementation details. Estimated sensitivities $\partial \hat{\Theta} / \partial Z_j$ are in Table 5. I supplement these results by estimating the model in subsamples formed on ten characteristics (Table 6).\textsuperscript{18} The benefit of the subsample approach is that it measures the effect of large changes in characteristics, whereas the approach in equation (16) only measures the effect of small, local changes. The benefit of equation (16) is that, like a multiple regression, it controls for multiple characteristics at once, whereas the subsample approach does not.

\textbf{INSERT TABLE 5 NEAR HERE}

\textbf{INSERT TABLE 6 NEAR HERE}

CEOs’ share of the surplus from good news ($\theta^{up}$) is significantly higher in the subsamples with insider CEOs, lower industry homogeneity, fewer directorships, smaller firms, higher institutional ownership, and older firms. One interpretation is that insider CEOs in heterogeneous industries have more bargaining power, since their firms have fewer potential replacement CEOs with similar expertise. Weak governance does not appear responsible for CEOs capturing positive surpluses, since $\theta^{up}$ is actually higher when there is more institutional ownership. After controlling for multiple characteristics in Table 5, the subsample differences lose significance. Instead, we see CEOs capturing more of the positive surplus in industries with more insider CEOs, possibly because their firms have fewer replacement CEOs. There is also a positive relation between $\theta^{up}$ and the number of similarly sized firms within the same industry, consistent with CEOs having a stronger bargaining position when there are more firms where they could potentially work.

Parameter $\theta^{down}$ is either constrained at zero or indistinguishable from zero in all 20 subsamples, indicating that downward rigid pay is pervasive. Weak governance does not

\textsuperscript{18}The model is not well identified in subsamples based on whether the CEO has an explicit contract, because in both subsamples return volatility does not decline with tenure. This is a small-sample problem, as data on CEO contracts are missing in roughly 80% of the sample. I impose the estimation constraint $\sigma_z \geq 0.01$, which binds in six of 20 subsamples in which the decline in return volatility occurs over just one year. Without this constraint the model is poorly identified in these six subsamples.
appear responsible for downward rigid pay, which is even more pronounced (although not significantly so) when there is more institutional ownership (Table 5). Table 5 shows that pay is also more downward rigid (although not significantly so) when the CEO has an explicit contract, which is consistent with downward rigidity being a feature of optimal contracts, as Harris and Hölmstrom (1982) predict.

Prior uncertainty about ability ($\sigma_0$) is significantly higher, and hence the surpluses from learning are larger in magnitude, in the subsamples with outsider CEOs, more directorships, smaller firms, and younger firms. After controlling for multiple characteristics in Table 5, only firm size remains significant, potentially because CEO ability matters more (as a fraction of assets) in smaller firms, where the CEO necessarily delegates less. We also see a negative relation between uncertainty and the fraction of industry CEOs who are insiders, consistent with insiders being more of a “known quantity.” Uncertainty is not significantly related to the CEO’s age, possibly because ability is match specific.

I also estimate the model in the sample’s largest four industries and in calendar year subsamples (Table 7). Pay is downward rigid in all subsamples. CEOs’ share of positive surpluses ranges from 18% in Business Equipment to 77% in Shops. CEOs capture just 19% of positive surpluses in financial firms. CEOs’ share of positive surpluses has dropped from 57% in 1992–1997 to 34% in 1998–2007, but the change is not statistically significant. Prior uncertainty ($\sigma_0$) has increased significantly over time, indicating the surpluses from learning have grown in magnitude.

The model fits the data well in all subsamples. The $J$-test cannot reject the model at the 10% level in any of the 26 subsamples.

To summarize, weak governance, as in Bertrand and Mullainathan (2001), does not seem to be driving the asymmetric response of pay to good and bad news. Downward rigid pay is pervasive. Some proxies for bargaining power are intuitively correlated with $\theta^p$, but other proxies have either a counterintuitive or insignificant correlation. One possible explanation
is that the proxies are noisy. Another is that several of these variables proxy not just for CEOs’ bargaining power but also for firms’ bargaining power, and these two effects offset each other.¹⁹ Results are quite different depending on whether I estimate in subsamples or control for multiple characteristics.

VII. Robustness

A. Gradual Vesting of CEO Compensation

In this subsection I show that the paper’s main conclusions are robust to using a different measure of CEO pay. So far, the measure has included stock- and option-based compensation in the year they are granted. These grants typically vest gradually over several years. If the firm and CEO renegotiate pay every year, then potentially only the pay vesting in year $t$ is relevant for the decision to continue the employment relationship in year $t$.²⁰ Following this logic, I create a new compensation measure that includes stock and option grants in the year they vest (details in Internet Appendix).

Table 3, row “Vesting measure,” shows results from estimating the model using this new measure. The paper’s main conclusions are robust: pay is downward rigid ($\theta^{\text{down}}$ is constrained at zero), and CEOs and shareholders both capture a large portion of positive surpluses. However, the CEO’s share of positive surpluses ($\theta^{\text{up}}$) has increased from 49% to 68%. Estimated prior uncertainty ($\sigma_0$) is also higher. The reason for these changes is that the 11th moment (the slope of changes in expected CEO pay on positive lagged excess

¹⁹For instance, CEOs in homogenous industries have many potential employers, but their firms also have many potential replacement CEOs.

²⁰The CEO may ignore pay vesting in $t + 1$ when deciding whether to stay in the firm in $t$, because the CEO anticipates that the firm will adjust salary and bonus in $t + 1$ to offset compensation vesting in $t + 1$. One argument for using the original, grant-based measure is that firms often compensate new CEOs for unvested shares and options they received from their previous employers (e.g., “Pay worries slow BofA search,” Wall Street Journal, Dec. 14, 2009.)
returns) has increased from 0.0023 to 0.0053.\textsuperscript{21} To fit this higher sensitivity while fitting the unchanged return volatility pattern, the model needs both larger surpluses (higher $\sigma_0$) and a larger share going to the CEO (higher $\theta^{up}$).

\section*{B. Endogenous CEO Dismissals}

The main model assumes CEOs are not fired after bad performance. There is considerable evidence to the contrary, extending from Coughlan and Schmidt (1985) through Jenter and Lewellen (2011). CEOs bear large, personal costs when they are fired: Fee and Hadlock (2004) find that dismissed CEOs who go on to run another firm end up at firms that are much smaller, which brings large decreases in compensation. While interesting and important for CEO incentives, these personal costs are unrelated to how CEOs and their current shareholders split the CEO’s surplus, which is this paper’s focus. Firing the CEO after bad performance does not create a negative surplus, according to this paper’s definition. As soon as the CEO and firm separate, the surplus the CEO brought to that firm ceases to exist, so we can no longer measure how it is shared.

Ignoring dismissals may nevertheless introduce estimation bias. I address this concern by estimating an extended model with endogenous firings. I follow the method of Taylor (2010), adding the following assumptions to the main model. The board of directors decides at the end of each year whether to fire or keep its CEO. An alternate interpretation is that shareholders decide each year whether the company should be acquired, in which event the CEO is replaced. The board of directors (in the first interpretation) or shareholders (in the second) maximize firm value. CEOs also retire each year with an exogenous, tenure-specific probability that I measure from the data.\textsuperscript{22} Replacing the CEO costs the firm a fraction $c$.

\textsuperscript{21}Part of the reason for the increase is that a high return last year will increase the value of shares granted in past years but vesting this year, which makes this new measure of pay more sensitive to positive lagged returns.

\textsuperscript{22}CEO turnover data for the Execucomp sample are from Jenter and Kanaan (2011) and Peters and Wagner (2009). CEO successions are classified as forced or voluntary according to Parrino’s (1997) rule.
of its assets.\textsuperscript{23} The board therefore faces a trade-off when deciding whether to fire a CEO with low perceived ability: firing the CEO allows the firm to hire a new CEO with uncertain but higher expected ability, but firing the CEO costs the firm $c$.

I numerically solve the board’s dynamic optimization problem. Details are in the Internet Appendix. The board optimally fires the CEO as soon as the cumulative CEO surplus captured by shareholders, which depends on perceived CEO ability and the \( \theta \) parameters, drops below an endogenous threshold. Stock prices now depend on the time-varying probability that the CEO is fired. I estimate this extended model by SMM. Identifying the turnover cost parameter $c$ requires an additional moment, namely, the average fraction of CEOs fired per year, which is 3.2 percent in this sample.

Parameter estimates are in Table 3, row labeled “With firings.”\textsuperscript{24} The main conclusions are unchanged: pay remains downward rigid (\( \theta_{\text{down}} \) is constrained at zero), and CEOs’ share of positive surpluses changes from 48.9% to just 49.4%.

\section*{C. Learning about Firm Quality}

So far the model has assumed that firm quality, denoted $a_i$ in equation (1), is constant and observable. This assumption implies that realized profitability is informative only about CEO ability, not about firm quality. I now relax this assumption, extending the model so that agents learn about firm quality and CEO ability at the same time. I show below that some parameter estimates change, but not the ones of interest.

\textsuperscript{23}Succession costs include severance or retirement pay, costs of searching for a new CEO, and any other costly disruption to the firm. Taylor (2010) splits the turnover cost into a cost to shareholders and a personal cost to directors. To keep this robustness exercise simple, I assume the firm, i.e. shareholders, bears the entire turnover cost.

\textsuperscript{24}The estimated turnover cost $c$ (not tabulated) is 10.75\% of assets (standard error 0.19\%). For comparison, Taylor (2010) estimates a turnover cost of 5.9\% of assets using a different model, dataset, and identification strategy. The main reason I find a higher turnover cost than Taylor (2010) is that I find more uncertainty about ability: \( \sigma_0 \) is 3.7\% of assets in this paper and 2.4\% of assets in Taylor (2010). A larger \( \sigma_0 \) means there is a bigger difference between a good and bad CEO and hence a stronger motive to fire bad CEOs. To continue fitting the low firing rate, the model offsets this stronger firing motive with a higher firing cost.
Firm profitability still follows equation (1), but now firm quality fluctuates over time according to

\[ a_{it} = \rho a_{it-1} + (1 - \rho) \overline{\pi}_i + u_{it}. \]  

(17)

Variables \( \eta_i \), \( a_{it} \), \( \varepsilon_{it} \), and \( u_{it} \) are all unobservable. I call uncertainty about \( a_{it} \) “firm uncertainty.” Shocks \( \varepsilon_{it} \) and \( u_{it} \) are normally distributed and uncorrelated. Agents learn about \( \eta_i \) (CEO ability) and \( a_{it} \) (firm quality) according to Bayes’ Rule, using information in realized profitability and the additional CEO signal, \( z_{it} \). High realized profitability now increases beliefs about both CEO ability and firm quality. All other model assumptions are the same as before. The Internet Appendix contains the solution. This extended model collapses to the main model when \( \sigma_u = 0 \), \( \rho = 0 \), and there is no prior uncertainty about firm quality.

This paper’s data offer little hope for identifying the parameters \( \sigma_u \) and \( \rho \) that govern learning about firm quality.\(^{25}\) Instead of estimating these additional parameters, I set them to values that imply a very high amount of firm uncertainty, then I estimate paper’s five main parameters using the same 12 moments as before. Specifically, I set \( \sigma_u \) to 10% per year and persistence \( \rho \) to 0.5.

Parameter estimates are in Table 3, row labeled “Learning about firm quality.” The main conclusions are unchanged: pay remains downward rigid (\( \theta^{down} \) is constrained at zero), and CEOs’ estimated share of positive surpluses changes from 48.9% to just 47.0%. The main parameter that changes is the estimated volatility of profitability shocks (\( \sigma_\varepsilon \)), which decreases from 36% to 30% when we add firm uncertainty. This change is expected: firm uncertainty now adds to stock return volatility, so we do not need such a high value of \( \sigma_\varepsilon \) to fit the observed level of return volatility.

\(^{25}\)The main challenge is disentangling \( \sigma_\varepsilon \) (the conditional volatility of shocks to realized profits) and \( \sigma_u \) (the volatility of shocks to expected profits).
D. Persistent Shocks to Profitability

In the main model, annual profitability $Y_{it}$ is hit by a firm-specific shock with volatility $\sigma_\varepsilon$. The estimate $\hat{\sigma}_\varepsilon = 36\%$ per year is unrealistically high (Section IV.B). In this section I show that we can rationalize this high estimate by reinterpreting the variable $Y_{it}$ (and hence $\sigma_\varepsilon$). I extend the model so that decisions and events in year $t$ may have a persistent effect on profitability. Persistence amplifies the relation between profitability shocks and returns, which in turn increases return volatility.

All model definitions and assumptions are the same as before, except the following. I now define $\pi_{it}$ to be year-$t$ profitability as a fraction of book assets. Profitability in year $t$ depends on actions and events during year $t$ and previous years. Specifically, $\pi_{it}$ depends on actions and events $x_{is\rightarrow t}$ that occurred and were observed in year $s \leq t$:

$$\pi_{it} \equiv \sum_{s=\infty}^{t} x_{is\rightarrow t}. \quad (18)$$

This extension allows a CEO’s actions during year $t$ to influence not just year-$t$ profits, but also profits in later years. Next, I re-define $Y_{it}$ to be the “value added” in year $t$, specifically, the present value of actions and events that occur in year $t$:

$$Y_{it} \equiv \sum_{\tau=0}^{\infty} \beta^\tau x_{it\rightarrow t+\tau}. \quad (19)$$

I assume $Y_{it}$ still follows equation (1), so that the value added to the firm in year $t$ is the sum of a firm fixed effect, CEO ability, an industry shock, and a firm-specific shock $\varepsilon_{it}$.

In the Internet Appendix I show that this version of the model makes predictions about excess stock returns and CEO pay that are identical to the main model’s predictions, albeit with a different interpretation of $Y_{it}$. Parameter estimates will therefore not change, but their interpretations will. In particular, $\sigma_\varepsilon$ is now the volatility of time-$t$ shocks to current plus discounted future profitability. Given this new interpretation, the 36% estimate of $\sigma_\varepsilon$
is not unreasonable.

VIII. Conclusion

I estimate a model in which agents learn gradually about a CEO’s ability, and the CEO and shareholders split the surplus resulting from a change in the CEO’s perceived ability. CEO pay responds asymmetrically to good and bad news about ability. The level of pay does not drop after bad news, implying the average CEO has downward rigid pay. This result is consistent with the model of Harris and Hölstrom (1982), in which firms optimally insure CEOs by offering a long-term contract with downward rigid pay. I find that offering downward rigid pay allows firms to pay risk-averse CEOs significantly less, on average. Following good news about CEO ability, the level of pay rises enough for the average CEO to capture roughly half of the positive surplus. This result implies that CEOs and firms have roughly equal bargaining power over these positive surpluses, on average. The asymmetric response is significantly stronger in firms with more institutional ownership, suggesting the result is not driven by weak governance.

This paper’s goal is to measure the surpluses from learning and how they are split. An important next step is to understand why surpluses are split the way they are. Section 6 begins to answer this question, but there is still important work to be done.

Also, this paper focuses on the level of CEO pay while abstracting from incentive compensation. Despite the large theoretical and empirical literature on incentive pay and optimal contracts, to my knowledge only Gayle and Miller (2009) and Page (2011) have estimated structural models of CEO incentive compensation. This is a fruitful area for future research.

Appendix

Variable Definitions

CEO yrs. inside firm: The number of years between Execucomp’s BECAMECEO (date became CEO) and JOINED_CO (date joined company). Winsorized at the 1st and 99th percentiles.

Fraction CEOs insiders: Fractions of CEOs within the data set and same Fama-French 49 industry that joined the company (JOINED_CO) less than 1 year before becoming CEO (BECAMECEO).

Industry homogeneity: The median, across Execucomp firms in the same Fama-French 49 industry, of the $R^2$ from time-series regressions of monthly stock returns on equal-weighted industry portfolio returns. Regressions use monthly return data from 1992-2007, exclude industry/month observations with fewer than 20 firms in the industry, and must have at least 30 monthly observations in the regression. Regressions and the industry portfolios only include firms in the Execucomp data, which contains primarily S&P1500 firms.

Number of similar firms: In the previous fiscal year, the number of Compustat firms in the same Fama-French 49 industry with assets within 20% of the given firm’s assets in a given year.

Outside directorships: The number of outside directorships held by the CEO in the previous fiscal year. The number of outside directorships is the number of firms in which the CEO appears in the Risk Metrics director database and is not classified as an employee of the firm. This variable is available starting in 1996.

Fraction pay unvested: The dollar value of unvested stock and option at the end of the previous year, as a fraction of the average total compensation (Execucomp’s TDC1) in the
previous four years. The value of unvested stock and option equals Execucomp’s “estimated value of in-the-money unexercised unexercisable options ($)” plus ”Restricted stock holdings ($)”.

CEO has explicit contract: Equals one if the CEO has an explicit employment agreement and equals zero otherwise. These data are available only for S&P 500 firms on January 1, 2000. To increase sample size I assume this variable is constant over all the years these CEOs were in office.

Institutional ownership: The fraction of shares owned by institutional investors. I compute this fraction of using Thomson Financial’s CDA/Spectrum Institutional (13-F) Holdings database. I set the fraction to one for less than 5% of observations in which the fraction exceeds one. Following Asquith, Pathek, and Ritter (2005), when the Thomson Financial database skips a quarter I impute shares owned by taking the minimum from the previous and next quarters. I impute a zero in 28% of observations in which the number of shares owned by institutional investors is missing.

CEO age in 1st year: The CEO’s age when he or she took office. Computed using Execucomp variables BECAMECEO (date the CEO took office), AGE (CEO’s current age), and YEAR.

Ln(firm age, in yrs): The natural log of the number of years since the firm first appeared in CRSP.

Estimating the Sensitivity of Parameters to Characteristics

This Appendix provides additional details on measuring $\frac{\partial \hat{\Theta}}{\partial Z_j}$, the sensitivity of parameter estimates to characteristic $Z_j$. First I describe how I measure $\frac{\partial \mathbf{M}}{\partial Z_j}$, the sensitivity of the 12 moments $\mathbf{M}$ to characteristic $Z_j$, holding all other characteristics constant. The 12 moments in vector $\mathbf{M} = \{M_i\}_{i=1}^{n_M}$ can be computed as the slopes $M_i$ from $n_M$ regressions of
the form

\[ Y_{im} = X_{im}'M_i + \delta_{im}, \quad i = 1, ..., n_m \quad (20) \]

where \( m \) indexes firm/year observations and \( X_{im} \) is \( k_i \times 1 \). I allow each moment to depend on an \( l \times 1 \) vector of firm, CEO, and industry characteristics \( Z_m \):

\[ M_{im}(Z_m) = [\Gamma_{i1} \ldots \Gamma_{ij} \ldots \Gamma_{il}]Z_m. \quad (21) \]

Each vector \( \Gamma_{ij} \) is \( k_i \times 1 \). From equation (21), \( \partial M/\partial Z_j = [\Gamma'_{ij} \ldots \Gamma'_{nMj}] \). I estimate the vectors \( \Gamma_{ij} \) in the OLS regression

\[ Y_{im} = (Z_m \otimes X_{im})' [\Gamma'_{i1} \ldots \Gamma'_{ij} \ldots \Gamma'_{il}]' + \delta_{im}. \quad (22) \]

The variance of \( \partial \hat{\Theta}/\partial Z_j \) comes from taking the variance of equation (16), which yields

\[ \text{var} \left( \frac{\partial \hat{\Theta}}{\partial Z_j} \right) = \frac{\partial \hat{\Theta}}{\partial M} \text{var} \left[ \Gamma'_{ij} \ldots \Gamma'_{nMj} \right] \frac{\partial \hat{\Theta}'}{\partial M}. \quad (23) \]

The OLS regression outputs the 12×12 matrix \( \text{var} \left[ \Gamma'_{1j} \ldots \Gamma'_{nMj} \right] \).
REFERENCES


Cremers, Martijn and Darius Palia, 2011, Tenure and CEO pay, Working paper, Yale School of Management.


Morellec, Erwan, Boris Nikolov, and Norman Schürhoff, 2011, Corporate governance and


Murphy, Kevin J. and Jan Zabojnik, 2007, Managerial capital and the market for CEOs. Unpublished working paper. Queen’s University, Kingston.


The solid line plots the combinations of parameters $\theta = \theta^{up} = \theta^{down}$ (the CEO’s fraction of the surplus) and $\sigma_0$ (prior uncertainty about CEO ability) that allow the model to match a given drop in excess stock return volatility. The dashed line plots the combinations of parameters $\theta$ and $\sigma_0$ that allow the model to match a given sensitivity of scaled changes in expected CEO pay to lagged excess stock returns.
This figure plots the relation between excess stock return volatility and CEO tenure. The vertical axis is the difference in the average variance of annual excess stock returns \( \text{var}_t(r_{it}) \) between CEO tenure \( \tau \) and tenure=1 year. The difference is mechanically zero for tenure equal one year. Actual moments with their 95% confidence interval are computed from a regression of excess return variance on tenure fixed effects (normalizing the fixed effect in year one to zero), log firm size, log firm age, and industry \( \times \) year fixed effects. The solid line plots the return variance predicted by the model, using main parameter estimates from Table 3.
Table 1: **Summary Statistics**

This table shows summary statistics for the sample of 4,545 CEOs. The measure of CEO pay is Execucomp’s TDC1, measured from 1992-2007. Scaled change in pay is pay in year $t$ minus year $t-1$ divided by the firm’s market cap at the beginning of year $t-1$. Excess annual return equals annual return minus industry return. Variance of returns is the annualized variance of weekly industry-adjusted returns. The remaining variables’ definitions are in the Appendix.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Stdev.</th>
<th>25th pctl.</th>
<th>Median</th>
<th>75th pctl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realized pay ($M)</td>
<td>20495</td>
<td>4.13</td>
<td>5.60</td>
<td>1.03</td>
<td>2.10</td>
<td>4.68</td>
</tr>
<tr>
<td>Scaled change in pay</td>
<td>16492</td>
<td>0.00029</td>
<td>0.00398</td>
<td>-0.00040</td>
<td>0.00000</td>
<td>0.00067</td>
</tr>
<tr>
<td>Excess annual return</td>
<td>20700</td>
<td>0.013</td>
<td>0.471</td>
<td>-0.245</td>
<td>-0.034</td>
<td>0.193</td>
</tr>
<tr>
<td>Variance of excess returns</td>
<td>20700</td>
<td>0.151</td>
<td>0.153</td>
<td>0.050</td>
<td>0.125</td>
<td>0.208</td>
</tr>
<tr>
<td>CEO tenure (years)</td>
<td>20700</td>
<td>7.9</td>
<td>7.4</td>
<td>3</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Market cap ($B)</td>
<td>20700</td>
<td>6.04</td>
<td>20.88</td>
<td>0.42</td>
<td>1.16</td>
<td>3.74</td>
</tr>
<tr>
<td>Assets ($B)</td>
<td>20700</td>
<td>10.85</td>
<td>53.43</td>
<td>0.40</td>
<td>1.27</td>
<td>4.85</td>
</tr>
<tr>
<td>CEO yrs. inside company</td>
<td>11567</td>
<td>8.6</td>
<td>10.4</td>
<td>0.0</td>
<td>3.5</td>
<td>15.8</td>
</tr>
<tr>
<td>Fraction of CEOs insiders</td>
<td>20695</td>
<td>0.59</td>
<td>0.13</td>
<td>0.50</td>
<td>0.61</td>
<td>0.73</td>
</tr>
<tr>
<td>Industry homogeneity</td>
<td>19481</td>
<td>0.265</td>
<td>0.082</td>
<td>0.210</td>
<td>0.250</td>
<td>0.330</td>
</tr>
<tr>
<td>Number of similar firms</td>
<td>20589</td>
<td>13</td>
<td>11</td>
<td>5</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Outside directorships</td>
<td>12073</td>
<td>0.54</td>
<td>0.86</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Fraction pay unvested</td>
<td>16118</td>
<td>0.91</td>
<td>1.10</td>
<td>0.06</td>
<td>0.49</td>
<td>1.29</td>
</tr>
<tr>
<td>CEO has explicit contract</td>
<td>3496</td>
<td>0.42</td>
<td>0.49</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Institutional ownership</td>
<td>20694</td>
<td>0.433</td>
<td>0.311</td>
<td>0.097</td>
<td>0.461</td>
<td>0.691</td>
</tr>
<tr>
<td>CEO age in first year</td>
<td>20509</td>
<td>48</td>
<td>8</td>
<td>43</td>
<td>49</td>
<td>54</td>
</tr>
<tr>
<td>Ln (firm age, in years)</td>
<td>20700</td>
<td>2.80</td>
<td>0.90</td>
<td>2.18</td>
<td>2.92</td>
<td>3.47</td>
</tr>
</tbody>
</table>
Table 2: **Moments from SMM Estimation**

Panel A shows the 12 moments used in SMM estimation. The actual moments are computed from the empirical sample of 4,545 Execucomp CEOs. Simulated moments are computed using the parameter estimates in Table 3, row “Main Results.” Empirical tenure fixed effects in return variance are estimated in a regression of realized return variance on tenure fixed effects, log lag firm assets, log firm age, and industry×year fixed effects. Predicted tenure fixed effects are computed from equation (IA.46) in the Internet Appendix. The sensitivity of changes in level of pay on lagged excess returns is from a regression of changes in realized CEO pay (scaled by lagged market cap) on contemporaneous and lagged excess stock returns (equation (11)). The \( J \)-test is the \( \chi^2 \) test for the model’s overidentifying restrictions. Panel B presents statistics on realized CEO pay, fifth year in office minus first year, divided by the firm’s assets at the beginning of year 1.

### A. Moments used in SMM estimation

<table>
<thead>
<tr>
<th>Actual moments</th>
<th>Simulated moments</th>
<th>( t )-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean excess return variance in CEOs’ first year</td>
<td>0.1709</td>
<td>0.1709</td>
</tr>
<tr>
<td>Tenure fixed effects in excess return variance:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenure = 2 years</td>
<td>-0.0178</td>
<td>-0.0184</td>
</tr>
<tr>
<td>Tenure = 3 years</td>
<td>-0.0228</td>
<td>-0.0221</td>
</tr>
<tr>
<td>Tenure = 4 years</td>
<td>-0.0229</td>
<td>-0.0236</td>
</tr>
<tr>
<td>Tenure = 5 years</td>
<td>-0.0274</td>
<td>-0.0244</td>
</tr>
<tr>
<td>Tenure = 6 years</td>
<td>-0.0228</td>
<td>-0.0249</td>
</tr>
<tr>
<td>Tenure = 7 years</td>
<td>-0.0244</td>
<td>-0.0252</td>
</tr>
<tr>
<td>Tenure = 8 years</td>
<td>-0.0267</td>
<td>-0.0254</td>
</tr>
<tr>
<td>Tenure = 9 years</td>
<td>-0.0246</td>
<td>-0.0255</td>
</tr>
<tr>
<td>Tenure = 10+ years</td>
<td>-0.0257</td>
<td>-0.0259</td>
</tr>
<tr>
<td>Sensitivity of changes in level of pay to lagged excess returns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive excess returns (( M^{(11)} ))</td>
<td>0.00233</td>
<td>0.00233</td>
</tr>
<tr>
<td>Negative excess returns (( M^{(12)} ))</td>
<td>0.00077</td>
<td>0.00077</td>
</tr>
</tbody>
</table>

| \( J \)-test | 1.91 |
| \( p \)-value | 0.96 |

### B. Scaled changes in realized CEO pay, years one to five

<table>
<thead>
<tr>
<th>Actual values</th>
<th>Simulated values</th>
</tr>
</thead>
<tbody>
<tr>
<td>25th percentile</td>
<td>-0.00019</td>
</tr>
<tr>
<td>50th percentile</td>
<td>0.00017</td>
</tr>
<tr>
<td>75th percentile</td>
<td>0.00100</td>
</tr>
</tbody>
</table>
Table 3: Parameter Estimates

This table shows parameter estimates for different specifications of the model. Section 7 describes the alternate specifications. All specifications use the grant-based measure of CEO pay, except for the “Vesting measure” specification, which measures pay in the year in which it vests rather than is granted. Specification “With firings” allows the firm to optimally fire CEOs. Specification “With firm uncertainty” includes learning about both CEO ability and firm quality. $\sigma_0$ is the standard deviation of prior beliefs about CEO ability. $\sigma_\epsilon$ is the volatility of shocks to firm profitability. Following good (bad) news about the CEO’s ability, the CEO captures a fraction $\theta^\text{up} (\theta^\text{down})$ of the resulting positive (negative) surplus. $\sigma_z$ is the volatility of the additional signal about CEO ability. Standard errors are in parentheses. The $J$-test is the $\chi^2$ test for the model’s overidentifying restrictions. All except “Main results” impose the constraint $\theta^\text{down} \geq 0$; standard errors are unavailable if the constraint binds.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Prior uncertainty $\sigma_0$</th>
<th>Volatility of profitability $\sigma_\epsilon$</th>
<th>CEO’s share of (+) surplus $\theta^\text{up}$</th>
<th>CEO’s share of (-) surplus $\theta^\text{down}$</th>
<th>Volatility of $z$ signal $\sigma_z$</th>
<th>$J$-test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main results</td>
<td>0.041</td>
<td>0.360</td>
<td>0.489</td>
<td>-0.051</td>
<td>0.033</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.108)</td>
<td>(0.045)</td>
<td>(0.009)</td>
<td>(0.965)</td>
</tr>
<tr>
<td>Constrained $\theta$</td>
<td>0.042</td>
<td>0.358</td>
<td>0.436</td>
<td>0.000</td>
<td>0.032</td>
<td>3.15</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.078)</td>
<td>N/A</td>
<td>(0.009)</td>
<td>(0.925)</td>
</tr>
<tr>
<td>Vesting measure</td>
<td>0.052</td>
<td>0.346</td>
<td>0.681</td>
<td>0.000</td>
<td>0.042</td>
<td>24.27</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.005)</td>
<td>(0.094)</td>
<td>N/A</td>
<td>(0.008)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>With firings</td>
<td>0.037</td>
<td>0.354</td>
<td>0.494</td>
<td>0.000</td>
<td>0.045</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.005)</td>
<td>(0.052)</td>
<td>N/A</td>
<td>(0.009)</td>
<td>(0.866)</td>
</tr>
<tr>
<td>Learning about firm quality</td>
<td>0.043</td>
<td>0.300</td>
<td>0.470</td>
<td>0.000</td>
<td>0.040</td>
<td>14.83</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.006)</td>
<td>(0.115)</td>
<td>N/A</td>
<td>(0.008)</td>
<td>(0.063)</td>
</tr>
</tbody>
</table>
Table 4: **The Value of Downward Rigid Pay**

Panel A presents statistics on realized pay simulated from the estimated model. Realized pay is scaled so that all CEOs’ average first-year pay is $2.84M, the empirical sample median for first-year CEOs in 2011 dollars. The net present cost of pay over years 1-5 is calculated by averaging over all realized pay paths and using an annual discount factor of 0.9. Panel B shows results from counterfactual simulations, which are the same as in Panel A but set $\phi^{\text{down}} = \hat{\phi}^{\text{up}} = 0.489$, and set average first-year pay to the level that makes the CEO indifferent between the paths in Panels A and B. The increase in net present cost is the ratio of net present cost in the counterfactual and actual models. Panel C describes empirical pay paths, where we scale all paths so that first-year pay is $2.84$, the sample median. Panel D describes counterfactual pay paths in which pay decreases as often as it increases, and first-year pay is set so that the CEO is indifferent between the paths in Panel C and D. CEOs in panels B and D have CRRA preferences over realized pay with the given risk aversion coefficient.

<table>
<thead>
<tr>
<th>A. Estimated model</th>
<th></th>
<th></th>
<th></th>
<th>Net present cost of pay, years 1-5 ($M)</th>
<th>Increase in net present cost (ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk aversion coefficient</td>
<td>Average 1st year pay ($M)</td>
<td>25th pct.</td>
<td>50th pct.</td>
<td>75th pct.</td>
<td>Net present cost of pay, years 1-5 ($M)</td>
</tr>
<tr>
<td>2.84</td>
<td>12.42</td>
<td>20.53</td>
<td>32.45</td>
<td>55.3</td>
<td>62.6</td>
</tr>
<tr>
<td>B. Counterfactual model without downward rigid pay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>13.1</td>
<td>0.1</td>
<td>13.0</td>
<td>31.9</td>
<td>62.6</td>
</tr>
<tr>
<td>1</td>
<td>21.1</td>
<td>2.0</td>
<td>21.0</td>
<td>39.9</td>
<td>86.3</td>
</tr>
<tr>
<td>2</td>
<td>55.1</td>
<td>36.1</td>
<td>55.1</td>
<td>74.0</td>
<td>203.5</td>
</tr>
<tr>
<td>4</td>
<td>113.8</td>
<td>94.7</td>
<td>113.7</td>
<td>132.6</td>
<td>419.0</td>
</tr>
<tr>
<td>C. Empirical distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.84</td>
<td>2.30</td>
<td>3.77</td>
<td>5.84</td>
<td>15.46</td>
<td>15.50</td>
</tr>
<tr>
<td>D. Counterfactual distribution without downward rigid pay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>3.58</td>
<td>1.32</td>
<td>3.58</td>
<td>5.84</td>
<td>15.50</td>
</tr>
<tr>
<td>1</td>
<td>3.92</td>
<td>1.66</td>
<td>3.92</td>
<td>6.18</td>
<td>16.7</td>
</tr>
<tr>
<td>2</td>
<td>7.49</td>
<td>5.23</td>
<td>7.49</td>
<td>9.75</td>
<td>29.2</td>
</tr>
<tr>
<td>4</td>
<td>33.6</td>
<td>31.3</td>
<td>33.6</td>
<td>35.8</td>
<td>124.6</td>
</tr>
</tbody>
</table>
Table 5: Sensitivity of Parameters to Firm, CEO, and Industry Characteristics
Panel A shows the parameter estimates from Table 3, row “Main Results.” Panel B shows each characteristic’s sample mean and standard deviation, and the change in parameter value associated with a one standard deviation increase in 11 CEO/firm/industry characteristics, holding other characteristics constant. \( t \)-statistics are in parentheses. These sensitivities are computed using equation (16), multiplying each sensitivity by the characteristic’s standard deviation. The Appendix provides detailed definitions of the 11 characteristics. Three characteristics (CEO yrs inside company, outside directorships, and CEO has explicit contract) are missing so often that I include them only in their own specification; hence 8 specifications control for 8 characteristics, and 3 specifications control for 9 characteristics.

A. Parameter Estimates (from Table 3)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean ( \sigma_0 )</th>
<th>Volatility of uncertainty ( \sigma_z )</th>
<th>Volatility of (+) surplus ( \theta_{up} )</th>
<th>Volatility of (-) surplus ( \theta_{down} )</th>
<th>( z ) signal change in parameter value associated with a one standard deviation increase in characteristic ( t )-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEO yrs. inside company</td>
<td>8.62 (10.36)</td>
<td>-0.009 (-1.62)</td>
<td>-0.020 (-4.44)</td>
<td>0.119 (0.70)</td>
<td>-0.073 (-1.12)</td>
</tr>
<tr>
<td>Fraction CEOs insiders</td>
<td>0.593 (0.133)</td>
<td>-0.008 (-2.02)</td>
<td>-0.045 (-14.06)</td>
<td>0.236 (1.98)</td>
<td>-0.017 (-0.38)</td>
</tr>
<tr>
<td>Industry homogeneity</td>
<td>0.265 (0.082)</td>
<td>0.007 (1.85)</td>
<td>0.000 (-0.02)</td>
<td>-0.198 (-1.73)</td>
<td>0.015 (0.34)</td>
</tr>
<tr>
<td>Number of similar firms</td>
<td>12.9 (11.4)</td>
<td>-0.002 (-0.45)</td>
<td>0.020 (6.3)</td>
<td>0.256 (2.18)</td>
<td>-0.126 (-3.01)</td>
</tr>
<tr>
<td>Outside directorships</td>
<td>0.540 (0.858)</td>
<td>0.003 (0.50)</td>
<td>-0.009 (-2.01)</td>
<td>0.025 (0.44)</td>
<td>0.025 (0.44)</td>
</tr>
<tr>
<td>Fraction pay unvested</td>
<td>0.907 (1.10)</td>
<td>-0.002 (-0.39)</td>
<td>0.006 (1.78)</td>
<td>0.147 (1.10)</td>
<td>-0.068 (-1.41)</td>
</tr>
<tr>
<td>CEO has explicit contract</td>
<td>0.420 (0.494)</td>
<td>0.009 (1.09)</td>
<td>-0.012 (-2.01)</td>
<td>0.044 (0.41)</td>
<td>-0.068 (-1.01)</td>
</tr>
<tr>
<td>Ln(firm assets)</td>
<td>0.41 (1.83)</td>
<td>-0.012 (-2.69)</td>
<td>-0.044 (-12.76)</td>
<td>0.053 (0.41)</td>
<td>-0.064 (-1.26)</td>
</tr>
<tr>
<td>Institutional ownership</td>
<td>0.433 (0.311)</td>
<td>-0.004 (-1.08)</td>
<td>0.001 (0.44)</td>
<td>0.113 (1.08)</td>
<td>-0.034 (-0.86)</td>
</tr>
<tr>
<td>CEO age in 1st year</td>
<td>48.0 (7.9)</td>
<td>-0.004 (-0.86)</td>
<td>-0.016 (-4.80)</td>
<td>0.028 (0.23)</td>
<td>-0.019 (-0.44)</td>
</tr>
<tr>
<td>Ln(firm age, in yrs)</td>
<td>2.80 (0.90)</td>
<td>-0.001 (-0.29)</td>
<td>-0.038 (-11.38)</td>
<td>0.074 (0.64)</td>
<td>-0.084 (-1.74)</td>
</tr>
</tbody>
</table>

B. Sensitivity of Parameter Values to Characteristics
Table 6: Estimation in Subsamples (I)

This table shows results from estimating the model in subsamples formed by splitting the full sample into two subsamples based on the median value of the indicated characteristic. The Appendix provides detailed definitions of the ten characteristics. I impose the constraints $\theta_{down} \geq 0$ and $\sigma_z \geq 0.01$; standard errors are unavailable if the constraint binds. Standard errors are in parentheses. The $J$-test is the $\chi^2$ test for the model’s overidentifying restrictions.

<table>
<thead>
<tr>
<th>Prior uncertainty</th>
<th>Volatility of profitability</th>
<th>CEO’s share of (+) surplus $\theta_{up}$</th>
<th>CEO’s share of (-) surplus $\theta_{down}$</th>
<th>Volatility of $z$ signal $\sigma_z$</th>
<th>$J$-test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less yrs. inside firm</td>
<td>0.0645</td>
<td>0.387</td>
<td>0.174</td>
<td>0.042</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.012)</td>
<td>(0.041)</td>
<td>(0.037)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>More yrs. inside firm</td>
<td>0.0345</td>
<td>0.326</td>
<td>0.614</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.010)</td>
<td>(0.141)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low fraction insiders</td>
<td>0.0443</td>
<td>0.414</td>
<td>0.605</td>
<td>0.000</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.154)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>High fraction insiders</td>
<td>0.0396</td>
<td>0.287</td>
<td>0.296</td>
<td>0.000</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.009)</td>
<td>(0.061)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low ind. homogeneity</td>
<td>0.0420</td>
<td>0.360</td>
<td>0.764</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.240)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>High ind. homogeneity</td>
<td>0.0488</td>
<td>0.356</td>
<td>0.224</td>
<td>0.000</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.008)</td>
<td>(0.045)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low # matching firms</td>
<td>0.0392</td>
<td>0.323</td>
<td>0.200</td>
<td>0.010</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.053)</td>
<td>(0.041)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>High # matching firms</td>
<td>0.0454</td>
<td>0.390</td>
<td>0.602</td>
<td>0.000</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.008)</td>
<td>(0.224)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Few directorships</td>
<td>0.0344</td>
<td>0.339</td>
<td>0.895</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.007)</td>
<td>(0.102)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>More directorships</td>
<td>0.0480</td>
<td>0.292</td>
<td>0.180</td>
<td>0.000</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.015)</td>
<td>(0.055)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low unvested wealth</td>
<td>0.0440</td>
<td>0.366</td>
<td>0.293</td>
<td>0.093</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.022)</td>
<td>(0.131)</td>
<td>(0.060)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>High unvested wealth</td>
<td>0.0439</td>
<td>0.369</td>
<td>0.355</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.028)</td>
<td>(0.131)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low firm assets</td>
<td>0.0500</td>
<td>0.427</td>
<td>0.765</td>
<td>0.000</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.007)</td>
<td>(0.214)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>High firm assets</td>
<td>0.0341</td>
<td>0.286</td>
<td>0.182</td>
<td>0.000</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.040)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low institutional own.</td>
<td>0.0454</td>
<td>0.364</td>
<td>0.360</td>
<td>0.000</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.008)</td>
<td>(0.067)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>High institutional own.</td>
<td>0.0369</td>
<td>0.353</td>
<td>0.745</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.006)</td>
<td>(0.148)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low age in 1st year</td>
<td>0.0460</td>
<td>0.388</td>
<td>0.455</td>
<td>0.000</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.009)</td>
<td>(0.137)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>High age in 1st year</td>
<td>0.0426</td>
<td>0.335</td>
<td>0.226</td>
<td>0.000</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.045)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Low firm age</td>
<td>0.0503</td>
<td>0.410</td>
<td>0.243</td>
<td>0.018</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.009)</td>
<td>(0.052)</td>
<td>(0.041)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>High firm age</td>
<td>0.0312</td>
<td>0.283</td>
<td>0.455</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.006)</td>
<td>(0.073)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 7: **Estimation in Subsamples (II)**
This table shows results from estimating the model in four Fama-French 12 industries and in two calendar year subsamples. Additional details are in Table 6.

<table>
<thead>
<tr>
<th>Industry/Year</th>
<th>Prior uncertainty $\sigma_0$</th>
<th>Volatility of profitability $\sigma_\epsilon$</th>
<th>CEO’s share of (+) surplus $\hat{\theta}_{up}$</th>
<th>CEO’s share of (-) surplus $\hat{\theta}_{down}$</th>
<th>Volatility of $z$ signal $\sigma_z$</th>
<th>$J$-test $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business equipment</td>
<td>0.0667</td>
<td>0.474</td>
<td>0.179</td>
<td>0.000</td>
<td>0.156</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td>(0.0095)</td>
<td>(0.013)</td>
<td>(0.088)</td>
<td>N/A</td>
<td>(0.086)</td>
<td>(0.958)</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.0352</td>
<td>0.322</td>
<td>0.351</td>
<td>0.131</td>
<td>0.010</td>
<td>8.58</td>
</tr>
<tr>
<td></td>
<td>(0.0029)</td>
<td>(0.015)</td>
<td>(0.143)</td>
<td>(0.104)</td>
<td>N/A</td>
<td>(0.379)</td>
</tr>
<tr>
<td>Shops</td>
<td>0.0506</td>
<td>0.368</td>
<td>0.768</td>
<td>0.000</td>
<td>0.027</td>
<td>11.57</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(0.012)</td>
<td>(0.146)</td>
<td>N/A</td>
<td>(0.026)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Finance</td>
<td>0.0368</td>
<td>0.274</td>
<td>0.193</td>
<td>0.000</td>
<td>0.029</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>(0.0050)</td>
<td>(0.018)</td>
<td>(0.096)</td>
<td>N/A</td>
<td>(0.018)</td>
<td>(0.991)</td>
</tr>
<tr>
<td>1992-1999</td>
<td>0.0382</td>
<td>0.326</td>
<td>0.574</td>
<td>0.000</td>
<td>0.028</td>
<td>5.72</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.008)</td>
<td>(0.171)</td>
<td>N/A</td>
<td>(0.012)</td>
<td>(0.678)</td>
</tr>
<tr>
<td>2000-2007</td>
<td>0.0457</td>
<td>0.382</td>
<td>0.342</td>
<td>0.000</td>
<td>0.038</td>
<td>3.22</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td>(0.008)</td>
<td>(0.050)</td>
<td>N/A</td>
<td>(0.012)</td>
<td>(0.920)</td>
</tr>
</tbody>
</table>
Internet Appendix for “CEO Wage Dynamics: Estimates from a Learning Model”

February 10, 2012

Contents

1 Model Solution
   1.1 Learning and first steps .................................................. 1
   1.2 Excess returns .............................................................. 2
   1.3 Return volatility ............................................................ 5
   1.4 CEO pay .................................................................. 8

2 Estimating the value of vesting options and restricted stock 9

3 Forecasting Final CEO Tenure $T_j$ 11

4 Details on Cleaning the Data 11

5 Model extension: Endogenous CEO turnover 12

6 Model extension: Learning about firm quality 15

7 Model extension: Persistent earnings shocks 17

8 CEO Tenure, Return Volatility, and the Variance of Profitability 19
1 Model Solution

1.1 Learning and first steps

First I solve the learning problem, which is a Kalman filtering problem. Since prior beliefs and signals are normally distributed, Bayes’ rule tells us that agents’ posterior beliefs about CEO ability will also be normally distributed. At the end of year $t$, agents’ beliefs are distributed as

$$\eta_i \sim N(m_{it}, \sigma^2_{\tau_{it}}),$$  \hspace{1cm} (IA.1)

where $\tau_{it}$ as the number of years completed by CEO of firm $i$ as of the end of year $t$. For simplicity I drop the subscripts on $\tau$. Agents update their beliefs about CEO ability by observing the mean-zero surprises in profitability and the additional signal:

$$\tilde{Y}_{it} = Y_{it} - a_i - v_t - m_{t-1} = \eta_i + \epsilon_{it} - m_{it-1}$$ \hspace{1cm} (IA.2)

$$\tilde{z}_{it} = z_{it} - m_{it-1}.$$ \hspace{1cm} (IA.3)

Applying Bayes’ rule, the posterior variance follows

$$\sigma^2_{\tau} = \sigma^2_0 \left( 1 + \tau \left( \frac{\sigma^2_0}{\sigma^2_\epsilon} + \frac{\sigma^2_\epsilon}{\sigma^2_z} \right) \right)^{-1},$$  \hspace{1cm} (IA.4)

which goes to zero in the limit where tenure $\tau$ becomes infinite. The posterior mean belief $m_{it}$ follows a martingale:

$$m_{it} = m_{it-1} + \frac{\sigma^2_{\tau}}{\sigma^2_\epsilon} \tilde{Y}_{it} + \frac{\sigma^2_{\tau}}{\sigma^2_z} \tilde{z}_{it}.$$  \hspace{1cm} (IA.5)

Next I solve for the changes in expected pay. From assumption 6, we have

$$\Delta E_t[w_{it}] = \theta_t B_{it} (m_{it-1} - m_{it-2}).$$ \hspace{1cm} (IA.6)

Substituting in equation (IA.5) yields

$$\Delta E_t[w_{it}] = \theta_t B_{it} \left( \frac{\sigma^2_{\tau-1}}{\sigma^2_\epsilon} \tilde{Y}_{it-1} + \frac{\sigma^2_{\tau-1}}{\sigma^2_z} \tilde{z}_{it-1} \right).$$ \hspace{1cm} (IA.7)

This equation relates changes in expected compensation to the previous year’s earnings surprise $\tilde{Y}_{it-1}$ and additional signal surprise $\tilde{z}_{it-1}$.

The dividend at the end of year $t$ equals profits minus CEO pay

$$D_{it} = B_{it} Y_{it} - w_{it},$$ \hspace{1cm} (IA.8)
and the firm’s value at the beginning of year \( t \) equals
\[
M_{it} = E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} D_{it+s} \right]. \tag{IA.9}
\]
From this equation I derive the firm’s stock return, the average industry return (which equals a constant plus \( v_t \)), and the stock return in excess of the industry.

### 1.2 Excess returns

**Prediction 1 (excess returns):** The excess stock return (firm minus industry) in year \( t \) equals
\[
r_{it} \approx \frac{B_{it}}{M_{it}} \tilde{Y}_{it} + \frac{B_{it}}{M_{it}} (1 - \theta_{t+1}) \beta \left( \frac{1 - \beta^{T-\tau}}{1 - \beta} \right) (m_{it} - m_{it-1}) - \text{median} (r_{it}). \tag{IA.10}
\]

This equation uses the approximation that the pay-performance sensitivity \( b_{it} \) is much less than the firm’s market value, which I confirm empirically.

**Proof:**

First I show that the industry shock to profitability, \( v_t \), is observable. I adjust profitability and average across the \( N_i \) firms \( k \) in firm \( i \)’s industry:
\[
\lim_{N_i \to \infty} \frac{1}{N_i} \sum_{k=1}^{N_i} (Y_{kt} - a_k - m_{kt-1}) = v_t + \lim_{N_i \to \infty} \frac{1}{N_i} \sum_{k=1}^{N_i} (\eta_k - m_{kt-1} + \varepsilon_{kt}) \tag{IA.11}
\]
\[
= v_t. \tag{IA.12}
\]
The model assumes agents know or can observe all quantities on the left-hand side, so it follows that they can also observe the right-hand side, which converges to the industry shock \( v_t \) since \( \eta_k - m_{kt-1} + \varepsilon_{kt} \) has mean zero.

In the remainder of this section I drop the firm subscript \( i \), for convenience. Also, since assets \( B_{it} \) are constant over time, I denote them \( B \).

The unexpected stock return is
\[
R_{t} - E_t [R_t] = M_{t}^{-1} (D_{t} - E_t [D_t] + M_{t+1} - E_t [M_{t+1}]) \tag{IA.13}
\]
The unexpected dividend is
\[
D_{t} - E_t [D_t] = B (\eta - m_{t-1} + v_t + \varepsilon_t) - (w_t - E_t [w_t]) \tag{IA.14}
\]
\[
= B (\tilde{Y}_{t} + v_t) - b_{t} r_t, \tag{IA.15}
\]
since (as I show later) the expected excess return $r_t$ equals zero, implying $w_t - E_t [w_t] = b_t r_t$.

Recalling from the learning results that

$$m_t = m_{t-1} + \frac{\sigma^2_x}{\sigma^2_z} \tilde{Y}_t + \frac{\sigma^2_x}{\sigma^2_z} \tilde{z}_t$$  \hspace{1cm} (IA.16)

we have

$$D_t - E_t [D_t] = B \left( \frac{\sigma^2_x}{\sigma^2_z} (m_t - m_{t-1}) - \frac{\sigma^2_x}{\sigma^2_z} \tilde{z}_t + v_t \right) - b_t r_t.$$  \hspace{1cm} (IA.17)

The surprise in future market value is

$$M_{t+1} - E_t [M_{t+1}] = E_{t+1} - E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} D_{t+1+s} \right]$$  \hspace{1cm} (IA.18)

$$= E_{t+1} - E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} (B \eta - w_{it+1+s}) \right].$$  \hspace{1cm} (IA.19)

because firm fixed effect $a_i$ is known (hence no change in its expected value), and shocks $\varepsilon$ and $\nu$ have conditional mean zero. Combining the results above yields

$$R_t - E_t [R_t] = M_t^{-1} \left( B \left( \tilde{Y}_t + v_t \right) - b_t r_t \right)$$  \hspace{1cm} (IA.20)

$$+ M_t^{-1} \left( E_{t+1} - E_t \left[ \sum_{s=1}^{\infty} \beta^s (B \eta - w_{it+s}) \right] \right).$$  \hspace{1cm} (IA.21)

The CEO’s last period is $T$, so there are $T - \tau_{t+1}$ periods left at the beginning of period $t + 1$. In periods $T + 1$ and later, a new CEO is in office. Before period $T$, agents learn nothing about this new CEO’s ability or his expected pay. Therefore we have

$$M_{t+1} - E_t [M_{t+1}] = E_{t+1} - E_t \left[ \sum_{s=0}^{T-\tau_{t+1}-1} \beta^{s+1} (B \eta - w_{it+1+s}) \right].$$  \hspace{1cm} (IA.22)

Decomposing into the two pieces and using fact that firm size and $\eta$ are constant over time,

$$M_{t+1} - E_t [M_{t+1}] = B \left( E_{t+1} - E_t [\eta] \right) \sum_{s=0}^{T-\tau-1} \beta^{s+1} - \sum_{s=0}^{T-\tau-1} \beta^{s+1} (E_{t+1} - E_t [w_{it+1+s}])$$

$$= B \left( m_{it} - m_{it-1} \right) \beta \left( \frac{1 - \beta^{T-\tau}}{1 - \beta} \right) - \sum_{s=0}^{T-\tau-1} \beta^{s+1} (E_{t+1} - E_t [w_{it+1+s}]).$$

Starting with $s = 0$, we want to know

$$E_{t+1} - E_t [w_{it+1}] = E_{t+1} [w_{it+1}] - E_t [E_{t+1} [w_{it+1}]].$$  \hspace{1cm} (IA.23)
Recall that $\theta_{t+1}$ is known at the beginning of period $t + 1$ but not at the beginning of $t$. We therefore need to treat $\theta_{t+1}$ as a random variable at time $t$. It is possible to show that

$$E_t [w_{t+1}] = \theta_{t+1} B (m_t - m_{t-1}) + B E_t [m_t - m_{t-1} | m_t - m_{t-1} > 0] \left( \frac{\theta^{\text{down}} - \theta^{\text{up}}}{2} \right).$$

(IA.24)

Using results for the truncated normal distribution, and denoting $\phi(0)$ the pdf of the standard normal distribution evaluated at zero, we have

$$E_t [w_{t+1}] = \theta_{t+1} B (m_t - m_{t-1}) + B \kappa(\tau).$$

(IA.25)

$$\kappa(\tau) \equiv (\text{Var}_t (m_{ijt} - m_{ijt-1}))^{1/2} \phi(0) (\theta^{\text{down}} - \theta^{\text{up}}).$$

Using backwards induction, it follows that

$$E_{t+1} - E_t [w_{t+1+s}] = E_{t+1} - E_t [w_{t+1}].$$

Plugging this result in, we have

$$M_{t+1} - E_t [M_{t+1}] = B (m_t - m_{t-1}) \beta \left( \frac{1 - \beta^{T-\tau}}{1 - \beta} \right) - (E_{t+1} - E_t [w_{t+1}]) \sum_{s=0}^{T-\tau-1} \beta^{s} \left( \frac{1 - \beta^{T-\tau}}{1 - \beta} \right) \kappa(\tau),$$

(IA.26)

and the firm’s stock return is

$$R_t = E [R_t] + \frac{B}{M_t} v_t - \frac{b_t}{M_t} r_t$$

(IA.28)

$$+ \frac{B}{M_t} (m_t - m_{t-1}) \left[ \sigma_z^2 + \beta \left( \frac{1 - \beta^{T-\tau}}{1 - \beta} \right) (1 - \theta_{t+1}) \right]$$

$$+ \frac{B}{M_t} \left( \frac{-\sigma_z^2 (z_t - m_{t-1})}{2} - \beta \left( \frac{1 - \beta^{T-\tau}}{1 - \beta} \right) \kappa(\tau) \right).$$

It is possible to show that the average of excess returns $r_t$ across industry firms goes to zero in the limit as the number of industry firms becomes infinite. Since all firms in the industry have the same assumed expected return $E [R]$, then the average realized industry return $\overline{R}_t$ equals

$$\overline{R}_t = E [R_t] + \left( \frac{B}{M_t} \right) v_t.$$
and the return in excess of the industry return equals\footnote{Here we use the approximation that the firm’s time-\(t\) book to market ratio, \(B/M_t\), approximately equals the industry average ratio, \((B/M_t)\). We also use the approximation \(b_{ijt} \ll M_{it}\).}

\[
    r_t \approx \frac{B}{M_t} \left[ \frac{\sigma_z^2}{\sigma^2} + \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) (1 - \theta_{t+1}) \right] (m_t - m_{t-1}) \tag{IA.29}
\]

\[
    - \frac{B}{M_t} \left( \frac{\sigma_z^2}{\sigma^2} (\bar{z}_t + \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) \kappa (\tau) \right).
\]

We can also write the excess return as

\[
    r_t \approx \frac{B}{M_t} \left( \left[ \frac{\sigma_z^2}{\sigma^2} + \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) (1 - \theta_{t+1}) \right] \frac{\sigma_z^2}{\sigma^2} \bar{Y}_t \right)
    + \frac{B}{M_t} \left( \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) (1 - \theta_{t+1}) \frac{\sigma_z^2}{\sigma^2} \bar{z}_t - \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) \kappa (\tau) \right)
\]

\[
    = \frac{B}{M_t} \left( \left[ 1 + \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) \frac{\sigma_z^2}{\sigma^2} (1 - \theta_{t+1}) \right] \bar{Y}_t \right)
    + \frac{B}{M_t} \left( \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) (1 - \theta_{t+1}) \frac{\sigma_z^2}{\sigma^2} \bar{z}_t - \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) \kappa (\tau) \right) \tag{IA.31}
\]

\[
    = \frac{B}{M_t} \left( \bar{Y}_t + \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) (1 - \theta_{t+1}) (m_t - m_{t-1}) - \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) \kappa (\tau) \right) \tag{IA.33}
\]

The various forms of this equation will be useful in various places later in the Appendix.

While \(\bar{Y}_t\) and \(\bar{z}_t\) are normally distributed with mean zero, the excess return \(r_t\) is not normally distributed, because \(\theta_{t+1}\) is a binary discrete random variable perfectly correlated with the sign of \(\bar{Y}_t \sigma_z^2 + \bar{z}_t \sigma_z^2\). The expected excess return is zero, by construction. The median of \(\bar{Y}_t \sigma_z^2 + \bar{z}_t \sigma_z^2\) is zero, and so is the median of \(\theta_{t+1} \left( \bar{Y}_t \sigma_z^2 + \bar{z}_t \sigma_z^2 \right)\). The median return is therefore

\[
    median \left( r_{ul} \mid T, \frac{B}{M_t}, \beta, \sigma, \sigma_0, \sigma_z, \theta^{down}, \theta^{up} \right) = - \frac{B}{M_t} \beta \left( \frac{1 - \beta^{T-t}}{1 - \beta} \right) \kappa (\tau), \tag{IA.34}
\]

which has the same sign as \(- (\theta^{down} - \theta^{up})\). Substituting this expression into the equations above yields Prediction 1. End of proof.

### 1.3 Return volatility

**Prediction 2 (excess return volatility):**
1. In the special case with no learning, i.e., $\sigma^2 = 0$, or in the limit when tenure goes to infinity, then the variance of excess stock returns equals

$$\text{var}_t (r_{it}) = \left( \frac{B_{it}}{M_{it}} \right)^2 \sigma^2_{\epsilon}$$.  

(IA.35)

2. In the special case in which $\theta^{up} = \theta^{down} = 1$, meaning the CEO captures the entire surplus from, then the variance equals

$$\text{var}_t (r_{it}) = \left( \frac{B_{it}}{M_{it}} \right)^2 \left( \sigma^2_{\tau - 1} + \sigma^2_{\epsilon} \right)$$,

(IA.36)

where $\sigma^2_{\tau - 1}$ is the uncertainty about CEO ability at the beginning of year $t$, given in equation (IA.4).

3. If prior uncertainty $\sigma^2 > 0$, the CEO’s share $\theta^{up} = \theta^{down} = \theta$, and $0 < \theta < 1$, then the variance of excess stock returns decreases with CEO tenure, increases with prior uncertainty $\sigma_0$, and decreases with the CEO’s share of the surplus $\theta$.

**Proof:** The proof uses the following result on asymmetric random normal variables. If $X$ is distributed as $N(0, \sigma)$ and

$$Y = \begin{cases} 
\theta_+ X & \text{if } X \geq 0 \\
\theta_- X & \text{if } X < 0,
\end{cases}$$

then the variance of $Y$ equals

$$\text{Var}(Y) = \sigma^2 \left[ \frac{\theta^2_+ + \theta^2_-}{2} - \phi(0)^2 (\theta_+ - \theta_-)^2 \right]$$.  

(IA.39)

I use equation (IA.33) to compute return volatility. Random variable $\theta_{t+1}$ depends on the sign of $(m_t - m_{t-1})$. I introduce notation that will come in handy soon:

$$\begin{align*}
\theta_{+/} & = \beta \left( \frac{1 - \beta^{T - \tau}}{1 - \beta} \right) (1 - \theta_{t+1}) \\
\theta_+ & = \beta \left( \frac{1 - \beta^{T - \tau}}{1 - \beta} \right) (1 - \theta^{up}) \\
\theta_- & = \beta \left( \frac{1 - \beta^{T - \tau}}{1 - \beta} \right) (1 - \theta^{down}),
\end{align*}$$

(IA.40)

$$\lambda(T - \tau; \theta^{up}, \theta^{down}, \beta) \equiv \frac{\theta^2_+ + \theta^2_-}{2} - \phi(0)^2 (\theta_+ - \theta_-)^2$$

(IA.43)
The variance of the second term in equation (IA.33) is therefore

\[
\left( \frac{B}{M_t} \right)^2 \left( \frac{\sigma_t^4}{\sigma_z^4} (\sigma_{\tau-1}^2 + \sigma_z^2) + \frac{\sigma_t^4}{\sigma_z^4} (\sigma_{\tau-1}^2 + \sigma_z^2) + 2 \frac{\sigma_t^4}{\sigma_z^4} \sigma_{\tau-1}^2 \right) \lambda (T - \tau; \theta^{up}, \theta^{down}, \beta) \tag{IA.44}
\]

The variance of the first term in equation (IA.33) above is

\[
\left( \frac{B_{it}}{M_{it}} \right)^2 (\sigma_{\tau-1}^2 + \sigma_z^2).
\]

The variance of returns therefore equals

\[
var_t (r_{it}) = \left( \frac{B}{M_t} \right)^2 (\sigma_{\tau-1}^2 + \sigma_z^2) + \left( \frac{B}{M_t} \right)^2 \left( \frac{\sigma_t^4}{\sigma_z^4} (\sigma_{\tau-1}^2 + \sigma_z^2) \right) \lambda (T - \tau; \theta^{up}, \theta^{down}, \beta) + \left( \frac{B}{M_t} \right)^2 \left( \frac{\sigma_t^4}{\sigma_z^4} (\sigma_{\tau-1}^2 + \sigma_z^2) + 2 \frac{\sigma_t^4}{\sigma_z^4} \sigma_{\tau-1}^2 \right) \lambda (T - \tau; \theta^{up}, \theta^{down}, \beta)
\]

\[
2 \frac{B}{M_t} \beta \left( \frac{1 - \beta^{T - \tau}}{1 - \beta} \right) \text{cov}_t (\hat{Y}_t, (1 - \theta_{t+1}) (m_t - m_{t-1})).
\]

It is possible to show that the covariance term above equals

\[
\text{cov}_t (\hat{Y}_t, \theta_{+/-} (Z_t)) = \frac{\theta_+ + \theta_-}{2} \left[ \frac{\sigma_t^4}{\sigma_z^4} (\sigma_{\tau-1}^2 + \sigma_z^2) + \frac{\sigma_t^4}{\sigma_z^4} \sigma_{\tau-1}^2 \right],
\]

so return volatility equals

\[
var_t (r_{it}) = \left( \frac{B}{M_t} \right)^2 (\sigma_{\tau-1}^2 + \sigma_z^2) + 2 \left( \frac{B}{M_t} \right)^2 \left( \frac{\theta_+ + \theta_-}{2} \left[ \frac{\sigma_t^4}{\sigma_z^4} (\sigma_{\tau-1}^2 + \sigma_z^2) + \frac{\sigma_t^4}{\sigma_z^4} \sigma_{\tau-1}^2 \right] \right) + \frac{2 B}{M_t} \beta \left( \frac{1 - \beta^{T - \tau}}{1 - \beta} \right) \text{cov}_t (\hat{Y}_t, (1 - \theta_{t+1}) (m_t - m_{t-1})).
\]

Claim: Holding constant \(T - \tau\) (number of years left in office) or setting \(T - \tau\) to infinity, then we have

\[
\lim_{\tau \to \infty} var_t (r_{it}) = \left( \frac{B_{it}}{M_{it}} \right)^2 \sigma_z^2.
\]

This results follows by noting

\[
\lim_{\tau \to \infty} \sigma_{\tau-1}^2 = 0.
\]
Claim: If \( \theta^{up} = \theta^{down} = \theta, \sigma_0^2 > 0, \) and \( 0 \leq \theta \leq 1, \) then \( \text{var}_{it}(r_{it}) \) decreases in tenure.

Proof: Posterior variance \( \sigma_{\tau}^2 \) is strictly positive and decreasing in \( \tau \) if \( \sigma_0^2 > 0. \) The quantity \( (1 - \beta^{T-\tau}) \) is also decreasing in \( \tau. \) The result follows by inspecting the expression for return variance, noting that quantities decreasing in \( \tau \) are multiplied by weakly positive quantities. Even if \( \theta = 1, \) the first term in equation (IA.46) still decreases in \( \tau. \)

Claim: If \( \theta^{up} = \theta^{down} = \theta, \) and \( 0 \leq \theta \leq 1, \) \( \text{var}_{it}(r_{it}) \) is increasing in \( \sigma_0^2. \)

Proof: Posterior variance \( \sigma_{\tau}^2 \) is increasing in \( \sigma_0^2. \) Terms multiplying \( \sigma_{\tau} \) in the expression above for return variance are all weakly positive, so the conclusion follows.

Claim: If \( \theta^{up} = \theta^{down} = \theta, \sigma_0 > 0, 0 < \theta < 1, \) then return variance is strictly decreasing in \( \theta. \)

Proof: Inspecting the expression for return variance, the term multiplying \( (1 - \theta) \) is positive, so the entire term is strictly decreasing in \( \theta. \)

1.4 CEO pay

Prediction 3 (CEO pay): The change in expected CEO compensation, scaled by the firm’s lagged market value, equals

\[
\frac{\Delta E_t[w_{it}]}{M_{it-1}} \approx \gamma r_{it-1} + \gamma B_{it} \frac{\sigma_0^2}{\sigma_{\tau}^2} \bar{z}_{it-1} + g(\cdot) \quad \text{(IA.49)}
\]

\[
\gamma(\tau, T; \beta, \sigma_{\epsilon}, \sigma_0, \theta_t) = \frac{\sigma_{\tau-1}^2 \theta_t}{\sigma_{\tau-1}^2 + \sigma_{\tau-1}^2 \beta \frac{1-\beta^{T-\tau+1}}{1-\beta} (1-\theta_t)} \quad \text{(IA.50)}
\]

where \( g \) is a deterministic function given in the Appendix.

Proof:

Using the last model assumption, we have

\[
\frac{\Delta E_t[w_{i}]}{M_{t-1}} = \theta_t \frac{B}{M_{t-1}} (m_{t-1} - m_{t-2})
\]

Rearranging equation (IA.29) yields

\[
(m_{t-1} - m_{t-2}) \approx r_{it-1} \frac{M_{t-1}}{B} + \beta \frac{1-\beta^{T-\tau+1}}{1-\beta} \kappa (\tau - 1) + \frac{\sigma_0^2 \bar{z}_{t-1}}{\sigma_{\tau-1}^2 + \beta \frac{1-\beta^{T-\tau+1}}{1-\beta} (1-\theta_t)},
\]
so

\[ \Delta E_t[w_t] \approx \theta_t B \frac{r_{t-1} M_{t-1}}{B_{t-1}} + \beta \left( \frac{1 - \beta^{T-t+1}}{1 - \beta} \right) \kappa (\tau - 1) + \frac{\sigma^2_{\epsilon}}{\sigma_{\epsilon}^2} \tilde{z}_{i,t-1} \] (IA.51)

\[ \frac{\Delta E_t[w_t]}{M_{t-1}} \approx \gamma r_{t-1} + \gamma B \frac{B_{t-1}}{M_{t-1}} \left( \frac{\sigma^2_{\epsilon}}{\sigma_{\epsilon}^2} \right) \tilde{z}_{i,t-1} + g(\cdot) \] (IA.52)

\[ \gamma (\tau, T; \beta, \sigma_{\epsilon}, \sigma_0, \theta_t) = \frac{\sigma^2_{\tau-1}}{\sigma_{\tau-1}^2 + \sigma^2_{\tau-1} \beta \left( \frac{1 - \beta^{T-t+1}}{1 - \beta} \right)} (1 - \theta_t) \] (IA.53)

\[ g \left( \tau, T; \beta, \sigma_{\epsilon}, \sigma_0, \theta_t, \frac{B}{M_{t-1}} \right) = \frac{B}{M_{t-1}} \kappa (\tau - 1) \beta \left( \frac{1 - \beta^{T-t+1}}{1 - \beta} \right) \gamma (\cdot) \] (IA.54)

Comparative statics for \( \gamma \): If there is no learning, i.e., \( \sigma_0 = 0 \), then \( \gamma = 0 \) since \( \sigma^2_{\tau-1} = 0 \). Also, in limit where \( \tau \) goes to infinity then we have \( \sigma^2_{\tau-1} = 0 \) and hence \( \gamma = 0 \). By inspection, for \( \sigma_0 > 0 \), slope \( \gamma \) is increasing in \( \theta_t \), independent of firm size \( M_{t-1} \) or \( B \), decreasing in signal noise \( \sigma^2_{\epsilon} \), increasing in initial uncertainty \( \sigma^2_0 \), and independent of the additional signal’s precision \( 1/\sigma_{\epsilon} \). It is straightforward to show that \( \gamma \) is decreasing in tenure.

2 Estimating the value of vesting options and restricted stock

This Appendix explains how I estimate \( vovest_{jt} \), the value of CEO \( j \)'s options that vest during year \( t \), and \( vsvest_t \), the value of a CEO shares that vest during year \( t \). The value of options vesting equals the number of options vesting \( (novest_t) \) times the price of each option vesting \( (pvest_t) \):

\[ vovest_t = novest_t \times pvest_t. \] (IA.55)

A similar formula applies to shares vesting:

\[ vsvest_t = nsvest_t \times psvest_t. \] (IA.56)

The number of options vesting during the year is

\[ novest_t = opt\_unex\_exer\_num_t \cdot \frac{ajex_t}{ajex_{t-1}} - opt\_unex\_exer\_num_{t-1} + opt\_exer\_num_t. \] (IA.57)

\( opt\_unex\_exer\_num_t \) is Execucomp’s number of unexercised exercisable options held by the CEO at the end of fiscal year \( t \). The ratio \( ajex_t/ajex_{t-1} \) (also Execucomp variables) adjusts for stock splits during year \( t \). \( opt\_exer\_num_t \) is Execucomp’s number of shares obtained upon exercising options during year \( t \). The explanation for the formula above is
that the CEO starts year $t$ with a supply of options $\text{opt}_{\text{unex}\_\text{exer}\_\text{num}}_{t-1}$ that have already vested but have not yet been exercised. An amount $\text{novest}_{jt}$ of new options vests, then the CEO gets rid of some of these options by exercising them ($\text{opt}_{\text{exer}\_\text{num}}_{t}$), so the CEO is left with a supply $\text{opt}_{\text{unex}\_\text{exer}\_\text{num}}_{t}$ of vested but unexercised options at the end of year $t$. The formula assumes that options are exercised before any stock splits occur. I set $\text{novest}$ equal to zero for fewer than 5% of observations that are negative.

The number of shares vesting during the year is given by

$$nsvest_t = stock\_unvest\_num_{t-1} - stock\_unvest\_num_{t} \frac{a_{ex}t}{a_{ex}t-1} + new\_granted\_num_{t}$$

(IA.58)

$stock\_unvest\_num_{t}$ is Execucomp’s number of shares of restricted stock held by the executive that had not yet vested by the end of year $t$. $new\_granted\_num_{t}$ is the number of new shares of restricted stock granted during the year, which I estimate by dividing the dollar value of newly granted options (Execucomp variable $rstkgrnt_{t}$ before 2006, $stock\_awards\_fv_{t}$ in 2006 and later) by $S_t$, the midpoint of the starting and ending share price for the year. To understand the formula for $nsvest_t$, the CEO starts with a supply of unvested shares at the beginning of the year ($stock\_unvest\_num_{t-1}$), then he or she receives some new shares ($new\_granted\_num_{t}$), then $nsvest_t$ shares vest, so the CEO is left with a supply $stock\_unvest\_num_{t}$ of unvested shares at the end of the year. I set $nsvest_t$ to zero if it takes a negative value. Since I do not know the exact date when the shares vest, I assume they vest at a share price $psvest_t$ midway between the starting and ending price for the year.

I estimate the price of the vesting options using the Black-Scholes formula, adjusted for dividends. I estimate the strike price $K_{t-1}$ for vesting options using the method of Core and Guay (2002), as described in Edmans, Gabaix, and Landier (2009):

$$K_{t-1} = S_t - \frac{\text{opt}_{\text{unex}\_\text{exer}\_\text{est}\_\text{val}}_{t-1}}{\text{opt}_{\text{unex}\_\text{exer}\_\text{num}}_{t-1}}.$$  

(IA.59)

$\text{opt}_{\text{unex}\_\text{exer}\_\text{est}\_\text{val}}_{t-1}$ is the Execucomp estimated value of unexercised exercisable options at the end of fiscal year $t - 1$. The dividend rate is Execucomp variable $bs\_yield$ measured at end of fiscal year $t$, divided by 100. I impute a zero if this variable is missing. I also winzorize this variable at the 95th percentile each year. Black-Scholes volatility is given by Execucomp variable $bs\_volatility$ at end of fiscal year $t$. If this variable is missing, I replace it with the year’s median value. I winzorize volatility at the 5th and 95th percentile each year. The risk free rate is the continuously compounded risk-free rate, derived from the one-month Treasury rate in July of year $t$. Following the method of Core and Guay
(2002) and Edmans, Gabaix, and Landier (2008), I set the average maturity of maturing options equal to the maturity of options granted during year $t$ (computed using Execucomp option maturity date, $exdate$), minus four years. If there were no new grants in year $t$ then I set $T_t = 5.5$ years. In the case of multiple new grants during year $t$, I take the longest maturity option. If maturity becomes negative then I set maturity equal to 1 day.

3 Forecasting Final CEO Tenure $T_j$

This section explains how I forecast $T_j$ (the total number of years CEO $j$ spends in office) for CEOs who have not left office by the end of the sample period. Forecasted $T_j$ equals the CEO’s tenure in his last record in Execucomp plus the forecasted number of years left in office, denoted $YearsLeft_{jt}$. The forecast is based on the following regression:

$$\log (1 + YearsLeft_{jt}) = \log a_0 + b_1 \log Age_{jt} + b_2 \log Tenure_{jt} + \varepsilon_{jt}. \quad (IA.60)$$

$Age_{jt}$ is CEO $j$’s age in year $t$ (Execucomp variable AGE). I estimate the regression by taking CEOs whose last year in office is in the database, and then creating one regression observation for each year the CEO spent in office, potentially including years before 1992. The regression uses 14,111 observations and has an $R^2$ value of 0.23. Forecasted $T_j$ is then

$$\hat{T}_j = Tenure_{jt^*} + \hat{a}_0 Age_{jt^*}^{\hat{b}_1} Tenure_{jt^*}^{\hat{b}_2} - 1 \quad (IA.61)$$

$$= Tenure_{jt^*} + e^{12.5 Age_{jt^*}^{2.75} Tenure_{jt^*}} - 1, \quad (IA.62)$$

where $t^*$ denotes CEO $j$’s last year in the database. $\hat{T}_j$ is missing if $Age_{jt}$ or $Tenure_{jt}$ is missing.

4 Details on Cleaning the Data

I clean the data as follows. First I fill in missing CEO indicators in Execucomp. I label an individual to be CEO in a firm/year observation if (i) Execucomp lists no one as CEO in the given firm/year, and (ii) either (a) this individual was CEO of the firm in previous and following year; (b) this individual was CEO in previous year, and we don’t know who was CEO in following year; or (c) this person was CEO in following year, and we don’t know who was CEO in previous year. I assume the CEO’s first fiscal year is the one when he completes at least 6 full months in office. I use Execucomp variable BECAMECEO as the date the CEO started in office. I exclude observations where BECAMECEO is missing.
Next I exclude all observations for those CEOs whose start date (BECAMECEO) is more than one year after their first yearly record as CEO in Execucomp; I assume these are data mistakes in Execucomp. Next I exclude firm/year observations where the CEO’s first fiscal year in office is less than 6 months long; I keep these CEOs’ later years in office. I cannot compute the vesting compensation measure in the CEO’s first year in Execucomp, because computing the value of shares and options vesting in year $t$ requires Execucomp data from year $t - 1$. Therefore, I cannot compute the change in this pay measure in a CEO’s first two years in Execucomp. In years when change in pay is missing for mechanical reasons, I keep the years’ stock return observation but treat the change in pay variable as missing. For other years, I delete firm/year records where change in pay is missing. I exclude firm/years where I cannot observe or forecast the CEO’s total tenure $T_j$. Next I exclude firm/years where I cannot find the firm’s lagged market cap in CRSP, and then I eliminate firm/years in which the variance of excess returns is missing.

5 Model extension: Endogenous CEO turnover

This robustness section extends the model to allow endogenous CEO turnover and then estimates the extended model. The extended model is similar to Taylor (2010). All the assumptions are the same as in the original model, but now we assume the board chooses whether or not to fire the CEO at the beginning of each year. Firing the CEO costs the firm a fraction $c$ of its assets. The board’s goal is to maximize firm value.

The firm’s dividend is now

$$D_{it} = Y_{it} - 1(fire_{it})B_ic - w_{it}$$

(IA.63)

where $1(fire_{it})$ is an indicator equal to 1 if the firm fires the CEO at the end of year $t$. The board makes CEO firing decisions that maximize firm value.

The board’s optimization problem is

$$\max_{\{fire_{it+j}\}_{j=0}^\infty} M_t,$$

(IA.64)

where $M_t$ is the firm’s market value at the beginning of year $t$, before the firing decision has
been made. We therefore have
\[
M_t^* = \max_{\{fire_{it+s}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} D_{it+s} \right]  
\]
\[
= \max_{\{fire_{it+s}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} (B_i Y_{it+s} - 1 \{fire_{it+s}\} B_i c - w_{it+s}) \right] \quad \text{(IA.66)}
\]
\[
= \max_{\{fire_{it+s}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} (B_i (a_i + \eta_i) - 1 \{fire_{it+s}\} B_i c - w_{it+s}) \right], \quad \text{(IA.67)}
\]
since shocks \(v_t\) and \(\varepsilon_t\) have mean zero. We therefore have
\[
\frac{M_t^*}{B_t} = a_i \frac{\beta}{1 - \beta} + \max_{\{fire_{it+s}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} \left( \eta_i - 1 \{fire_{it+s}\} c - \frac{w_{it+s}}{B_i} \right) \right]  
\]
\[
= a_i \frac{\beta}{1 - \beta} + \max_{\{fire_{it+s}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} \left( m_{it-1+s} - 1 \{fire_{it+s}\} c - \frac{E_{t+s} \{w_{it+s}\}}{B_i} \right) \right], \quad \text{(IA.69)}
\]

I define \(\Delta_t\) as the difference between posterior and the prior belief:
\[
m_{it} = m_{i0} + \Delta_{it}. \quad \text{(IA.70)}
\]
Also, I denote initial expected pay when the CEO enters office \(E [w_{i0}]\). Substituting in yields
\[
\frac{M_t^*}{B_t} = (a_i + m_{i0} - E [w_{i0}]) \frac{\beta}{1 - \beta} + \\
\max_{\{fire_{it+s}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} \left( \Delta_{it-1+s} - 1 \{fire_{it+s}\} c - \frac{E_{t+s} \{w_{it+s}\} - E [w_0]}{B_i} \right) \right]  
\]
\[
\frac{M_t^*}{B_t} = (a_i + m_{i0} - E [w_{i0}]) \frac{\beta}{1 - \beta} + V_t^* \quad \text{(IA.72)}
\]
where \(V_t^*\) is the value function:
\[
V_t^* = \beta \max_{fire_{it}} \left[ \Delta_{it-1} - \frac{E_t [w_{it}] - E [w_0]}{B_i} - 1 \{fire_{it}\} c \right] + \beta E_t [V_{t+1}^*]. \quad \text{(IA.73)}
\]
The state variable is
\[
x_{it} \equiv \Delta_{it-1} - \frac{E_t [w_{it}] - E [w_0]}{B_i}, \quad \text{(IA.74)}
\]
which equals zero when the CEO first starts in office. State variable \(x\) equals the cumulative CEO surplus captured by the shareholders, plus a constant. We know the dynamics for \(\Delta_{it}\) and \(E_t [w_{it}]\):
\[
\Delta_{it} = \Delta_{it-1} + (m_{it} - m_{it-1}) \quad \text{(IA.75)}
\]
\[
\frac{E_{t+1} [w_{it+1}]}{B_i} = \frac{E_t [w_{it-1}]}{B_i} + \theta_{t+1} (m_{it} - m_{it-1}) \quad \text{(IA.76)}
\]
so

\[ x_{it+1} = x_{it} + (m_{it} - m_{it-1}) (1 - \theta_{t+1}), \]  

(IA.77)

where

\[ m_{it} - m_{it-1} \sim N \left( 0, \sigma_{\tau-1}^2 \sigma_{\tau}^2 \left( \frac{1}{\sigma_{\varepsilon}^2} + \frac{1}{\sigma_{z}^2} \right) \right) \]  

(IA.78)

by Bayes’ Rule. Since the dynamics of \( x \) depend on tenure \( \tau \), \( \tau \) is also a state variable. I therefore write \( V^* \) as a function: \( V^*(x, \tau) \). If the board fires its CEO, it hires a new one so that the value function resets to

\[ V_{\text{fire}} = V(x = 0, \tau = 1) - c. \]  

(IA.79)

If the CEO retires voluntarily, then the value function resets to

\[ V_{\text{ret}} = V(x = 0, \tau = 1). \]  

(IA.80)

CEOs voluntarily retire after tenure year \( \tau \) with probability \( p_{\text{ret}}(\tau) \), estimated using data on voluntary successions as in Taylor (2010). If the firm chooses not to fire its CEO, then

\[
V_{\text{keep}} = \beta x_{it} + \beta E_t \left[ V_{\text{ret}}^* \right] \\
= \beta x_{it} + \beta \left( p_{\text{ret}} V_{\text{ret}} + (1 - p_{\text{ret}}) E_t [V(x_{it+1}, \tau+1)] \right). 
\]

(IA.81)

(IA.82)

The firm chooses whether to fire the CEO at time \( t \) according to

\[
V(x_{it}, \tau_t)^* = \max \{ V_{\text{fire}}, V_{\text{keep}} \} \\
= \max \{ V(0, 1) - c, \beta x_{it} + \beta \left( p_{\text{ret}} V_{\text{ret}} + (1 - p_{\text{ret}}) E_t [V(x_{it+1}, \tau+1)] \right) \}. 
\]

(IA.83)

(IA.84)

(IA.85)

Collecting results, the firm’s market value equals

\[
\frac{M_{it}^*}{B_i} = k + V^*(x_{it}, \tau_{it}) \\
k = (a_i + m_{i0} - E[w_{i0}]) \frac{\beta}{1 - \beta} \\
V^*(x_{it}, \tau_{it}) = \max \{ V(0, 1) - c, \beta x_{it} + \beta \left( p_{\text{ret}}^* V(0, 1) + (1 - p_{\text{ret}}^*) E_t [V(x_{it+1}, \tau+1)] \right) \} \\
V_{\text{ret}} = V^*(0, 1) \\
x_{it+1} = x_{it} + (m_{it} - m_{it-1}) (1 - \theta_{t+1}) \\
m_{it} - m_{it-1} \sim N \left( 0, \sigma_{\tau-1}^2 \sigma_{\tau}^2 \left( \frac{1}{\sigma_{\varepsilon}^2} + \frac{1}{\sigma_{z}^2} \right) \right) \\
\theta_{t+1} = \theta^{\text{up}} \text{ if } m_{it} - m_{it-1} > 0, \text{ otherwise } \theta_{t+1} = \theta^{\text{down}}. 
\]

(IA.86)

(IA.87)

(IA.88)

(IA.89)

(IA.90)

(IA.91)

(IA.92)

(IA.93)
In simulations I choose \( k \) so that the simulated market-to-book ratio equals its empirical counterpart.

I simulate returns using the following equations:

\[
R_{it} = \frac{D_{it} + M_{it+1}}{M_{it}} - 1 \tag{IA.94}
\]

\[
= \frac{Y_{it} - E_t[w_{ijt}] / B - br_{it}/B + M_{it+1}/B}{M_{it}/B} - 1 \tag{IA.95}
\]

\[
= \frac{Y_{it} - E_t[w_{ijt}] / B + k + V(x_{t+1}, \tau + 1)}{k + V(x_t, \tau)} - 1. \tag{IA.96}
\]

Unexpected returns equal

\[
R_{it} = R_{it} - E_t[R_{it}] \tag{IA.97}
\]

\[
= \frac{Y_{it} + V(x_{t+1}, \tau + 1) - E[V(x_{t+1}, \tau + 1)|x_t, \tau]}{k + V(x_t, \tau)}. \tag{IA.98}
\]

6 Model extension: Learning about firm quality

I make the following changes in notation. For convenience I drop subscripts on several variables. \( \hat{a}_{t|s} \) and \( \hat{\eta}_{t|s} \) denotes the posterior mean of \( a_{it} \) and \( \eta_{it} \), respectively, at the end of period \( s \). Therefore, \( \hat{\eta}_{t|s} = m_{is} \) from the original notation. \( \Sigma_{a_{t|s}} \) and \( \Sigma_{\eta_{t|s}} \) are the posterior variance of beliefs about \( a_{it} \) and \( \eta_{it} \), respectively, at the end of period \( s \). I drop firm subscripts \( i \) for convenience.

I write the problem in vector form to apply the multivariate version of Bayes’ rule. State variable \( x_t \equiv [a_t \eta] \) follows (as long as CEO stays in office)

\[
x_t = \Phi x_{t-1} + (I - \Phi) \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \end{pmatrix} \tag{IA.99}
\]

\[
\Phi = \begin{pmatrix} \rho & 0 \\ 0 & 1 \end{pmatrix}. \tag{IA.100}
\]

Beliefs about \( x_t \) at the end of period \( t - 1 \) are distributed as \( N(\mu_{t|t-1}, \Omega_{t|t-1}) \), and beliefs about \( x_t \) at end of period \( t \) are distributed as \( N(\mu_{t|t}, \Omega_{t|t}) \). From the law of motion for \( x \) we
can immediately write

\[ \mu_{t|t-1} = \Phi \mu_{t-1|t-1} + (I - \Phi) \left( \begin{array}{c} \bar{\sigma} \\ 0 \end{array} \right) \quad \text{(IA.101)} \]

\[ \Omega_{t|t-1} = \Phi' \Omega_{t-1|t-1} \Phi + \left( \begin{array}{cc} \sigma_u^2 & 0 \\ 0 & 0 \end{array} \right) \cdot \text{(IA.102)} \]

When a new CEO takes office at the beginning of period \( t \) we set the off-diagonal elements of \( \Omega_{t|t-1} \) to zero and the diagonal element corresponding to \( \eta \) to \( \sigma_0^2 \), meaning that uncertainty about the CEO resets to the prior uncertainty while uncertainty about firm quality keeps its current value. The signal observed each period is

\[ X_t \equiv \begin{pmatrix} Y_t - v_t \\ z_t \end{pmatrix} = \Theta x_t + \varepsilon_t \quad \text{(IA.103)} \]

\[ \Theta = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_t \\ \delta_{zt} \end{pmatrix} \sim N(0, \Sigma) \quad \text{(IA.104)} \]

\[ \Sigma = \begin{pmatrix} \sigma_{\varepsilon}^2 & 0 \\ 0 & \sigma_{z}^2 \end{pmatrix} \quad \text{(IA.105)} \]

Bayes rule states that

\[ \mu_{t | t} = \Omega_{t | t} \left( \Omega_{t | t-1}^{-1} \mu_{t | t-1} + \Theta' \Sigma^{-1} X_t \right) \quad \text{(IA.106)} \]

\[ \Omega_{t | t} = \left[ \Omega_{t | t-1}^{-1} + \Theta' \Sigma^{-1} \Theta \right]^{-1} \quad \text{(IA.107)} \]

Next I provide an expression for excess stock returns. In the original model, excess returns are given in equation (\( \text{ eq}\{\text{eq_r_ztilde}\} \)). The only difference in this model is that the firm’s market value moves due to changes in beliefs about \( a_t \), firm quality. The contribution of firm quality to market value at the beginning of period \( t \) is

\[ B \beta \sum_{s=0}^{\infty} \beta^s E_t [a_{t+s}] \quad \text{(IA.108)} \]

It is straightforward to show that

\[ E_t [a_{t+s}] = \rho^{s+1} \tilde{a}_{t-1|t-1} + \bar{\alpha} \left( 1 - \rho^{s+1} \right) \quad \text{(IA.109)} \]

hence

\[ B \beta \sum_{s=0}^{\infty} \beta^s E_t [a_{t+s}] = B \beta \sum_{s=0}^{\infty} \beta^s \left( \rho^{s+1} \tilde{a}_{t-1|t-1} + \bar{\alpha} \left( 1 - \rho^{s+1} \right) \right) \quad \text{(IA.110)} \]

\[ = B \beta \sum_{s=0}^{\infty} \beta^s \left( \rho^{s+1} \left( \tilde{a}_{t-1|t-1} - \bar{\alpha} \right) + \bar{\alpha} \right) \quad \text{(IA.111)} \]
$$B\beta \sum_{s=0}^{\infty} \beta^s E_t [a_{t+s}] = B\beta \left[ \frac{\bar{a}}{1 - \beta} + \rho \frac{\hat{a}_{t-1|t-1} - \bar{a}}{1 - \beta \rho} \right]. \quad \text{(IA.112)}$$

The change in this contribution from the end of period $t$ to the end of period $t-1$ is

$$B\beta \rho \frac{\hat{a}_{t|t} - \hat{a}_{t-1|t-1}}{1 - \beta \rho}. \quad \text{(IA.113)}$$

The unexpected change in the contribution is

$$B\beta \rho \frac{\hat{a}_{t|t} - E_t [\hat{a}_{t|t}]}{1 - \beta \rho}, \quad \text{(IA.114)}$$

where $E_t$ denotes expectations conditioning on information known at the beginning of period $t$. Using equations (\ref{eq_mu_dynamics1})-(\ref{eq_Omega_dynamics}) one can show that the unexpected change in contribution to market value is

$$B\beta \rho \frac{\hat{a}_{t|t} - \rho \hat{a}_{t-1|t-1} - (1 - \rho) \bar{\pi}}{1 - \beta \rho}. \quad \text{(IA.115)}$$

Therefore, the excess stock return in year $t$ is given by equation (\ref{eq_r_ztiled}) plus the following term:

$$\frac{B_{it}}{M_{it}} \beta \rho \frac{\hat{a}_{t|t} - \rho \hat{a}_{t-1|t-1} - (1 - \rho) \bar{\pi}}{1 - \beta \rho} \quad \text{(IA.116)}$$

that comes from learning about firm quality.

I obtain predicted moments by first simulating values of state variable $x_t$, then simulating values of the signals $X_t$, updating beliefs according to the equations above, computing excess returns, and then taking the variance of simulated returns. I begin simulations with the variance of $a_t$ at its long-run value.

## 7 Model extension: Persistent earnings shocks

This Appendix proves the following claim: Given the definitions of $x$, $\pi$, and $Y$ in robustness Section ?, the extended model’s predictions are identical to those in the main model.

The firm’s unexpected return is

$$R_t - E_t [R_t] = M_t^{-1} (D_t - E_t [D_t] + M_{t+1} - E_t [M_{t+1}]) \quad \text{(IA.117)}$$

The unexpected dividend equals

$$D_t - E_t [D_t] = B (\pi_{it} - E_t [\pi_{it}]) - (w_t - E_t [w_t]) = B \left( \sum_{s=-\infty}^{t} x_{is-t} - E_t \left[ \sum_{s=-\infty}^{t} x_{is-t} \right] \right) - b_t r_t.$$
Since contributions made at $s < t$ are known at time $t$, we have

$$D_t - E_t[D_t] = B (x_{it\rightarrow t} - E_t[x_{it\rightarrow t}]) - b_t r_t.$$  

The surprise in future market value is

$$M_{t+1} - E_t[M_{t+1}] = E_{t+1} - E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} D_{t+1+s} \right] = E_{t+1} - E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} (B \pi_{it+1+s} - w_{it+1+s}) \right] = E_{t+1} - E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} \left( B \sum_{\tau=-\infty}^{t+1+s} x_{i\tau\rightarrow t+1+s} - w_{it+1+s} \right) \right].$$

(IA.118)  

(IA.119)  

(IA.120)  

(IA.121)

Since beliefs about past contributions $x_{is\rightarrow t}$, $s < t$, do not change during period $t$, we have

$$M_{t+1} - E_t[M_{t+1}] = E_{t+1} - E_t \left[ \sum_{s=0}^{\infty} \beta^{s+1} \left( B \sum_{\tau=t}^{t+1+s} x_{i\tau\rightarrow t+1+s} - w_{it+1+s} \right) \right].$$

Since

$$\sum_{s=0}^{\infty} \beta^{s+1} \left( \sum_{\tau=t}^{t+1+s} x_{i\tau\rightarrow t+1+s} \right) = -x_{it\rightarrow t} + \sum_{s=0}^{\infty} \beta^s \sum_{\tau=0}^{\infty} \beta^\tau x_{it+s\rightarrow it+s+\tau}$$

$$= -x_{it\rightarrow t} + \sum_{s=0}^{\infty} \beta^s Y_{it+s} = -x_{it\rightarrow t} + Y_{it} + \sum_{s=1}^{\infty} \beta^s Y_{it+s}$$

we have

$$M_{t+1} - E_t[M_{t+1}] = E_{t+1} - E_t \left[ B \left( -x_{it\rightarrow t} + Y_{it} + \sum_{s=1}^{\infty} \beta^s Y_{it+s} \right) - \sum_{s=1}^{\infty} \beta^s w_{it+s} \right]$$

$$= -B (x_{it\rightarrow t} - E_t[x_{it\rightarrow t}]) + Y_{it} - E_t[Y_{it}] + E_{t+1} - E_t \left[ \sum_{s=1}^{\infty} \beta^s (BY_{it+s} - w_{it+s}) \right]$$

so

$$R_t - E_t[R_t] = M_t^{-1} (B (Y_{it} - E_t[Y_{it}]) - b_t r_t) + M_t^{-1} \left( E_{t+1} - E_t \left[ \sum_{s=1}^{\infty} \beta^s (BY_{it+s} - w_{it+s}) \right] \right).$$
Substituting in equation (1), the assumption that the firm fixed effect is known, and the industry shock is i.i.d. with mean zero, we have

\[ R_t - E_t[R_t] = M_t^{-1} \left( B \left( \bar{Y}_t + v_t \right) - b_t r_t \right) + M_t^{-1} \left( E_{t+1} - E_t \left[ \sum_{s=1}^{\infty} \beta^s (B \eta - w_{it+s}) \right] \right). \]

This equation is identical to equation \ref{eq_app_connect_exvol} in Appendix ??, so the predictions about excess returns will be identical as in the main model. Since the equation for \( Y_{it} \) has not changed, the equations for learning dynamics will not change, and neither will the equations for wage dynamics.

### 8 CEO Tenure, Return Volatility, and the Variance of Profitability

Figure 2 in the main paper shows that excess stock return volatility declines after a new CEO takes office. The model attributes this decline to learning about CEO ability. In this section I test an alternate explanation, which is that earnings volatility declines with CEO tenure. First I estimate the shocks to profitability, then I check whether the volatility of these shocks changes with CEO tenure. For comparison, I confirm that return volatility declines with tenure even after including additional controls.

I compute annual return on assets (ROA) for every firm/year in the sample. I estimate earnings shocks \( \varepsilon_{it} \) using the following panel model:

\[ ROA_{it} = \beta_0 + \beta_1 ROA_{it-1} + \beta_2 \log(\text{Assets}_{it-1}) + \beta_i + \beta_t + \beta_{\tau} + \varepsilon_{it}, \quad (IA.122) \]

where \( \beta_i \) is a firm fixed effect, \( \beta_t \) is a year fixed effect, and \( \beta_{\tau} \) is a CEO tenure fixed effect for tenure categories \( \tau = 1, \ldots, 10+ \) years. The conditional mean of the squared residuals, \( E[\varepsilon_{it}^2|\text{regressors}] \), equals the conditional variance of profitability. I estimate this conditional variance from the following regression:

\[ \varepsilon_{it}^2 = \gamma_0 + \gamma_1 \log(\text{Assets}_{it-1}) + \gamma_i + \gamma_t + \gamma_{\tau} + u_{it}, \quad (IA.123) \]

where \( \varepsilon_{it}^2 \) is estimated from regression (IA.122), \( \gamma_i \) is a firm fixed effect, \( \gamma_t \) is a year fixed effect, and \( \gamma_{\tau} \) is a CEO tenure fixed effect.

Table 1 shows the estimated tenure fixed effects \( \gamma_{\tau} \) for the conditional variance of ROA. The fixed effect for tenure = 10+ years is normalized to zero. None of the tenure fixed
effects is significantly different from zero. The conditional variance of profitability shows no significant pattern with CEO tenure.

For comparison, I measure tenure fixed effects in excess return volatility. I regress RETVAR (the annualized variance of excess stock returns) on log lag assets, firm fixed effects, year fixed effects, tenure fixed effects, and (in one specification) the squared shocks to ROA ($\bar{\varepsilon}_{it}^2$). Results are in Table 1. The fixed effect for tenure equal one (two) years is significantly positive at the one (ten) percent confidence level, and the remaining fixed effects are indistinguishable from zero, consistent with the result in Figure 3. In sum, return volatility declines significantly with tenure, but earnings volatility does not.
Table 1: **CEO Tenure, Return Volatility, and the Variance of Profitability**

This table shows the variance of firm profitability and excess stock returns, conditional on CEO tenure and other controls. The variance for CEOs with tenure = 10+ is normalized to zero. First I estimate shocks to return on assets (ROA) by regressing ROA on its lag, log(lag assets), firm fixed effects, year fixed effects, and CEO tenure fixed effects (results not shown). I then square the estimated residuals and regress these on log lag assets, firm fixed effects, year fixed effects, and tenure fixed effects; estimates are below. The table also shows the tenure fixed effects from a regression of RETVAR (annualized variance of excess stock returns) on log lag assets, firm fixed effects, year fixed effects, tenure fixed effects, and the squared shock to ROA. The sample contains is described in section IV.B. Standard errors are in parentheses.

<table>
<thead>
<tr>
<th>CEO tenure (yrs)</th>
<th>Squared ROA shock</th>
<th>Variance of excess returns</th>
<th>Variance of excess returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0009</td>
<td>0.0264</td>
<td>0.0263</td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td>(0.0033)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>2</td>
<td>0.0025</td>
<td>0.0060</td>
<td>0.0050</td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td>(0.0034)</td>
<td>(0.0034)</td>
</tr>
<tr>
<td>3</td>
<td>0.0009</td>
<td>0.0034</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0035)</td>
<td>(0.0035)</td>
</tr>
<tr>
<td>4</td>
<td>-0.0012</td>
<td>0.0027</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>(0.0034)</td>
<td>(0.0036)</td>
<td>(0.0036)</td>
</tr>
<tr>
<td>5</td>
<td>0.0001</td>
<td>-0.0001</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0037)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>6</td>
<td>-0.0015</td>
<td>0.0038</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>(0.0036)</td>
<td>(0.0038)</td>
<td>(0.0038)</td>
</tr>
<tr>
<td>7</td>
<td>-0.0024</td>
<td>-0.0003</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.0040)</td>
<td>(0.0040)</td>
</tr>
<tr>
<td>8</td>
<td>-0.0027</td>
<td>-0.0017</td>
<td>-0.0014</td>
</tr>
<tr>
<td></td>
<td>(0.0039)</td>
<td>(0.0041)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td>9</td>
<td>-0.0028</td>
<td>-0.0025</td>
<td>-0.0022</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0043)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>10+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>log(lag assets)</td>
<td>0.0044</td>
<td>-0.0189</td>
<td>-0.0183</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.0019)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Squared ROA shock</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0584</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0079)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>20,400</td>
<td>20,482</td>
<td>20,400</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.183</td>
<td>0.612</td>
<td>0.614</td>
</tr>
</tbody>
</table>