CEO Wage Dynamics: Evidence from a Learning Model

October 1, 2010

Abstract: Good news about a CEO's ability creates a positive surplus. Empirically, CEOs capture most of this surplus by bidding up their pay. However, CEOs bear almost none of the negative surplus resulting from bad news about ability. These results are consistent with the optimal contracting benchmark of Harris and Hölmstrom (1982). Since CEOs do not capture their entire surplus, CEO ability matters more for shareholders, which is supported by predictions and data on unanticipated CEO deaths. The model helps explain the sensitivity of CEO pay to lagged returns, and also the changes in return volatility around CEO successions.

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

Keywords: CEO, compensation, learning, SMM

Note: The Internet Appendix is at the end of this document.

I. Introduction

This paper measures how a CEO's compensation changes over time due to learning about the CEO's ability. News about a CEO's ability creates a surplus. How the CEO and shareholders split this surplus depends on their relative bargaining power, which in turn depends on their outside options in the labor market and contractual constraints. For instance, what is the next best job for the CEO, and how costly is switching jobs? Who is the next best CEO for the firm, and how costly is switching CEOs? Does the CEO have a long-term compensation contract, and if so, how costly is renegotiating it? By measuring bargaining outcomes, this paper contributes to our understanding of the CEO labor market, CEOs' bargaining power, and the nature of CEO compensation contracts. Also, the paper quantifies the effects of bargaining and wage dynamics on shareholder value.

The approach in this paper is to estimate a dynamic learning model. In the model, a firm's profitability depends on unobservable CEO ability, a constant firm-specific component, industry shocks, and firm-specific shocks. There is no asymmetric information. The CEO and shareholders start with prior beliefs about the CEO's ability, and they use Bayes' rule to update their beliefs each year after observing realized profits and an additional, latent signal. The firm's stock returns depend endogenously on beliefs about CEO ability. A change in beliefs creates a surplus, defined as the change in the firm's expected annual profits. The CEO and shareholders split this surplus by adjusting the level of CEO pay. If the level of pay does not fully adjust, then shareholders capture some of the surplus via a higher or lower share price. Separate parameters control how positive and negative surpluses are split.

The model nests the predictions of Harris and Hölmstrom (1982), who provide an optimal-contracting benchmark when there is learning about the ability of a worker— in the current paper, a CEO. Harris and Hölmstrom show that it is optimal for the firm to insure the risk-averse CEO against changes in perceived ability. The firm does so by offering a long-term compensation contract with downward rigid wages. This contract guarantees the CEO's expected pay never drops. As a result, shareholders bear the costs of bad news about the CEO's ability, ex post. In contrast, the CEO/worker in their model has enough bargaining power to capture all the benefits from good news. The reason is that the CEO can threaten to leave and work at an identical firm after good news arrives, so the CEO can renegotiate the long-term contract and obtain a higher level of pay. In sum, Harris and Hölmstrom (1982) predict CEOs capture 100% of a positive surplus and none of a negative surplus. I measure how close the data come to their benchmark, and also to the related models of Jovanovic (1979), Gibbons and Murphy (1992), and Hölmstrom (1999).

I highlight two predictions that help explain how I estimate the model. For certain parameter values, the model predicts that changes in expected CEO pay are positively correlated with lagged excess stock returns.[1] This prediction obtains because abnormally high profits generate a high abnormal stock return and also an increase in the CEO's perceived ability, which in turn causes expected CEO pay to increase in the next year (and vice versa). I find strong empirical support for this prediction, which is consistent with the empirical findings of Boschen and Smith (1995).

Second, the model predicts that stock return volatility declines with CEO tenure. The reason is that uncertainty about CEO ability contributes to uncertainty about dividends, and this uncertainty drops over time as agents learn the CEO's ability. I find strong empirical support for this prediction, which Clayton, Hartzell, and Rosenberg (2005) first documented.

I use the simulated method of moments (SMM) to estimate the model's five parameters: the CEO's share of the surplus from good and bad news, the volatility in profitability, the precision of the additional signal of CEO ability, and the prior uncertainty about CEO ability. Estimation uses data on excess stock returns and total CEO compensation from Execucomp between 1992 and 2007. I estimate the five parameters by matching 12 moments. The first 10 moments are stock return volatilities at different CEO tenure levels, and the last two are sensitivities of changes in expected CEO pay to lagged excess returns.

There are two main identification challenges. The first is to separately measure the CEO's share of the surplus from good and bad news. The model prescribes a simple solution: split the sample depending on whether the lagged excess return is high (indicating good news) or low (indicating bad news), and measure the sensitivity of CEO pay to lagged returns in each subsample. The second and larger challenge is to separately measure CEOs' share of the surplus and prior uncertainty about CEO ability. What allows identification is that these two parameters have different effects on two model predictions. First, the two parameters have a similar, positive effect on the predicted sensitivity of CEO pay to lagged returns. The reason is that a higher CEO share makes CEO pay move more in response to a given surplus, and higher prior uncertainty creates larger surpluses. Second, the two parameters have opposite effects on the the predicted drop in return volatility with CEO tenure. Intuitively, there is more uncertainty about dividends when there is either more initial uncertainty about CEO ability, or when shareholders bear a larger share (and hence CEOs bear a smaller share) of this uncertainty.

The estimated model fits several features of the data. It exactly fits the positive relation

[1] CEO pay is also sensitive to contemporaneous stock returns, but this sensitivity is exogenous in the model. The sensitivity to lagged returns results endogenously from learning.

between changes in expected CEO pay and lagged excess stock returns, both in the high- and low-return subsamples. CEO pay is more sensitive to returns in the high-return subsample, which the model matches. Return volatility drops during CEOs' first three years in office, both in the data and the model.

Parameter estimates are consistent with Harris and Hölmstrom's (1982) optimal-contracting benchmark. Depending on which compensation measure I use, a CEO bears an estimated 6 to 7% of the negative surplus resulting from bad news about his or her ability. Neither estimate is significantly positive. In other words, I cannot reject the hypothesis that CEOs have completely downward rigid wages, consistent with Harris and Hölmstrom's (1982). Shareholders, not the CEO, bear almost the entire cost of bad news about the CEO's ability. This result is not consistent with the models of Jovanovic (1979) and Gibbons and Murphy (1992), in which workers bear all the costs of bad news about their ability.

In contrast, CEOs capture an estimated 54 to 94% of the surplus from good news about their ability. Shareholders capture the small remaining portion of the benefits from good news. The estimates imply that a CEO's expected pay increases 0.54 for one or 0.94 for one with increases in the CEO's perceived contribution to firm profits. For the larger point estimate, I cannot reject the hypothesis that CEOs capture 100% of the surplus from good news. This result suggests that CEOs' outside options and bargaining power over a positive surplus are quite strong, as Harris and Hölmstrom (1982), Jovanovic (1979), and Gibbons and Murphy (1992) all assume.

To explore the determinants of CEO wage dynamics, I measure whether the model's parameters vary with firm, CEO, and industry characteristics. Allowing heterogeneity in parameters is rare in the structural corporate finance literature,[2] probably due to computational costs. Taylor (2010) and Nikolov and Whited (2009) avoid these costs by estimating their models in subsamples. However, that approach makes it difficult to control for several characteristics at once. I develop an approach that solves this problem while imposing minimal computational costs. This methodological contribution could apply to any project that uses GMM or SMM estimation.

Parameters indeed vary with certain characteristics. Gillan, Hartzell, and Parrino (2009) show that less than half of S&P500 CEOs have explicit employment agreements. I find that CEO pay is more downward rigid for those CEOs with explicit contracts, supporting the idea that firms offer explicit contracts in part to insure the CEO against bad news. CEOs appear to have more bargaining power over positive surpluses in industries that hire more

[2]Exceptions include Morellec, Nikolov, and Schuerhoff (2008) and Albuquerque and Schroth (2010).

CEOs from inside the firm, potentially because these firms have fewer potential replacement CEOs. Bargaining outcomes are not significantly related to other proxies for bargaining power, including the number of years the CEO works in the firm before becoming CEO, industry homogeneity, the number of similar-sized firms in the industry, the number of outside directorships the CEO holds, and the amount of unvested wealth the CEO would lose by leaving the firm. I offer several explanations for these insignificant results. As expected, there is less initial uncertainty about the ability of CEOs who spend more time inside the firm before becoming CEO, or in industries that typically hire insiders.

The skimming story of Bertrand and Mullainathan (2001) is an alternate explanation for why CEO pay rises after good news but does not fall after bad news. CEOs may simply grab resources in good times but insulate themselves from bad times, due to weak governance. If this skimming story is true, then the asymmetric response of CEO pay to good and bad news should be stronger in firms with weaker governance. I find the opposite result, using low institutional ownership as a proxy for weak governance. However, this opposite relationship is not statistically significant.

The model and parameter estimates have implications for the question, do CEOs matter for shareholders? I address this question using unanticipated CEO deaths as a natural experiment. Intuitively, the more CEOs matter for shareholders, the more stock prices should change around these deaths. I compare three model predictions to data on 81 unexpected CEO deaths, collected by Nguyen and Nielsen (2010). These comparisons also provide a useful out-of-sample test for the model. First, the model predicts an average event return of zero, consistent with the data. Second, the estimated model predicts an 11.3 percentage point increase in return volatility around the deaths. Volatility in the data increases by 6.8 percentage points. The model predicts no increase in volatility around deaths if CEOs capture 100% of their surplus. In this special case, differences in pay exactly offset differences in CEO ability, making shareholders indifferent about CEO ability. Third, the model predicts a negative relation between the event return and the deceased CEO's abnormal compensation, which is a proxy for CEO ability according to the model. Consistent with this prediction, Nguyen and Nielsen (2010) find a significantly negative regression slope. The model can match the slope's empirical magnitude if CEOs on average capture between 50 and 70% of their surplus, which is in the neighborhood of my parameter estimates. The model performs well in these three tests and suggests that CEOs matter for shareholders only if shareholders capture some of the CEO's surplus.

The baseline model ignores that CEOs are more likely to be fired after bad news about their ability. By focusing only on CEO pay, this paper measures just one dimension of CEO bargaining power. Also, ignoring firings makes the model ill suited for measuring the effect of news on CEO wealth and incentives, which is not this paper's focus. An important question, however, is whether ignoring firings results in biased parameter estimates, an issue I address in Section 4.

For robustness I solve a more general model in which agents simultaneously learn about CEO ability and firm quality, i.e., the firm fixed effect in profitability. I argue that the estimation method, which ignores learning about firm quality, biases the volatility of profitability but not the other four parameters.

This paper belongs to the literature on the dynamics of executive pay . Like my model, the models of Jovanovic (1979), Hölmstrom (1999), Harris and Hölmstrom (1982), and Gibbons and Murphy (1992) examine how a worker's pay changes over time when agents learn about the worker's ability. Gibbons and Murphy (1992) test directional predictions from their model, but this is the first paper to directly estimate this class of models. As such, I take a first step towards evaluating how well and what inputs these models need to fit the magnitudes in the executive pay data. Unlike the earlier models, my model allows CEOs to capture any fraction of the surplus from good and bad news, which generates new reduced-form predictions. Also, I derive new predictions about how stock return volatility varies with CEO tenure and model parameters. Edmans, Gabaix, Sadzik, and Sannikov (2009) also model the dynamics of executive pay and find that pay should be sensitive to lagged performance. Their result is due to consumption smoothing, whereas mine comes from learning. Milbourn (2003) develops a model of CEO pay and learning about ability, but his goal is to explain cross-sectional variation in stock-based compensation.

Gabaix and Landier (2008) and Tërvio (2008) also study how CEOs and shareholders split the CEO's surplus. Their evidence comes from the cross section, whereas my evidence comes from the time series. Specifically, their papers examine the matching between CEO ability and firms of different sizes, whereas this paper focuses on learning about CEO ability over time. Gabaix and Landier find that CEOs capture only 2% of the value they create, which is considerably less than this paper's estimates. Alder (2009) applies different functional forms and finds a higher capture rate. Tërvio (2008) finds that CEOs capture roughly 20% of the value they add to their firms. Examining returns around CEO deaths, Nguyen and Nielsen (2010) conclude that executives capture 80% of their surplus. Johnson et al. (1985), Hayes and Schaeffer (1999), and others also examine returns around executive deaths.

Several authors have asked variants of the question, do CEOs matter? Bennedsen, Perez-Gonzalez, and Wolfenzen (2008) find that firm performance deteriorates when CEOs are distracted by deaths or illness in their families. Bertrand and Schoar (2003) and Graham, Li, and Qiu (2009) show that managers carry certain attributes with them as they move between firms. Murphy and Zabojnik (2007) develop a model in which shareholders capture no rents from the CEO's general human capital and only a fraction of rents from firm-specific human capital. Malmendier and Tate (2009) find that CEOs extract higher pay after their perceived ability rises, proxied by the CEO winning a business award. My results are consistent with CEOs extracting benefits after an increase in their perceived ability.

The paper is structured as follows. Section 2 presents the learning model's assumptions and predictions. Section 3 describes the data, identification strategy, and estimation method. Section 4 presents parameter estimates and results on model fit. Section 5 describes how the structural parameters vary with firm, CEO, and industry characteristics. Section 6 discusses implications for shareholder value and unanticipated CEO deaths. Section 7 describes the extension with learning about firm quality, and section 8 concludes.

II. Model

In this section I develop and solve a dynamic model of CEO pay. In the model, some CEOs have higher ability than others, meaning they can produce higher average profitability. No one directly observes a CEO's ability. CEOs and shareholders learn about CEO ability over time by observing the firm's realized profits and an additional signal. When a CEO's perceived ability changes, so does his or her perceived contribution to future firm profits. The level of CEO pay then adjusts to a degree determined by parameters θ^{up} and θ^{down}, which reflect the CEO's bargaining power in response to good and bad news, respectively. The firm's market value and stock returns depend endogenously on beliefs about the CEO's ability.

The following example illustrates how the model works. Suppose a firm has \$1 billion in assets, the CEO's expected pay this year is \$6 million, the CEO's perceived contribution to annual excess (i.e., firm-specific) profits is \$10 million = +1% of assets, and the firm's contribution is –2%. Expected excess profitability is therefore –1%=–2%+1%. Suppose realized excess profitability this year is 0%, higher than expected. This is good news about the CEO's ability. Suppose Bayes' Rule tells us the CEO's perceived contribution increases from 1% to 1.2% of assets, an increase of 0.2% $\times$ \$1 billion = \$2 million. If the CEO captures $\theta^{up} = 75\%$ of this \$2 million surplus, for instance, by renegotiating a long-term compensation contract, then the CEO's expected pay rises by \$1.5 million. The CEO's expected pay next year is 6+1.5=\$7.5 million.

A. Assumptions

The model features firms i that live an infinite number of discrete 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

$$Y_{it} = a_i + \eta_i + v_t \left(\frac{M_{it}}{B_{it}} \right) + \varepsilon_{it}. \tag{1}$$

Variable η_i is the unobservable ability of the CEO in firm i at time t. A given CEO's ability is constant over his tenure. To be precise, η_i should have a CEO-specific subscript, since different CEOs within the firm can have different abilities. Parameter a_i reflects the contribution of non-CEO factors in firm i. For now I assume a_i is known and constant. Section 7 extends the model to allow learning about a_i. Shock v_t has conditional mean zero; this shock is common to all firms in the industry. M_{it} and B_{it} are the market and book assets of the firm at the beginning of year t. Scaling the industry shock v_t by the firm's market-to-book ratio simplifies the math, as I show later. Shock ε_{it} is an unobservable i.i.d. firm-specific shock distributed as $\mathcal{N}(0, \sigma_\varepsilon^2)$. There are very many firms in the same industry as firm i, which allows me to prove later that industry shock v_t is observable even though η_i and ε_{it} are not.

Assumption 2: Investors use exogenous discount factor β 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. (For this reason I sometimes drop the t subscript on assets B_{it}.) 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.

CEO j spends a total of T_j years in office. T_j is exogenous and known when the CEO is hired. For robustness, later I extend the model to allow endogenous succession. For simplicity I drop the j subscript on T_j.

Assumption 3: Agents have common, normally distributed prior beliefs about the ability of a newly hired CEO in firm i: $\eta_i \sim \mathcal{N}(m_{0i}, \sigma_0^2)$.

Different firms i may hire from different CEO talent pools, so the prior mean ability of CEOs, m_{0i}, is firm specific. For instance, if high-quality CEOs match with large firms, as in Gabaix and Landier (2008), then prior mean ability m_{0i} is increasing in firm size. The prior mean m_{0i} will drop out of the analysis.

Assumption 4: Investors use Bayes' Rule to update beliefs about CEO ability η_i 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 $z_{it} \sim \mathcal{N}\left(\eta_i, \sigma_z^2\right)$.

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 agents 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 σ_z is lower.

Assumption 5: Realized total compensation for the CEO in firm i and year t equals

$$w_{it} = E_t[w_{it}] + b_{it} r_{it}, \tag{2}$$

where r_{it} is the firm's industry-adjusted stock return, which is endogenous in the model. Expected pay $E_t[w_{it}]$ and the contemporaneous pay-performance sensitivity b_{it} are both known at the beginning of period t, but r_{it} is not.

The model makes predictions about changes over time in a CEO's expected pay, and I use these predictions to estimate the model. The contemporaneous pay-performance sensitivity b_{it} 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 the optimal amount of 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 on firm and CEO characteristics in a flexible way. Finally, making b_{it} exogenous simplifies the model solution and estimation.

Assumption 6: The change in the CEO's expected annual pay equals equals

$$\begin{aligned}
\Delta E_t\left[w_{it}\right] &= E_t\left[w_{it}\right] - E_{t-1}\left[w_{it-1}\right] && (3)\\
&= \theta_t B_{it}\left(E_t\left[\eta_i\right] - E_{t-1}\left[\eta_i\right]\right) && (4)\\
\theta_t &= \theta^{up} \quad \text{if } E_t\left[\eta_i\right] \geq E_{t-1}\left[\eta_i\right] \quad \text{(beliefs increase)} && (5)\\
\theta_t &= \theta^{down} \quad \text{if } E_t\left[\eta_i\right] < E_{t-1}\left[\eta_i\right] \quad \text{(beliefs decrease).} && (6)
\end{aligned}$$

By equation (1), the CEO's expected contribution to firm dollar profits in period t is $B_{it}E_t\left[\eta_i\right]$. The change in this expected contribution, due to learning about CEO ability η_i, is $B_{it}(E_t[\eta_i] - E_{t-1}[\eta_i])$. Assumption 6 therefore states that the CEO captures a fraction θ_t of the change in his expected contribution to firm profits. When beliefs increase (decrease), the CEO captures a fraction θ^{up} (θ^{down}) of this surplus. The parameters θ^{up} and θ^{down} measure the CEO's bargaining power over changes in level of pay. The model makes predictions about changes over time in expected pay, but not about the level of expected pay.

The remainder of this subsection provides a few rationales for Assumption 6, which is admittedly ad hoc. The purpose here is to illustrate that there are many economic factors that affect CEO bargaining power. Section 5 explores several of these factors. However, the 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, and also measure the other parameters that affect CEO wage dynamics.

A special case of the model is when the CEO's expected pay each year equals the CEO's expected contribution to firm profits, $BE_t[\eta_i]$. 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 illustrate what economic setup can produce this special case. They assume that every period there are multiple identical firms competing for the CEO. The firms offer the CEO a one-period contract, 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 CEO pay 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 benchmark with multi-period contracts. Their model predicts that the CEO's expected pay never drops (i.e., $\theta^{down} = 0$), but the CEO has enough bargaining power to capture the entire positive surplus (i.e., $\theta^{up} = 100\%$). I show that the data are much closer to this benchmark.

These stylized equilibrium models omit several determinants of CEO bargaining power. If the CEO's human capital is specific to the firm, then the CEO cannot make a strong threat to leave firm, as in 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 Tërvio (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 threat to leave the firm is weaker.

B. Solving the model

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

$$\eta_i \sim N\left(m_{it}, \sigma^2_{\tau_{it}}\right), \tag{7}$$

where τ_{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 τ. Agents update their beliefs about CEO ability by observing the mean-zero surprises in profitability and the additional signal:

$$\widetilde{Y}_{it} = Y_{it} - a_i - \left(\frac{M_{it}}{B_{it}}\right) v_t - m_{t-1} = \eta_i + \varepsilon_{it} - m_{it-1} \tag{8}$$

$$\widetilde{z}_{it} = z_{it} - m_{it-1}. \tag{9}$$

Applying Bayes' rule, the posterior variance follows

$$\sigma^2_\tau = \sigma^2_0 \left(1 + \tau\left(\frac{\sigma^2_0}{\sigma^2_\varepsilon} + \frac{\sigma^2_0}{\sigma^2_z}\right)\right)^{-1}, \tag{10}$$

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

$$m_{it} = m_{it-1} + \frac{\sigma^2_\tau}{\sigma^2_\varepsilon}\widetilde{Y}_{it} + \frac{\sigma^2_\tau}{\sigma^2_z}\widetilde{z}_{it}. \tag{11}$$

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

$$\Delta E_t\left[w_{it}\right] = \theta_t B_{it}\left(m_{it-1} - m_{it-2}\right). \tag{12}$$

Substituting in equation (11) yields

$$\Delta E_t\left[w_{it}\right] = \theta_t B_{it}\left(\frac{\sigma^2_{\tau-1}}{\sigma^2_\varepsilon}\widetilde{Y}_{it-1} + \frac{\sigma^2_{\tau-1}}{\sigma^2_z}\widetilde{z}_{it-1}\right). \tag{13}$$

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

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

$$D_{it} = B_{it}Y_{it} - w_{it}, \tag{14}$$

and the firm's value at the beginning of year t equals

$$M_{it} = E_t\left[\sum_{s=0}^{\infty}\beta^{s+1}D_{it+s}\right]. \tag{15}$$

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.

C. Model predictions

First I present predictions about stock returns and return volatility, and then I present predictions about CEO pay. All proofs are in the Internet Appendix.[3]

Prediction 1 (excess returns): The excess stock return (firm minus industry) in year t equals

$$r_{it} \approx \frac{B_{it}}{M_{it}}\widetilde{Y}_{it} + \frac{B_{it}}{M_{it}}\left(1-\theta_{t+1}\right)\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(m_{it}-m_{it-1}\right) - median\left(r_{it}\right). \tag{16}$$

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. The equation for excess returns depends on the median return, because returns are skewed and the expected excess return is zero.[4] Equation (16) decomposes excess returns into an unexpected dividend (the $\widetilde{Y}_{it}$ term) and a change in market value due to learning about CEO ability (the $m_{it}-m_{it-1}$ term). To understand the learning term, note that when the CEO's perceived contribution to dollars profits changes by $B_{it}\left(m_{it}-m_{it-1}\right)$ and the CEO has $T-\tau$ years left in office, then the present value of this news (i.e. the total surplus) equals

$$\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)B_{it}\left(m_{it}-m_{it-1}\right). \tag{17}$$

[3]The Internet Appendix is currently attached to this document. Eventually it can live on the Internet somewhere.

[4]Excess returns are skewed when and only when $\theta^{up} \neq \theta^{down}$, in which case θ_{t+1} is correlated with unexpected profitability $\widetilde{Y}_{it}$. The Appendix contains an expression for the predicted median excess return, which is a function of current and final tenure, depends on all model parameters, and equals zero when $\theta^{up} = \theta^{down}$.

Since shareholders capture only a fraction $(1-\theta_{t+1})$ of this surplus, we obtain the second term in equation (16) after scaling by market cap M_{it}. Unexpectedly high profits $\widetilde{Y}_{it}$ generate high excess returns by producing a high dividend (the first term in (16)) and by pushing up beliefs about the CEO (the second term). From equation (11), the change in beliefs depends on both the unexpected dividend and the additional signal z_{it}.

The sensitivity of excess returns to changes in beliefs is decreasing in θ_t, the CEO's share of the surplus. In other words, perceived CEO ability matters more for shareholders when the CEO captures less (and hence shareholders capture more) of the surplus from perceived CEO ability.[5]

I use stock return volatility to estimate the model. The Internet Appendix provides a closed-form expression for excess stock return volatility, and also proves the following limits, special cases, and comparative statics:

Prediction 2 (excess return volatility):

1. In the special case with no learning, i.e., $\sigma_0^2 = 0$, or in the limit when tenure goes to infinity, then the variance of excess stock returns equals

$$var_t\left(r_{it}\right) = \left(\frac{B_{it}}{M_{it}}\right)^2 \sigma_\varepsilon^2. \tag{18}$$

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

$$var_t\left(r_{it}\right) = \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\varepsilon^2\right), \tag{19}$$

 where $\sigma_{\tau-1}^2$ is the uncertainty about CEO ability at the beginning of year t, given in equation (10).

3. If prior uncertainty $\sigma_0^2 > 0$, 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 σ_0, and decreases with the CEO's share of the surplus θ.

[5] To see why, imagine the CEO's share θ is below 1. Good news about CEO ability this year increases expected future profits. Since shareholders capture a positive fraction $1-\theta$ of this surplus, expected future dividends (=profits minus CEO pay) are also higher, so the firm's market value increases. In the special case in which the CEO captures his entire surplus (i.e. $\theta^{up} = \theta^{down} = 1$), the firm's market value is constant over time (result in Appendix); good news may coincide with a high dividend this year but not a higher market value at the end of the year, because the increase in CEO pay exactly offsets the benefits to shareholders.

The intuition for Prediction 2 is that higher uncertainty about CEO ability leads to more uncertainty about dividends and hence higher return volatility. There is more uncertainty when the CEO starts out with more uncertainty (i.e., higher prior uncertainty σ_0^2) or when fewer years of learning have occurred. Clayton, Hartzell, and Rosenberg (2005) show empirically that stock return volatility drops after CEO successions, which is consistent with this prediction. When there is no uncertainty about CEO ability, which occurs when prior uncertainty is zero or in the limit when many years of learning have occurred, then the only source of return volatility is the shock to profitability, which has volatility σ_ε^2 (equation (1)). In the special case in which $\theta^{up} = \theta^{down} = 1$, uncertainty about CEO ability affects only the riskiness of this period's net profits, because changes in CEO pay exactly offset changes in expected future profits. Return volatility is decreasing in the CEO's share θ, because lower values of θ make uncertainty about ability affect not only this year's dividends, but also future years' dividends, since shareholders capture a larger fraction $1-\theta$ of the CEO's expected contribution to future profits.

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

$$\frac{\Delta E_t\left[w_{it}\right]}{M_{it-1}} \approx \gamma r_{it-1} + \gamma \frac{B_{it}}{M_{it-1}}\left(\frac{\sigma_\varepsilon^2}{\sigma_z^2}\right)\widetilde{z}_{it-1} + g\left(\cdot\right) \tag{20}$$

$$\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t\right) = \frac{\sigma_{\tau-1}^2\theta_t}{\sigma_\varepsilon^2 + \sigma_{\tau-1}^2\beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)}, \tag{21}$$

where g is a deterministic function given in the Appendix.

These equations use the approximation that the pay-performance sensitivity b_{it} is much less than the firm's market value, which I confirm empirically. The scaled change in expected CEO pay depends on lagged stock returns r_{it-1} and lagged values of the additional signal $\widetilde{z}_{it-1}$. Variable γ is the sensitivity of changes in expected pay to lagged returns, which depends on the current and final tenure, θ_t, and other model parameters. The sensitivity of changes in expected pay to the lagged z signal depends on γ as well as $\sigma_\varepsilon^2/\sigma_z^2$, which is the relative precision of the profitability signal and additional signal. Next I examine several special cases to explain the intuition behind this prediction.

The first case assumes the CEO captures the entire surplus from changes in beliefs, so $\theta^{up} = \theta^{down} = 1$. In this special case the sensitivity γ from equation (21) simplifies to $\gamma = \sigma_{\tau-1}^2/\sigma_\varepsilon^2$. The sensitivity γ is positive as long as there is some initial uncertainty about CEO ability ($\sigma_0 > 0$). To see why, imagine 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 perceived ability, which causes expected CEO pay to rise in period t, which means $\Delta E_t [w_{it}]$ is positive. Therefore, we have a positive correlation between change in expected pay and last period's excess stock return r_{it-1}. An unexpectedly high additional signal z_{it-1} has the same two effects, because a high additional signal increases stock returns (equation (16)) and raises beliefs about the CEO.

Another special case is when there is no learning, which occurs when prior uncertainty σ_0=0. In this special case $\gamma = 0$, meaning there is no relation between the change in expected pay and lagged returns. The reason is that there are no changes in beliefs about CEO ability, and hence there are no surpluses and no changes in expected pay. This special case illustrates that the sensitivity of expected pay to lagged returns is due to exclusively to learning about CEO ability.

Yet another special case is when $\theta_t = 0$, meaning the CEO captures none of the surplus from changes in his or her perceived ability. In this case we obtain $\gamma = 0$. The CEO's expected pay does not change over time, even though his or her perceived ability does change over time.

The sensitivity γ depends on θ_t but not on both θ^{up} and θ^{down}. If beliefs increased in period $t-1$, then $\theta_t = \theta^{up}$, and the sensitivity γ in that period depends on θ^{up} but not θ^{down}. This result will help later to disentangle θ^{up} and θ^{down} empirically.

Corollary to Prediction 3: If $\sigma_0 > 0$ and $\theta_t > 0$, then the sensitivity γ is increasing in prior uncertainty σ_0, decreasing in profitability volatility σ_ε, increasing in the CEO's share θ_t, and independent of firm size.

The intuition for these comparative static results is similar to the intuition above. CEO pay is more sensitive to lagged stock returns when there is more uncertainty, because higher uncertainty causes beliefs to move more in response to any given signal. Therefore, surpluses are larger and the sensitivity γ is higher when there is more initial uncertainty (σ_0). Expected CEO pay moves less with lagged stock returns when profitability is more volatile (higher σ_ε) because beliefs change less in response to noisier signals. Expected CEO pay moves more with lagged stock returns when θ_t is higher, because the CEO receives a larger fraction of the surplus from changes in beliefs. The sensitivity γ does not depend on firm size, because the change in expected pay is scaled by firm size in equation (20).

III. Estimation

First I estimate the pay-performance sensitivity b_{it} using a simple OLS regression. Then I estimate the model's five parameters by SMM, which involves matching empirical and predicted moments. The five parameters are the volatility of profitability shocks σ_ε, the volatility of the additional signal σ_z, the prior uncertainty about CEO ability σ_0, and θ^{up} and θ^{down}, the CEO's fraction of the surplus from good and bad news (respectively) about the CEO's ability. First I estimate the model assuming these parameters are constant across firms, CEOs, and years. In Section 5 I allow the parameters to depend on observable characteristics like firm size and CEO age. Estimation uses annual data on realized CEO pay, excess stock returns, and excess stock return volatility. Before describing the data and estimator in detail, I provide intuition for how the parameters are identified, which informs the choice of moments used in SMM estimation.

A. Identification

To see how the pay-performance sensitivity b_{it} is identified, recall assumption 5:

$$w_{it} = E_t\left[w_{it}\right] + b_{it} r_{it}. \tag{22}$$

Since excess return r_{it} is orthogonal to expected pay $E_t\left[w_{it}\right]$, we can estimate b_{it} from an OLS regression of realized CEO pay w_{it} on the contemporaneous excess return r_{it} and information known at the beginning of period t. Using the estimate $\widehat{b}_{it}$, I estimate expected pay according to

$$\widehat{E}_t\left[w_{it}\right] = w_{it} - \widehat{b}_{it} r_{it}. \tag{23}$$

The volatility of profitability, σ_ε, is identified off of excess stock return volatility for long-tenured CEOs. From Prediction 2, return volatility drops with CEO tenure and approaches a limit that depends only on parameter σ_ε.

Parameter σ_z, the additional signal's noise, is mainly identified off how quickly return volatility drops with CEO tenure. To illustrate the intuition, Figure 1 plots predicted return volatility versus CEO tenure for three values of σ_z. Return variance drops faster with tenure when σ_z is lower, meaning the additional signal z_{it} is more precise. The reason is that a more precise signal allows agents to learn more quickly about CEO ability, which in turn causes uncertainty about CEO ability— and hence uncertainty about dividends— to drop more quickly. When σ_z is extremely high, so there is effectively no additional signal[6], return volatility drops extremely slowly, because the remaining signal (firm profits) is very noisy. Reducing σ_z increases return volatility in the CEO's first year, because beliefs move more during the CEO's first year in office when the additional signal z is more precise.

INSERT FIGURE 1 HERE

Disentangling prior uncertainty σ_0 and the fraction of surplus going to the CEO (θ^{up} and θ^{down}) is more challenging. First I consider the special case in which $\theta^{up} = \theta^{down} = \theta$, then I explain how I separately estimate θ^{up} and θ^{down}. The total drop in return volatility during a CEO's tenure depends on both σ_0 and θ. Figure 2 plots the infinite number of pairs $\{\theta, \sigma_0\}$ that allow the model to match a given drop in return volatility. For instance, the model can match it if there is high uncertainty σ_0 (which increases the predicted drop in volatility), but a large fraction θ going to the CEO (which reduces the drop in volatility). Alternatively, lower uncertainty σ_0 with a lower CEO surplus θ can also match the same drop in volatility.

INSERT FIGURE 2 HERE

We need at least one additional empirical moment to uniquely identify θ and σ_0. The obvious candidate is γ, the sensitivity of changes in expected CEO pay to lagged stock returns. From the Corollary to Prediction 3, γ depends on parameters θ and σ_0, as well as other parameters. Crucially, γ is increasing in both σ_0 and θ, whereas the drop in return volatility is increasing in σ_0 and decreasing in θ. Since the parameters drive these moments in different directions, we can use the two moments to uniquely pin down the two parameters. Figure 2 illustrates this strategy by plotting the infinite combinations of σ_0 and θ that allow the model to match a given sensitivity γ. The model matches this moment either by choosing high uncertainty σ_0 and low θ (so surpluses from learning are large, but the CEO captures less of them), or by choosing low uncertainty and high θ (so surpluses are small, but the CEO captures more of them). The lines in Figure 2 cross at a unique point, meaning there is a unique pair of parameters $\{\sigma_0, \theta\}$ that can match both the drop in return volatility and the sensitivity of expected pay to lagged returns.

To separately estimate θ^{up} and θ^{down}, I split the sample depending on whether lagged excess returns are above or below their median. Beliefs about CEO ability mostly decreased in the subsample with low lagged excess returns, and vice-versa. The model predicts that when beliefs decrease, the sensitivity of expected pay to lagged returns depends on θ^{down} but not θ^{up} (equation (21)), and vice versa when beliefs increase. By measuring this sensitivity

[6]From equation (11), when the additional signal's volatility is extremely high, agents effectively ignore the additional signal and rely only on the profitability signal.

in these two subsamples, I can separately measure θ^{up} and θ^{down}.

I do not estimate the discount factor β, but instead set it to a plausible value, 0.9.[7]

B. Data

Data come from Execucomp, CRSP, Compustat, Kenneth French's website, Thomson Financial, Risk Metrics, and Gillan, Hartzell, and Parrino (2009). 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.

Annual stock returns are computed from monthly CRSP returns and information on firms' fiscal calendars. I use Kenneth French's classification of 49 industries throughout. Equal weighted annual industry returns are computed for each firm to take into account firms' different fiscal calendars. Excess return r_{it} equals the firm's annual stock return minus the corresponding industry return.

Next I discuss measurement of total annual CEO pay, w_{it}. It is common to measure pay as total compensation plus changes in CEO wealth (e.g. Core, Guay, and Verrecchia (2003)). That measure is ideal for studying CEO incentives. Since this paper studies the CEO labor market rather than CEO incentives, this paper needs a different measure of pay. A plausible interpretation of the model is that the firm and CEO consider renegotiating 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. Clearly, salary and bonus should be included in w_{it}. It is less obvious how to treat stock and option grants, which typically vest gradually over several years (Kole (1997)). Any grants awarded and vested before year t are sunk from the CEO and firm's perspectives, so they should not affect the decision to continue the employment relationship during year t. This consideration rules out measures that include grants at the time they are exercised (e.g. Execucomp's TDC2), since these grants may have vested in previous years. This consideration also rules out using changes in CEO wealth, because CEO wealth may have accrued and vested in previous years. While this previously vested wealth likely affects CEO incentives to work hard and make value-maximizing decisions, it should not affect a CEO's decision to remain in the firm, which is the focus of this paper.

Having ruled out shares and options that vested in the past, the question becomes, should

[7]This discount factor implies that expected stock returns equal $(1-\beta)/\beta$ =11% per year.

we include stock and option grants in the year they were awarded or the year when they vested? I use both methods rather than taking a stand. The first measure of annual CEO pay, denoted $w^{(grant)}$, is Execucomp's total compensation variable TDC1, which is comprised of the following: salary, bonus, other annual, total value of restricted stock granted, total value of stock options granted (using Black-Scholes), long-term incentive payouts, and "all other total." The second measure, denoted $w^{(vest)}$, is the same as the first, except it includes stock and options that vest during year t, valued at the time they vest, regardless of when they were granted. The Internet Appendix contains details on constructing $w^{(vest)}$, which to my knowledge is new to the literature.

The measure of excess stock return variance for firm i in fiscal year t is $RVAR_{it}$, which equals the annualized variance of weekly industry-adjusted stock returns during the fiscal year. I annualize by multiplying the weekly variance by 52. I remove year fixed effects in volatility by subtracting off each year's average volatility and adding back the full-sample average return volatility.

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.

Estimation uses data on the change in expected CEO pay scaled by lagged market value. Following equation (23), I estimate expected pay $\widehat{E}_t[w_{it}]$ by subtracting the unexpected portion $\widehat{b}_{it} r_{it}$ from realized pay w_{it}. As I explain above, b_{it} is identified in a regression of realized pay on contemporaneous returns and information at the beginning of the year t. I parameterize b_{it} as

$$b_{it} = a_0 + a_1 \log(M_{it-1}), \tag{24}$$

and then estimate coefficients a_0 and a_1 in the pooled OLS regression

$$w_{it} = c_1 w_{it-1} + (c_2 + c_3 \log(M_{it-2}))\, r_{it-1} + (a_0 + a_1 \log(M_{it-1}))\, r_{it} + u_{it}. \tag{25}$$

Regression estimates are tabulated in the Internet Appendix. Estimated expected pay then equals

$$\widehat{E}_t[w_{it}] = w_{it} - (\widehat{a}_0 + \widehat{a}_1 \log(M_{it-1}))\, r_{it}. \tag{26}$$

The estimated change in expected pay, scaled by lagged market value, equals

$$\frac{\Delta \widehat{E}_t[w_{it}]}{M_{it-1}} = \frac{\widehat{E}_t[w_{it}] - \widehat{E}_{t-1}[w_{it-1}]}{M_{it-1}}. \tag{27}$$

I winsorize $\Delta\widehat{E}_t\left[w_{it}\right]/M_{it-1}$ at the 1st and 99th percentiles, and I subtract off the yearly median across CEOs, since the model does not attempt to explain aggregate changes in CEO pay. I also winsorize excess returns at the 1st and 99th percentiles.

INSERT TABLE 1 HERE

Summary statistics are in Table 1. The database contains 20,636 firm/year observations and 4,540 CEOs. There are fewer observations of the change in expected pay, which cannot be computed in CEOs' first year in office. Mean realized pay is \$6 million using the $w^{(vest)}$ measure, with a standard deviation of \$10 million. There is only slightly less variation in the measure of expected pay, meaning the estimation procedure attributes almost all the variation in realized pay to variation in expected pay. Using $w^{(vest)}$, the standard deviation of changes in expected pay is around \$7 million or 0.55% of lagged market cap. $RVAR_{it}$, the annualized variance of weekly returns within a firm/year, has median 0.12, which corresponds to annualized return volatility of $\sqrt{0.12} = 35\%$. The median firm/year observation is for a CEO in his 6th year in office, and who is expected to complete a total of 12 years before leaving office. Section 5 discusses the remaining variables in Table 1.

C. Estimator

I estimate the five model parameters in $\Theta = \left[\begin{array}{ccccc}\sigma_\varepsilon^2 & \sigma_z^2 & \sigma_0^2 & \theta^{up} & \theta^{down}\end{array}\right]$ using SMM. Like GMM, SMM estimates parameter values by matching certain empirical moments and model-implied moments as closely as possible. Whereas GMM uses closed-form expressions to compute the model-implied moments, SMM estimates them using simulations. A few of the moments below are available in closed-form, so I do not estimate them by simulation. In this sense, the estimator is technically a hybrid between SMM and GMM. The estimator is

$$\widehat{\Theta} \equiv \arg\min_{\Theta}\left(\widehat{M} - \widehat{m}\left(\Theta\right)\right)' \mathbf{W}\left(\widehat{M} - \widehat{m}\left(\Theta\right)\right). \tag{28}$$

$\widehat{M}$ is a vector of moments estimated from the actual data, and $\widehat{m}\left(\Theta\right)$ is the corresponding vector of model-implied moments. The hat on $\widehat{m}$ indicates that some model-implied moments are estimated by simulation. For these simulations, I use parameter values Θ to simulate a sample 20 times larger[8] than the empirical sample, then I compute the moment from 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 M. Following McFadden (1989) and Pakes and Pollard (1989), I adjust standard errors to take

[8]Michaelides and Ng (2000) find that using a simulated sample 10 times as large as the empirical sample generates good small-sample performance. I use a simulated sample 20 times larger to be conservative.

into account that the simulated model-implied moments are estimated with error. Following Hennessy and Whited (2005) and others, I use simulated annealing to search for parameter values that solve (28).

I estimate the five parameters in Θ using 12 moments in vectors M and m. The first 10 moments are means, and the last two are regression slopes. The first 10 moments, which are available in closed form, are the average variance of excess returns for CEOs in their 1st, 2nd, ..., 9th, and 10+ year in office. The 11th and 12th moments are the slopes M_{11} and M_{12} from the following OLS regression:

$$\frac{\Delta \widehat{E}_t\left[w_{it}\right]}{M_{it-1}} = a_0 + a_1 \mathbf{1}(r_{it-1} > med) + M_{11} r_{it-1} \mathbf{1}(r_{it-1} > med) + M_{12} r_{it-1} \mathbf{1}(r_{it-1} \leq med) + e_{it}, \tag{29}$$

where med is the sample median lagged excess return. These last two moments are not available in closed form, so I estimate them from simulated data. Moment M_{11} (M_{12}) measures the sensitivity of changes in expected pay to high (low) lagged excess returns. Beliefs about CEO ability mostly increase when lagged returns are high. Therefore, M_{11} is most informative about $\theta^{(up)}$, the CEO's share of a surplus from good news. Similarly, M_{12} is most informative about $\theta^{(down)}$.

IV. Estimation results

I begin by describing how the model fits the data. I then present the main parameter estimates, which assume parameter values are constant across firms and CEO. In the next section I describe how the parameter values vary with firm, CEO, and industry characteristics.

A. Model Fit

Figure 3 plots the variance of excess stock returns against CEO tenure. Variance in years 1-10 are the first 10 moments used in estimation. I also include stock return variance in the previous CEO's last year in office (year 0) and the previous 3 years. The figure compares the actual data (dashed line) to the estimated model's predictions (solid line). The grey area indicates the 95% confidence interval for the empirical data. I present results using the $w^{(grant)}$ pay measure; results are similar using $w^{(vest)}$. In both the model and the data, return volatility peaks in the new CEO's first year in office and then drops with tenure. Taking square roots of the empirical variances, we see return volatility drop from 42% in CEOs' first year in office to 37% in the fifth year, and then rise to 39% for CEOs in years 10+. The model generates a drop in return variance due to learning, as expected. In both the data and the model, return variance drops rapidly and then levels off, indicating that agents are learning very fast about CEO ability. The model's predicted return variance is within the empirical 95% confidence interval in 8 out of 10 tenure categories.

INSERT FIGURE 3 HERE

Outside the model, an alternate explanation for the decline in return volatility 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 profitability. Also, I show that the decline in return volatility is robust to controlling for firm size, firm fixed effects, year fixed effects, and the magnitude of the shock to profitability in the same firm and year.

Figure 3 shows that, empirically, return volatility increases before the previous CEO leaves office, whereas the model predicts a drop in volatility. There are two main reasons for this model failure. First, there is no news about the replacement CEO's ability until he or she takes office, whereas uncertainty about the incumbent's ability continues dropping until he or she leaves office. Second, successions are exogenous and perfectly anticipated in the model. When a CEO is hired, everyone knows exactly when he or she will leave office.

To determine whether endogenous successions explain the run-up in volatility, I alter the model as follows: I still allow CEOs to retire if they reach their retirement date, but I now assume a CEO is fired as soon as his posterior mean ability m_{it} drops below an exogenous threshold, $\underline{\mu}$.[9] Taylor (2010) shows that a firing rule like this explains several features of the CEO turnover data and may be optimal for boards of directors. I choose $\underline{\mu}$ so that 2% of CEOs are fired per year on average, which matches the empirical rate, also from Taylor (2010). I simulate this altered model and plot average return volatility versus tenure in Figure 4. The altered model produces a rise and fall in volatility around successions, which qualitatively matches the data. The model now predicts an increase in realized volatility before successions, because some successions are due to firings, and firings are usually preceded by extremely low earnings in order to pull the posterior mean below the firing threshold. In terms of magnitudes, predicted pre-succession volatility is still outside the empirical 95% confidence region.

INSERT FIGURE 4 HERE

[9] I continue using equation (16) to simulate stock returns, so I assume investors irrationally fail to anticipate the possibility of firing the CEO. If investors are rational, pre-firing stock prices will reflect a positive probability of a firing. Correcting for this requires solving a dynamic program, as in Taylor (2010). I have chosen to keep this exercise simple since it is not a main result of the paper.

Although this model with endogenous firings fits the data better, I focus on estimation results from the simpler model, for three reasons. First, comparing Figures 3 and 4, pre-succession volatility looks quite different in the altered model, post-succession volatility does not. Since the estimation procedure uses only post-succession volatility, I do not expect my parameter estimates to change much if I used the altered model. Second, the altered model fits the data better at the expense of an extra degree of freedom ($\underline{\mu}$), which would need to be estimated and is not the focus of this paper. Finally, the altered model is much less tractable.

Next I examine how well the model fits the relation between CEO pay and lagged excess returns. I present results using the grant-date pay measure $w^{(grant)}$; results are similar using the vesting-date pay measure $v^{(vest)}$. Figure 5 plots the empirical scaled change in expected CEO pay ($\Delta E_t [w_{it}] / M_{it-1}$) versus the lagged excess return r_{it-1}. Each point is a firm/year observation. The solid vertical line splits the sample by the median lagged return. The figure shows the best fit line in each subsample. The slopes of these two lines are the 11th and 12th moments used in estimation. The slope is significantly positive in both subsamples, indicating that, on average, expected pay rises more (or falls less) after better news about the CEO. The slope is significantly higher in the subsample with high lagged returns (0.0027 versus 0.0014), suggesting that expected pay is more sensitive to good news than bad news.

INSERT FIGURE 5 HERE

Figure 6 shows the same plot, but with data simulated from the estimated model. The model exactly matches the empirical slopes in the two subsamples. In both the model and the real data, expected pay is less sensitive to lagged returns in the subsample with low excess stock returns. Also, the model produces many data points in the upper left quadrant (low lagged return but CEO pay increases) and lower right quadrant (vice versa). These data points are surprising, because stock prices and beliefs about the CEO are moving in opposite directions. Without the additional signal z_{it}, the model would not produce data points in these quadrants, because low returns would always coincide with decreasing beliefs, and vice versa. The additional signal z can create a wedge between returns and changes in beliefs. Equation (16) shows that if a firm delivers low profits yet receives a very positive signal z about the CEO, the stock return may be negative and yet beliefs about the CEO may increase, generating a data point in the upper left quadrant. By generating these "off-quadrant" data points, the additional signal z allows the model to produce a cloud of data points that is similar to the one we see in the real data.

INSERT FIGURE 6 HERE

Table 2 summarizes the 12 moments used in estimation. The p-values test whether the model-implied moment matches the empirical moment. Using both pay measures, equality is rejected at the 1% level for one of 12 moments, rejected at the 5% level for a second moment, and for the remaining moments equality cannot be rejected even at the 10% level. The table also provides a χ^2 statistic that jointly tests whether the model matches all 12 empirical moments. The p-value rejects joint equality at the 1% (5%) confidence level using the $w^{(grant)}$ ($w^{(vest)}$) measure. In other words, the data reject the model. I do not interpret this result as particularly damning, since we can reject any model with enough data. For instance, if I tried to fit 5 moments instead of 12, using 5 free parameters, model fit would have appeared perfect.

INSERT TABLE 2 HERE

B. Main Parameter Estimates

Table 3 presents the parameter estimates using both measures of CEO pay. The estimated standard deviation of profitability, σ_ε, is roughly 36% using both pay measures. The model needs this high value to match the high level of stock return volatility. This high volatility implies the profitability signal of CEO ability is extremely noisy. In contrast, the additional signal z is quite precise, with an estimated volatility of 1.8% or 3.3% depending on which pay measure I use. The precise z signal allows agents to learn very quickly, which allows the model to fit the sharp drop in stock return volatility during CEOs' first two years in office.

INSERT TABLE 3 HERE

The estimated standard deviation of prior beliefs about CEO ability (σ_0) is 4.2% using $w^{(grant)}$ and 5.5% using $w^{(vest)}$. This parameter is in units of annual profits (before subtracting CEO compensation) as a percent of assets. Using the estimate $\widehat{\sigma}_0 = 4.2\%$, the difference in average profitability between CEOs at the 5th and 95th ability percentiles is $2 \times 1.65 \times \widehat{\sigma}_0 = 14\%$ of assets per year, which is extremely large. For comparison, using different data and a different 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 larger than 2.4%. 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.

Using the pay measures $w^{(grant)}$ and $w^{(vest)}$, the estimates of CEOs' share of a negative surplus (θ^{down}) are 6.7% and 6.0%, respectively. In other words, the level of CEO pay moves 0.067 or 0.060 for one with decreases in the CEO's perceived contribution to firm profits. These estimates imply that shareholders, not the CEO, bear almost all the costs from bad news about a CEO's ability. Neither estimate of θ^{down} is significantly different from zero, so I cannot reject the hypothesis that CEOs have completely downward rigid wages, i.e., wages that do not drop following bad news about ability.

The estimates of θ^{up} (CEOs' share of a positive surplus) are 53.6% and 93.6% using $w^{(grant)}$ and $w^{(vest)}$, respectively. In other words, the level of CEO pay changes 0.536 or 0.936 for one with increases in the CEO's perceived contribution to firm profits. These estimates imply that CEOs capture 53.6–93.6% of the benefits from an improvement in their perceived ability. The higher estimate of θ^{up} is statistically indistinguishable from 100%, meaning I cannot reject the hypothesis that CEOs capture all the benefits from good news. The lower estimate, however, is significantly less than 100%.

In sum, I find that CEO pay responds asymmetrically to good and bad news. Using $w^{(grant)}$ ($w^{(vest)}$), I can reject the hypothesis that $\theta^{up} = \theta^{down}$ with a t-statistic of 3.6 (6.1). It is not surprising that I find $\theta^{up} > \theta^{down}$, since changes in CEO pay are more sensitive to high lagged returns than to low lagged returns, as shown in Figure 5.

The estimates of θ^{up} and θ^{down} are not consistent with the models of Jovanovic (1979) or Gibbons and Murphy (1992), which predict that workers capture the entire surplus resulting from both good and bad news. However, the estimates are consistent with the model of Harris and Hölmstrom (1982), in which firms optimally offer contracts with downward rigid wages, but workers' outside options are strong enough that they can capture the entire positive surplus from good news.

The estimated model ignores endogenous CEO firings, which potentially introduces estimation bias. Without firings, stock prices in the model react too much to bad news about CEO ability. The reason is that investors should rationally anticipate CEO firings following bad news, so stock prices should not drop so much as my model predicts.[10] After incorporating this effect, my parameter estimates would produce a pay/return sensitivity that is weaker than the empirical one. To continue fitting the data, the CEO's share of a negative surplus must be higher than my current estimates of 6 to 7%.

[10] Bond, Goldstein, and Prescott (2010) formalize this story.

V. Determinants of CEO Wage Dynamics

In theory, bargaining outcomes should depend on the CEO's and board's outside options, and on contractual constraints. For instance, CEOs should capture more of a positive surplus when the CEO can make a more credible threat to leave the firm, when the board lacks a good replacement CEO, and when renegotiating the CEO's contract is less costly. In this section I test whether model parameters are correlated with proxies for these outside options and contractual constraints. These tests provide a useful consistency check for the model, and they shed light on the determinants of CEO wage dynamics and bargaining power. An important caveat is that all the proxies are endogenous and I lack instruments, hence the correlations below do not have a strong causal interpretation.

I examine five proxies for the CEO's outside employment opportunities. Detailed definitions of these variables are in the Appendix, and summary statistics are in Table 1. One challenge is that most proxies for strong CEO outside options are also proxies for strong outside options of the firm, and these two effects act in opposite directions. A non-zero correlation therefore implies that one of these effects outweighs the other.

The first proxy for CEO outside options is the number of years the CEO spends in the firm before becoming CEO. CEOs who spend more years in the firm arguably have less general, transferable human capital, and hence have fewer outside opportunities. On the other hand, if this variable picks up whether the firm typically promotes CEOs from within, then it also measures the board's outside options: boards that promote from within have fewer potential replacement CEOs.

The second measure is the fraction of CEOs promoted from within the firm in the same industry. Parrino (1997) shows that some industries are more likely to promote their CEOs from within the firm. If industry firms usually hire from within, then a given CEO has fewer employment options outside his or her own firm. On the other hand, if firms hire from within, then the firm has fewer potential replacement CEOs, hence the firm's outside option is weaker.

The third proxy is the homogeneity of firms in the industry, also used by Parrino (1997) and Gillan, Hartzell, and Parrino (2009). This variable 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. CEOs likely have better outside employment options when there are similar firms in which they could work. However, the board's outside option is also stronger when it can hire replacement CEOs from similar firms.

The fourth measure is the number of "similar firms," defined as Compustat firms in the same industry with assets within 20% of the firm's assets, in the previous fiscal year. This variable proxies for the number of potential firms in which the CEO could feasibly get hired, and hence for strength of the CEO's outside options. However, it also proxies for strength of the board's outside option, since there are more potential replacement CEOs if there are more similar firms.

The fifth proxy is the number of outside directorships the CEO holds. If the CEO has more connections outside the firm, his or her outside employment options are arguably stronger. This is an appealing proxy for CEO outside options because, unlike the previous variables, it does not also proxy for the board's outside options.

I also include two contracting variables that potentially affect bargaining outcomes. The first is the value of the CEO's unvested stock and options at the end of previous fiscal year, as a fraction of the CEO's average total compensation in the previous four years. If the CEO would lose a large amount of unvested stock and options if he or she left the firm, it is more difficult for the CEO to make a credible threat to leave the firm. The second variable is an indicator for whether the CEO has an explicit employment agreement (EA), in other words, a contract that is a physical, legal document rather than an implicit understanding between the board and CEO. These data are from Gillan, Hartzell, and Parrino (2009), who show that less than half of S&P 500 CEOs have explicit EA's. Renegotiating a CEO's contract is arguably less costly when the contract is implicit rather than explicit. In these situations I expect CEOs to bid up their pay more following good news, and for boards to reduce CEO pay more after bad news.

As control variables, I also include the log of the firm's lagged assets, the fraction of shares held by institutional investors (a proxy for governance strength), the CEO's age in the first year in office (a proxy for prior uncertainty), and the log of the number of years since the firm first appeared in CRSP.

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

$$\frac{\partial\widehat{\Theta}}{\partial Z_j} = \frac{\partial\widehat{\Theta}}{\partial M}\frac{\partial M}{\partial Z_j}. \tag{30}$$

In words, the sensitivity of parameters to characteristics equals the sensitivity of parameters to moments, times the sensitivity of moments to characteristics. I measure $\partial\widehat{\Theta}/\partial M$ by perturbing one of the 12 empirical moments, re-estimating the model, measuring the change in estimated parameters, then repeating for the other 11 moments. I measure $\partial M/\partial Z_j$ using OLS regressions that interact the main variables with the characteristics above.[11] Details are in the Appendix. I multiply $\partial\widehat{\Theta}/\partial Z_j$ by the standard deviation of Z_j so its magnitude is easier to interpret. Estimates of sensitivities $\partial\widehat{\Theta}/\partial Z_j$ are in Table 4.

A CEO who spends more years inside the company before promotion is associated with lower uncertainty about ability, consistent with inside hires being more of a "known quantity." This sensitivity is significantly negative at the 6% level but not the 5% level. Insider CEOs are also associated with significantly less volatile earnings. A one standard deviation increase in years inside the company is associated with CEOs capturing 51% more of a positive surplus. This sensitivity is economically very large but statistically insignificant (p value = 0.16). A positive sensitivity would be consistent with firms having less bargaining power if they typically hire from within, because their pool of replacement CEOs is smaller.

A one standard deviation increase in the industry's fraction of CEOs hired from inside the firm is associated with CEOs capturing 58.6% more of a positive surplus. The sensitivity is economically very large and statistically significant at the 5% level. This correlation is consistent with firms having less bargaining power relative to the CEO when they typically promote CEOs from within, since these firms have fewer potential replacement CEOs. Internal promotion is also significantly associated with less uncertainty about ability, again reflecting that an internal hire is more of a known quantity.[12] The strongest result in Table 4 is that profitability is less volatile in industries that hire more from inside. This result lacks an obvious explanation and raises the concern that the inside hiring variable may be picking up the effects of an omitted, correlated variable such as industry maturity. It is difficult to address this concern without finding an instrument for this and the other characteristics.

Industry homogeneity is not significantly related to any model parameters. Parrino (1997) shows that homogeneity is negatively correlated with inside succession, so the effects of homogeneity may already be captured through the inside succession variable. Another possibility is that CEOs' and firms' outside options are both stronger in more homogenous

[11] Three characteristics (CEO years inside company, outside directorships, and CEO has explicit contract) are missing so often that I include them only in the specification used to compute their sensitivity. Therefore, 8 specifications use 8 characteristics, and 3 specifications use 9 characteristics.

[12] This specification does not control for the number of years the CEO spent inside the firm, as explained in footnote 11. Therefore, the only variable in this specification measuring insider vs. outsider is this industry-level variable.

industries, and these effects cancel each other out.

A one standard deviation increase in the number of similar firms is associated with CEOs capturing 32% more of a positive surplus. This result is consistent with CEOs having stronger outside options when there are more firms in which they can work. This sensitivity is economically large but statistically insignificant. A larger number of similar firms is correlated with more volatile profits, which again lacks an obvious explanation and raises concerns about bias from omitted variables like industry maturity or competitiveness.

A higher number of outside directorships is associated with slightly less volatile profits, but is unrelated to the other model parameters. Data on directorships are missing for almost half of the sample, so it is possible that I am not measuring these sensitivities precisely.

There is no significant relationship between CEOs' holdings of unvested stock and options and θ^{up} or θ^{down}. This result is inconsistent with the hypothesis that CEOs have less bargaining power when they stand to lose unvested wealth by leaving the firm. A potential explanation is that a CEO's new firm will pay a hiring bonus that compensates the CEO for lost, unvested wealth[13]. If this is the case, then unvested wealth does not reduce CEO bargaining power. The relation between unvested pay and earnings volatility is statistically significant but economically very small.

The only characteristic significantly related to θ^{down} (the CEO's share of a negative surplus) is the indicator for whether the CEO has an explicit employment agreement. I find negative relation, significant at the 1% level, between CEOs' share of a negative surplus and CEOs having an explicit contract. In other words, CEOs pay is more downward rigid when the CEO has an explicit contract. The negative relation makes sense: if contracts are indeed designed to insure CEOs against bad news, as in Harris and Hölmstrom (1982), then the existence of a contract should be associated with more strongly downward rigid wages. I find no significant relation between the contract indicator and CEOs' share of a positive surplus, contrary to the hypothesis that CEOs can more easily renegotiate their pay upwards when there is no explicit contract in place.

The skimming story of Bertrand and Mullainathan (2001) is an alternate explanation for why CEO pay rises after good news but does not fall after bad news. In other words, CEOs may simply grab resources in good times but insulate themselves from bad times, due to weak governance. If this skimming story is true, then the asymmetric response to good and bad news should be stronger in more weakly governed firms. I find the opposite: less institutional ownership, a proxy for weak governance, is associated with CEOs bearing more

[13]For example, (http://money.cnn.com/2007/04/05/news/companies/ford_execpay/).

costs from bad news (higher θ^{down}) and capturing fewer benefits from good news (lower θ^{up}). However, these sensitivities are not statistically significant.

Next I discuss the remaining control variables. Uncertainty about CEO ability is significantly lower in larger firms, suggesting CEOs matter less in larger firms, measured as a fraction of assets. Taylor (2010) finds a similar result. This result is plausible, since CEOs necessarily delegate more in larger firms. As expected, there is less uncertainty about older CEOs, although this result is not statistically significant. Not surprisingly, profits are more volatile in smaller and younger companies. Profits are also more volatile in companies with younger CEOs, possibly because younger CEOs take more risks or because risky companies hire younger CEOs.

To summarize, this exercise represents a first step toward using a structural model to understand the determinants of CEO wage dynamics and bargaining power. I find that the model's parameters indeed vary with certain firm, CEO, and industry characteristics. CEOs appear to have more relative bargaining power over a positive surplus in industries that hire more insiders, and CEO pay is more downward rigid when the CEO has an explicit employment agreement. Bargaining outcomes are not significantly related to the other proxies for bargaining power, although the sensitivities often have the expected sign and large magnitudes. One possible explanation is that several of these other variables proxy not just for the CEO's bargaining power but also the firm's bargaining power, and these two effects offset each other. As expected, there is less uncertainty about the ability of CEOs who have already worked in the company for several years, or in industries that typically hire insiders.

VI. Implications for Shareholder Value: CEO Deaths

The model has implications for the question, does CEO ability matter for shareholders? If shareholders capture more of the CEO's surplus, then shareholders benefit more from good news about CEO ability and, conversely, suffer more from bad news about CEO ability. I explore this issue using unanticipated CEO deaths as a natural experiment. Intuitively, the more CEO ability matters for shareholders, the more we expect stock prices to change around unanticipated CEOs deaths. If we assume the CEO's share $\theta^{up} = \theta^{down} = \theta$, then the model simplifies considerably and makes the following prediction:

Prediction 4 (death announcement return): If the CEO's share $\theta^{up} = \theta^{down} = \theta$ and the previous model assumptions hold, then the stock return in response to an unanticipated death of CEO at firm i at the beginning of period t equals

$$R_{it}^{death} = \frac{B_{it}}{M_{it}} \left(m_{i0} - m_{it-1}\right) \left(1 - \theta\right) \beta \left(\frac{1 - \beta^{T-\tau}}{1 - \beta}\right). \tag{31}$$

Proof in Internet Appendix.

This equation produces four testable hypotheses:

H1: For any value of θ, the average death announcement return is zero.

H2: If $\theta = 1$ then the death announcement return equals zero for every deceased CEO.

H3: If $\theta < 1$ then stock return volatility increases around unanticipated CEO deaths.

H4: If $\theta < 1$ then the event return is negatively correlated with $m_{it-1} - m_{0i}$, the difference between the deceased CEO's perceived ability and that of his or her replacement.

Here is the intuition for these predictions. Since the model allows no firings, the average CEO who dies in office is just as skilled as the as the average replacement CEO, so the average CEO death is neither good nor bad news to shareholders (H1). If the CEO captures his entire surplus ($\theta = 1$), then the deceased CEO's expected pay exactly offsets his or her expected contribution to profits. In this special case, shareholders are indifferent between the current CEO and his replacement, so we should never see stock prices change when a CEO unexpectedly dies (H2). If $\theta < 1$, meaning shareholders capture a positive fraction of the CEO's surplus, then the death of a high-ability CEO is bad news for shareholders (and vice-versa), hence H3 and H4.

I compare these predictions to the empirical evidence of Nguyen and Nielsen (2010), who identify 81 unanticipated deaths of CEO in listed U.S. firms from 1991-2008. They check various news sources to ensure the deaths are unanticipated and not due, for example, to prolonged illness, undisclosed cause, or suicide.

Consistent with H1, Nguyen and Nielsen (2010) find the average cumulative abnormal return from days -1 to +1 around unanticipated CEO deaths is indistinguishable from zero, with a z statistic of 0.96. H2 is easy to reject; Nguyen and Nielsen (2010) indeed find that not all event returns equal zero.

Nguyen and Nielsen (2010) find that excess stock return volatility is higher in the days around the death announcement, which is consistent with H3 and a CEO share $\theta < 1$. In addition to testing this directional prediction, I compare magnitudes from the data and the estimated model. The predicted increase in return variance comes from taking variances of both sides of equation (31); the formula is in the Internet Appendix. I compute the predicted increase using this formula and the following assumptions: parameters equal their estimated values from Table 3 using the $w^{(grant)}$ measure; the CEO's share θ is the average of $\widehat{\theta}^{up}$ and $\widehat{\theta}^{down}$; and the deceased CEO's current tenure and expected final tenure are 6 and 12 years, which equal the sample medians from Table 1. The model predicts an 11.3 percentage point increase in return volatility around the CEO's death, whereas Nguyen and Nielsen (2010) report that volatility increases by 6.8 percentage points.[14] The match is fairly close, especially considering this is an out-of-sample test for the estimated model.

We can also test H4 in terms of magnitude and not just direction. To test H4 we need a proxy for $m_{it-1} - m_{i0}$, the gap in perceived ability between the deceased and replacement CEOs. The model prescribes a proxy, namely, the deceased CEO's abnormal pay, defined as the residual from a regression of the CEO's pay on firm characteristics and the firm's contemporaneous stock return.[15] If we assume abnormal pay is the difference in expected pay between a CEO and his or her hypothetical replacement, then the model makes the following prediction, which I prove in the Internet Appendix:

$$R_{it}^{event} = \lambda \frac{Abnormal\ pay_{it-1}}{M_{it}} + u_{it} \tag{32}$$

$$\lambda = -\frac{1-\theta}{\theta}\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right), \tag{33}$$

where $u_{i,t}$ is independent of $Abnormal\ pay_{it-1}$. In other words, in a regression of death announcement returns on the deceased CEO's abnormal pay scaled by market cap, then the regression slope should equal λ, which depends on the CEO's share θ. The intuition is that, according to the model, CEOs with high perceived contribution relative to their replacement are paid more, especially if the CEO captures a large fraction θ of his or her surplus. The predicted slope λ is negative if $0 < \theta < 1$, since the death of a high-ability (hence highly paid) CEO is bad news for shareholders. Table 5 shows the predicted slope λ for multiple values of CEO share θ and years left in office $T - \tau$. The slope's magnitude is decreasing in θ, for two reasons. When shareholders' share of the surplus $1 - \theta$ is larger, differences in perceived CEO ability matter more to shareholders. Also, when CEOs capture a smaller

[14] I estimate a base-rate 3-day excess return volatility by taking the average RVAR (variance of annual excess returns) from Table 1, then multiplying by 3/365. After taking square roots, the base-case 3-day volatility is 3.5%. Nguyen and Nielsen (2010) report that the 3-day event CAR has a standard deviation of 10.3%, which is 6.8% higher than the base-rate volatility.

[15] My model assumes realized pay $w_{it} = E_t[w_{it}] + b_{it}r_{it}$. By including contemporaneous returns as a control variable, the abnormal pay regression soaks up the $b_{it}r_{it}$ term, so that the residual only captures variation in the level of pay, $E_t[w_{it}]$.

fraction θ of their perceived contribution, then a given difference in CEO pay translates into a larger difference in perceived ability.

Nguyen and Nielsen (2010) estimate the regression in equation (32), albeit with additional control variables. I compare their estimated slopes to the model's predicted slope λ, which provides a useful consistency check for the model and parameter estimates. Their estimated slopes range from –1.8 to –2.6, depending on which controls they include. These estimated slopes are significantly negative with t-statistics from –2.4 to –3.9, consistent with H4 above. In Table 5 I print an asterisk beside predicted slopes that fall within the empirical range of –1.8 to –2.6. For instance, the model produces a slope within the empirical range if CEOs capture 50% of their surplus and have 3 years remaining in office at the time of death, or if CEOs capture 70% of the surplus with 9 years remaining.[16] A CEO share θ between 50% and 70% is in the ballpark of my parameter estimates. For instance, using the parameter estimates from Table 3 corresponding to pay measure $w^{(vest)}$, I find $\widehat{\theta}^{up}$=94% and $\widehat{\theta}^{down}$=6%, which averages to 50%.

There are two main conclusions from this exercise. First, the model performs fairly well in these out of sample tests, not just in terms of directional predictions, but also in terms of magnitudes. Second, CEO ability matters for shareholders only if shareholders capture some of the CEO's surplus.

VII. Robustness: Learning about Firm Quality

So far I have assumed 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 a_{it} is unobservable and fluctuates over time, and agents learn about firm quality and CEO ability at the same time. I call uncertainty about a_{it} "firm uncertainty." I argue below that ignoring firm uncertainty biases some parameter estimates, but not the main ones of interest.

Firm profitability still follows equation (1), but now the firm fixed effect ("firm quality") evolves over time according to

$$a_{it} = \rho a_{it-1} + (1 - \rho)\, \overline{a}_i + u_{it}. \tag{34}$$

Variables η_i, a_{it}, ε_{it}, and u_{it} are all unobservable. All other parameters are known. Agents

[16] Nguyen and Nielsen's (2010) slopes likely suffer from attenuation bias, since their regressor, abnormal pay, is measured with error. The true slopes are therefore likely even more negative than –2.6, their lowest point estimate. To fit a more negative empirical slope, the model will require a lower CEO share θ.

learn about η_i (CEO ability) and a_{it} (firm quality) from realized profitability according to Bayes Rule. (For simplicity I assume there is no additional signal z.) High realized profitability will increase the beliefs about both CEO ability and firm quality, and vice versa. I assume the shocks ε_{it} and u_{it} are normally distributed and uncorrelated. Uncertainty about CEO ability is reset to σ_0 when a new CEO takes office. I consider the special case where the CEO's share $\theta^{up} = \theta^{down} = \theta$. All other model assumptions are the same as before. This extended model collapses to the main model in Section II when $\sigma_u = 0$, $\rho = 0$, and there is no prior uncertainty about firm quality.

I solve this extension numerically. Details are in the Internet Appendix. Figure 7 compares predictions from models with and without learning about firm quality. The dashed line shows the case with no learning about firm quality; this case assumes σ_u is (almost) zero and other parameters are at their estimated values from Table 3. The solid line introduces learning about firm quality by increasing σ_u to 0.06. Starting in the top left of Figure 7, raising σ_u raises the amount of firm uncertainty, as expected. Firm uncertainty exhibits almost no variation with CEO tenure, because firm uncertainty converges to a steady state level— agents never fully learn a_{it}, since it fluctuates over time. On the top right, uncertainty about CEO ability drops slower when there is learning about firm quality. The reason is that firm uncertainty makes profitability a less precise signal of CEO ability. For instance, when agents observe high realized profitability, they do not know whether it is due to luck (the ε shock), high CEO ability, or high firm quality. This third source of uncertainty is new and makes agents learn more slowly about CEO ability. The bottom left panel shows that increasing firm uncertainty shifts return volatility upwards. This is because there is a new, extra source of uncertainty about dividends. The bottom right panel shows that predicted changes in CEO pay are less sensitive to lagged returns when there is more firm uncertainty. There are two reasons why. First, higher firm uncertainty makes beliefs about CEO ability less sensitive to realized profitability, as described above. Second, returns are more sensitive to profits when there is more firm uncertainty, because high realized profits increases beliefs about future firm quality.

INSERT FIGURE 7 NEAR HERE

Since my main estimation results ignore learning about firm quality, my parameter estimates may be biased. Unfortunately, there is no obvious way to directly estimate the more general model in this section. Instead, I establish an upper bound for the amount of estimation bias in my main estimates. I set firm uncertainty to an extremely high level, and then I ask how much my parameter estimates must change to make the model still fit the data. To keep the exercise simple, I interpret the solid lines in Figure 7 as both the empirical moments and the predicted moments from the baseline model with no firm uncertainty.[17] I then add firm uncertainty by raising σ_u from zero to 0.06. This value is extremely high; it implies beliefs about firm quality have a standard deviation of roughly 9%, in units of annual profitability. The dashed line shows the new predicted moments, discussed above. The model no longer fits the empirical moments— the dashed line is far from the solid line. I can make the model fit almost as well as before by making one small change: reducing parameter σ_ϵ, the volatility of profitability, from 34% to 31%. The dotted line labeled "hand fit" shows predictions after making this change. Reducing σ_ϵ causes return volatility to drop back to the solid line (bottom left), and makes the predicted sensitivity of CEO pay to lagged returns increase to the solid line (bottom right). The latter change follows directly from the Corollary to Prediction 3. I conclude from this robustness exercise that introducing learning about firm quality has a large effect on parameter σ_ϵ (volatility of profitability), but will not necessarily change the other parameter values.

VIII. Conclusion

I solve and 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. I find that CEO pay responds asymmetrically to good and bad news about ability. CEO pay does not drop significantly after bad news. This result is consistent with the model of Harris and Hölmstrom (1982), in which firms optimally offer long-term contracts with downward rigid wages. CEO pay is especially downward rigid for CEOs with explicit contracts. After good news, compensation rises enough for the CEO to capture 54 to 94% of the positive surplus. This result suggests that CEOs have strong outside options and hence considerable bargaining power over a positive surplus, also consistent with Harris and Hölmstrom's (1982) predictions. The asymmetric response is not stronger in firms with less institutional ownership. This result suggests the asymmetric response is not a sign of weak governance, but simply reflects optimal contracting and labor market competition.

This paper takes a first step toward understanding the determinants of CEO wage dynamics by measuring how parameter values vary with firm, CEO, and industry characteristics. Since I lack instruments for these characteristics, the correlations do not have a strong causal interpretation. Finding instruments and quantifying the effects on CEO wage dynamics, potentially using an approach like the one here, is an interesting area for future work.

[17] This amounts to assuming the simple model with no firm uncertainty was fitting the data well, consistent with results in Table 2.

Appendix

Variable Definitions

CEO yrs inside company: 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.

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 year. 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 from Gillan, Hartzell, and Parrino (2009), and 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. In 28% of observations in which the number of shares owned by institutional investors is missing, I impute a zero.

CEO age in 1st year: The CEO's age when he 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.

$w_t^{(grant)}$: Realized total compensation in year t, including stock and options granted during year t. Equals Compustat variable TDC1.

$w_t^{(vest)}$: Realized total compensation in year t, including stock and options vested during year t. Computed using Execucomp data. Details in the Internet Appendix.

Estimating Sensitivity of Parameters to Characteristics

This Appendix provides additional details on measuring $\partial\widehat{\Theta}/\partial Z_j$, the sensitivity of parameter estimates to characteristic Z_j. First I describe how I measure $\partial M/\partial Z_j$, the sensitivity of the 12 moments to characteristic Z_j, holding all other characteristics constant. The 12 moments $M = [M_1 \ M_2]$ can be computed as slopes from two regressions i:

$$Y_{im} = X_{im}' M_i + \delta_{im}, \quad i = 1, 2, \tag{35}$$

where m indexes firm/year observations and X_{im} is $k_i \times 1$. For regression $i = 1$, Y_{im} equals the variance of excess returns in firm/year m, and X_{im} contains indicator variables for CEO tenure=1,2,...,10+ years. For regression $i = 2$, Y_{im} equals the scaled change in expected CEO pay, and X_{im} contains a constant, lagged excess returns times an indicator for excess return greater than the median, lagged excess returns times an indicator for return less or equal to the median, and an indicator for return above the median. M_2 contains the slopes but not the intercept or indicator from this regression.I allow each moment to depend on an $l \times 1$ vector of firm, CEO, and industry characteristics Z_m :

$$M_{im}\left(Z_m\right) = [\Gamma_{i1} \ ...\Gamma_{ij}...\Gamma_{il}] Z_m. \tag{36}$$

Each vector Γ_{ij} is $k_i \times 1$. From equation (36), $\partial M/\partial Z_j = \left[\Gamma'_{1j}\ \Gamma'_{2j}\right]$. I estimate the vectors Γ_{ij} from the following OLS model:

$$Y_{im} = (Z_m \otimes X_{im})' \left[\Gamma'_{i1}\ ...\Gamma'_{ij}...\Gamma'_{il}\right]' + \delta_{im}. \tag{37}$$

The variance of $\partial\widehat{\Theta}/\partial Z_j$ comes from taking the variance of equation (30), which yields

$$var\left(\frac{\partial\widehat{\Theta}}{\partial Z_j}\right) = \frac{\partial\widehat{\Theta}}{\partial M} var\left[\Gamma'_{1j}\ \Gamma'_{2j}\right] \frac{\partial\widehat{\Theta}'}{\partial M}. \tag{38}$$

The OLS procedure outputs the 12×12 matrix $var\left[\Gamma'_{1j}\ \Gamma'_{2j}\right]$.

REFERENCES

Albuquerque, Rui and Enrique Schroth, 2010, Quantifying private benefits of control from a structural model of block trades, *Journal of Financial Economics* 96(1): 33–55.

Alder, Simeon, 2009, In the wrong hands: Complementarities, resource allocation, and aggregate TFP. Unpublished working paper. UCLA, Los Angeles.

Bennedsen, Morton, Francisco Perez-Gonzalez, and Daniel Wolfenzon, 2008, Do CEOs matter?. Unpublished working paper. Stanford University, Palo Alto.

Bertrand, Marianne, and Sendhil Mullainathan, 2001, Are CEOs rewarded for luck? The ones without principals are, *Quarterly Journal of Economics* 116(3), 901–932.

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

Bond, Philip, Itay Goldstein, and Edward S. Prescott, 2010, Market-based corrective actions, *Review of Financial Studies* 23(2): 781–820.

Boschen, John F. and Kimberly J. Smith, 1995, You can pay me now and you can pay me later: The dynamic response of executive compensation to firm performance, *Journal of Business* 68(4), 577–608.

Clayton, Matthew C., Jay C. Hartzell, and Joshua Rosenberg, 2005, The impact of CEO turnover on equity volatility, *Journal of Business,* 78, 1779–1808.

Cornelli, Francesca, Zbigniew Kominek, and Alexander Ljungqvist, 2010, Monitoring managers: Does it matter? Unpublished working paper. London Business School, London.

Core, John E. and Wayne R. Guay, 2002, Estimating the value of employee stock option portfolios and their sensitivities to price and volatility, *Journal of Accounting Research* 40:613–630.

Core, John E., Wayne R. Guay, and Robert E. Verrecchia, 2003, Price versus non-price performance measures in optimal CEO compensation contracts, *The Accounting Review* 78(4): 957–981.

Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov, 2010, Dynamic incentive accounts. Unpublished working paper. University of Pennsylvania, Philadelphia.

Gabaix, Xavier, and Augustin Landier, 2008, Why has CEO pay increased so much?, *Quarterly Journal of Economics* 123(1): 49–100.

Gibbons, Robert and Kevin J. Murphy, 1992, Optimal incentive contracts in the presence of career concerns: Theory and evidence, *Journal of Political Economy* 100(3): 468–505.

Graham, John R., Si Li, and Jiaping Qiu, 2009, Managerial attributes and executive compensation. Unpublished working paper. Duke University, Durham.

Harris, Milton and Bengt Hölmstrom, 1982, A theory of wage dynamics, *Review of Economic Studies* 49: 315–333.

Hayes, Rachel M. and Scott Schaefer, 1999, How much are differences in managerial abilities worth? *Journal of Accounting and Economics* 27: 125–148.

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

Hölmstrom, Bengt, 1999, Managerial incentive problems: A dynamic perspective, *Review of Economic Studies* 66: 169–182.

Johnson, W. Bruce, Robert P. Magee, Nandu J. Nagarajan, and Harry A. Newman, 1985, An analysis of the stock price reaction to sudden executive deaths: Implications for the managerial labor market, *Journal of Accounting and Economics* 7: 151–174.

Jovanovic, Boyan, 1979, Job matching and the theory of turnover, *The Journal of Political Economy,* 87(5), 972–990.

Kole, Stacey, 1997, The complexity of executive compensation contracts, *Journal of Financial Economics* 43(1): 79–104.

Malmendier, Ulrike and Geoffrey Tate, 2009, Superstar CEOs, *Quarterly Journal of Economics* 124(4): 1593–1638.

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

Milbourn, Todd T., 2003, CEO reputation and stock-based compensation, *Journal of Financial Economics* 68: 233–262.

Murphy, Kevin J. and Jan Zabojnik, 2006, Managerial capital and the market for CEOs, 2006. Unpublished working paper. Queen's University, Kingston.

Nikolov, Boris and Toni Whited, 2008, Agency conflicts and cash: Estimates from a structural model. Unpublished working paper. University of Rochester, Rochester.

Nikolov, Boris, Erwan Morellec, and Norman Schuerhoff, 2008, Corporate governance and capital structure dynamics: Evidence from a structural estimation. Unpublished working paper. University of Rochester, Rochester.

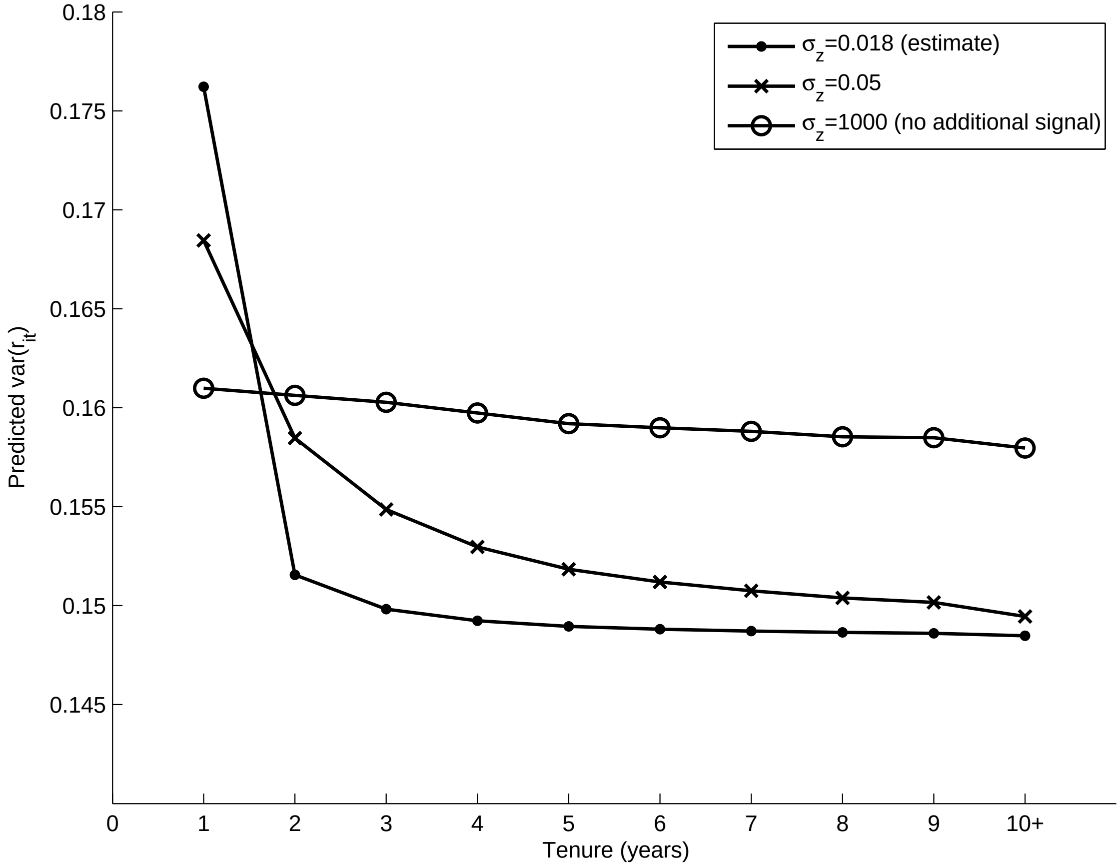
Nguyen, Bang D., and Kasper M. Nielsen, 2010, What death can tell: Are executives paid for their contributions to firm value? Unpublished working paper. Chinese University of Hong Kong, Hong Kong.

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

Taylor, Lucian A., 2010, Why are CEOs rarely fired? Evidence from structural estimation, *Journal of Finance,* forthcoming.

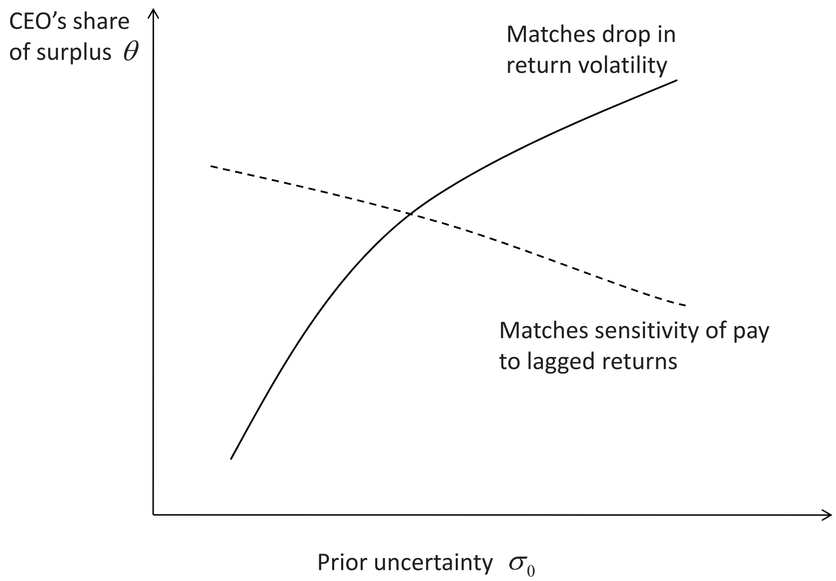
Tërvio, Marko, 2008, The difference that CEOs make: An assignment model approach, *American Economic Review,* 98(3): 642–668.

Figure 1: **Effect of the Additional Signal's Precision on Stock Return Volatility**



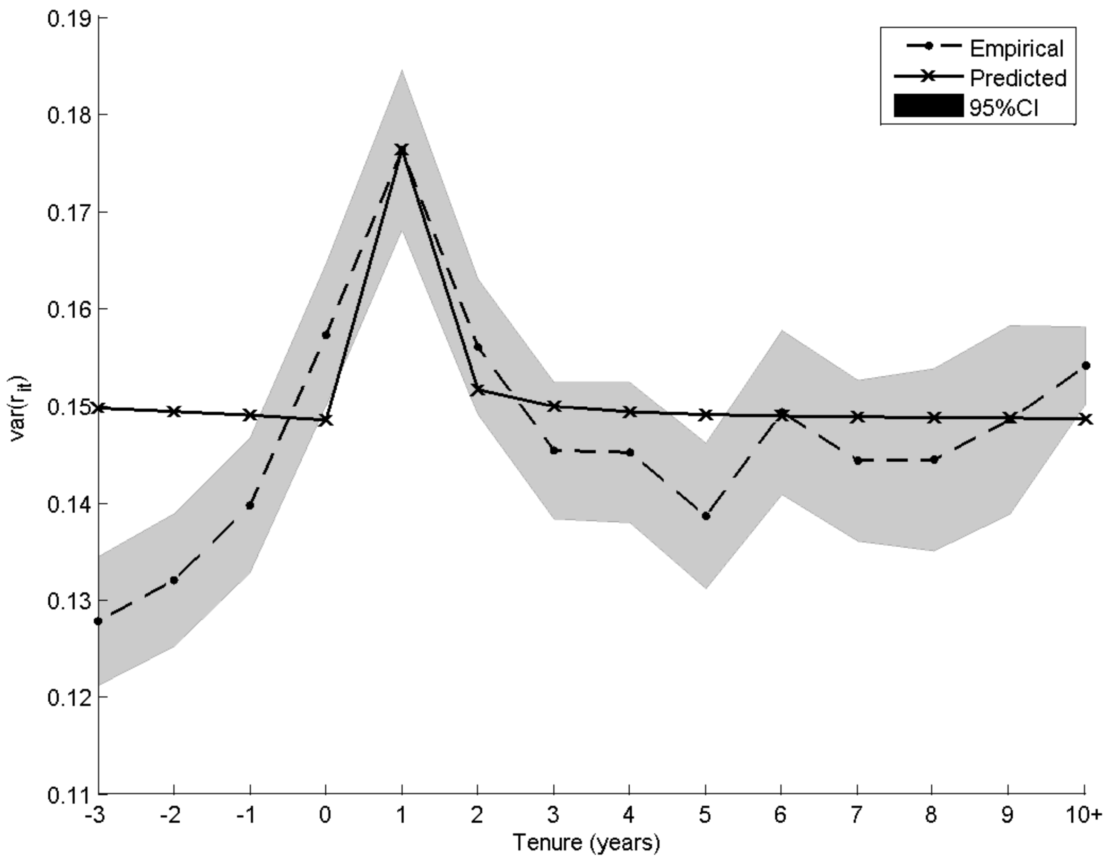
This figure plots the predicted variance of excess stock returns versus CEO tenure, for three values of σ_z, the volatility of the additional signal of CEO ability. Variance of excess return is computed using equation (A?) in the Internet Appendix. Remaining parameters (besides σ_z) are set to their estimated values from Table 3, row $w^{(grant)}$.

Figure 2: **Disentangling Prior Uncertainty and the CEO's Surplus**



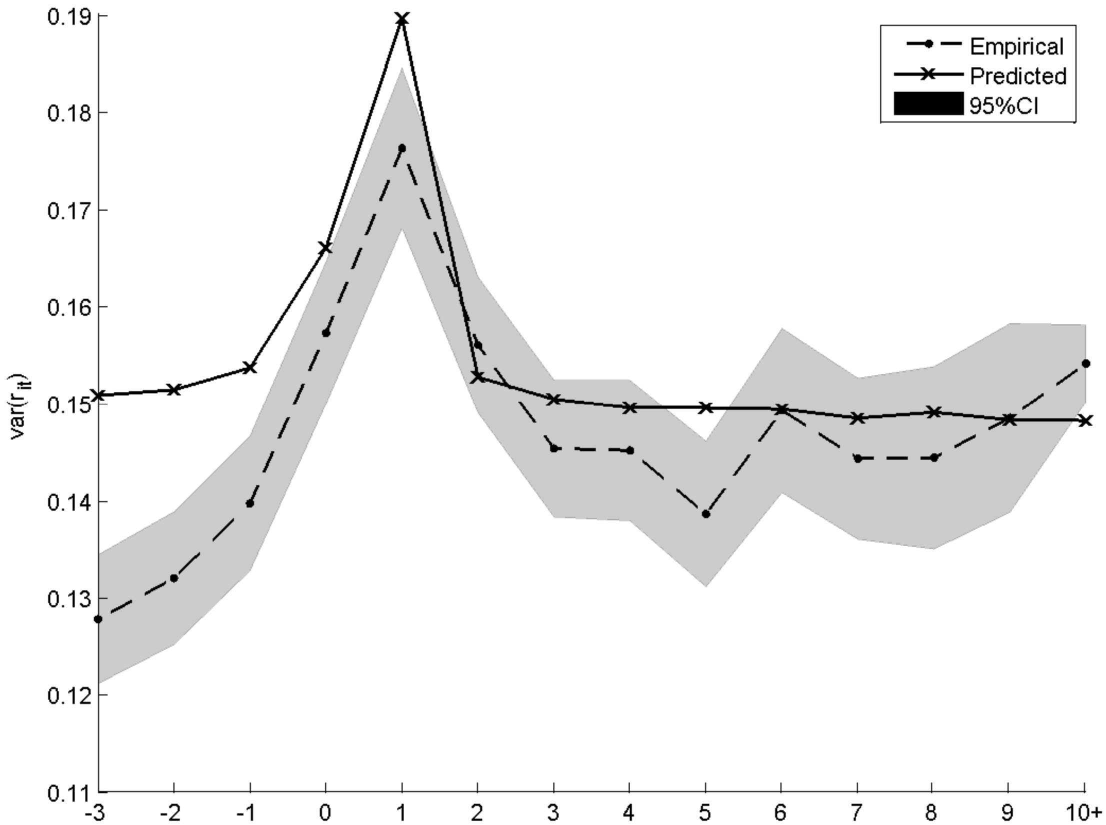
This figure illustrates the combinations of parameter values that allow the model to certain aspects of the data. The solid line plots the combinations of parameters $\theta = \theta^{up} = \theta^{down}$ (the CEO's fraction of the surplus) and σ_0 (prior uncertainty about CEO ability) that allow the model to match a given drop in excess stock return volatility. The dashed line plots the combinations of parameters θ and σ_0 that allow the model to match a given sensitivity of scaled changes in expected CEO pay to lagged excess stock returns.

Figure 3: **Stock Return Volatility and CEO Tenure**



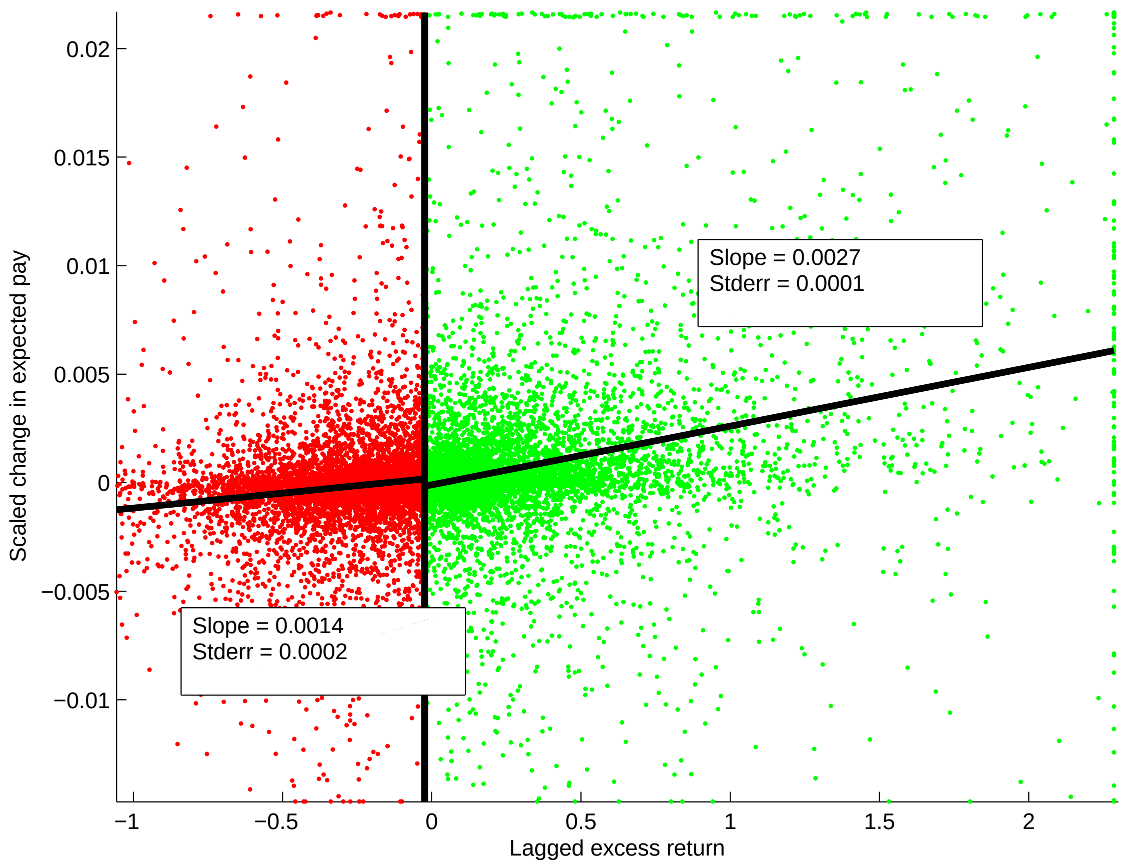
This figure plots the variance of annual excess stock returns ($var_t(r_{it})$) at various CEO tenure levels. Year 1 is the CEO's first year in office; year zero is the previous CEO's last year in office. The dashed line with its corresponding grey 95% confidence interval is computed from the empirical sample. The empirical measure is the annualized variance of weekly stock returns in excess of industry returns. The solid line plots the model's predicted return variance, using estimated parameter values in Table 3 with the w^{grant} measure.

Figure 4: **Stock Return Volatility and CEO Tenure, Allowing CEO Firings**



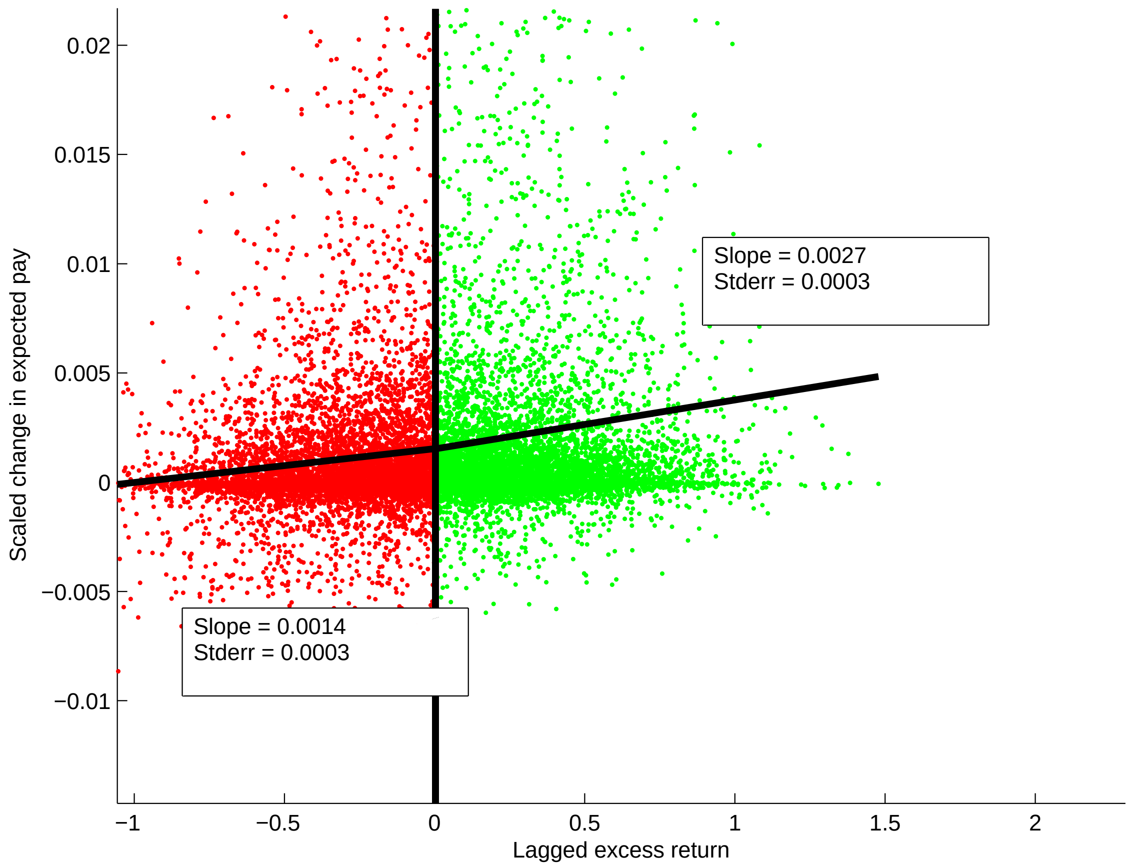
This figure is the same as figure 3, except this version of the model assumes the CEO is fired if his posterior mean ability m_t drops below $\underline{\mu} = -0.04$. All other features of the model are the same as before.

Figure 5: **Empirical Sensitivity of Pay to Lagged Returns**



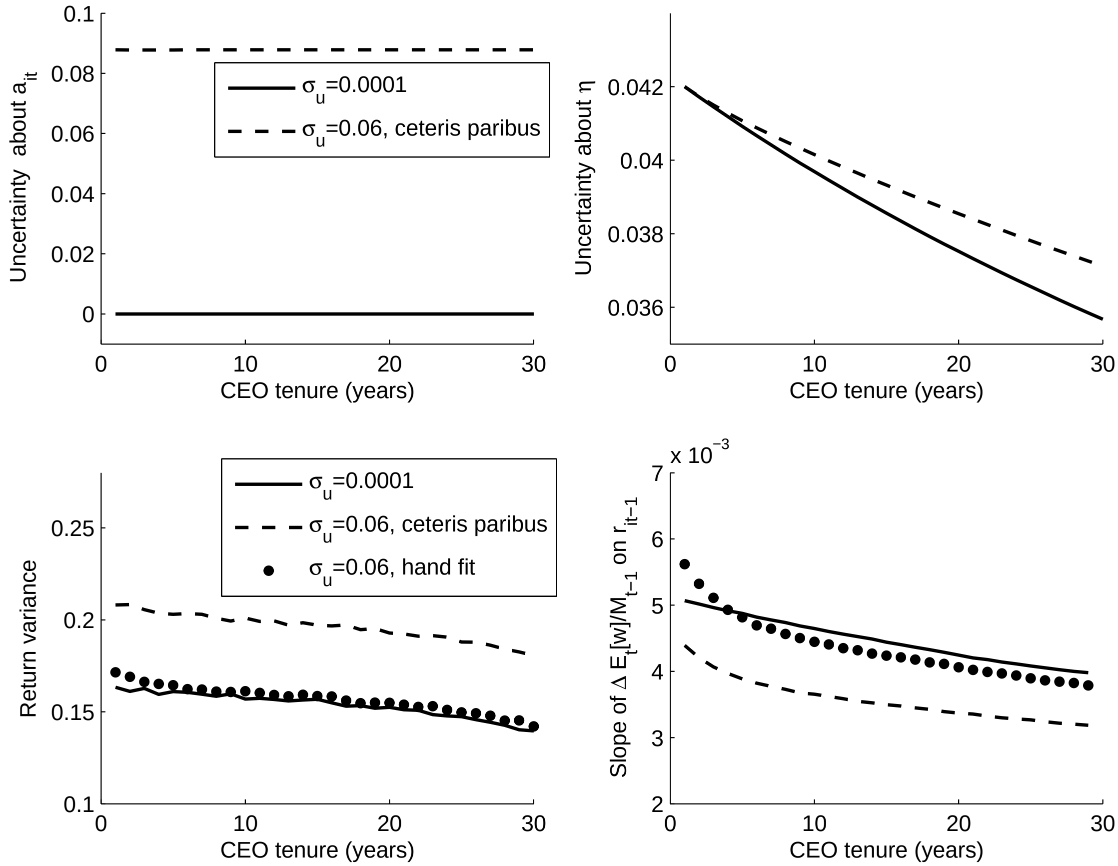
This figure plots the empirical scaled change in expected CEO pay $= \Delta E_t[w_{it}]/M_{it-1}$ versus the empirical lagged excess returns $= r_{it-1}$. Each point is a single firm/year observation from the empirical sample. The solid vertical line denotes the median lagged excess return. The other solid lines are the best-fit lines from a piecewise OLS regression split by the median lagged excess return. Text boxes show the estimated slope and its standard error.

Figure 6: **Predicted Sensitivity of Pay to Lagged Returns**



This figure is the same as Figure 5, except it shows data simulated from the estimated model instead of actual data.

Figure 7: **Model Predictions with Learning about Firm Quality**



This figure shows predictions from the model in Section VII, in which agents simultaneously learn about CEO ability and firm quality. "Uncertainty about a_{it}" is the standard deviation of beliefs about a_{it}, which is firm quality. "Uncertainty about η_i" is the standard deviation of beliefs about η_i, CEO ability. All results are from model simulations described in the Internet Appendix. The solid line uses $\sigma_u = 0.00001$, $\rho = 0.75$, and other parameter values taken from Table 3, using $w^{(grant)}$ and setting θ to the average of θ^{up} and θ^{down} . The dashed line uses the same parameter values but sets $\sigma_u = 0.06$. The dots use the same parameter values but sets $\sigma_u = 0.06$ and $\sigma_\epsilon = 0.31$.

Table 1: **Summary Statistics**

This table shows summary statistics for the sample of 4540 CEOs. Data on CEO pay come from Execucomp, 1992-2007. w_t^{vest} (w_t^{grant}) is CEO pay, including stock and option grants in the year when they vest (are granted). Expected pay is realized pay minus the portion correlated with contemporaneous excess returns. "Scaled Δ in E[pay]" equals the change in expected pay, divided by lagged market cap. Excess annual return equals annual return minus industry return. Δw_t denotes the dollar change in CEO pay over the previous year. Variance of returns is the annualized variance of industry-adjusted returns, computed from weekly returns. Current tenure is the CEO's tenure at the end of year t, and final tenure is the CEO's tenure when he/she leaves office. The remaining variables' definitions are in Section 5 and the Appendix.

Variable	Notation	N	Mean	Std	25th pctl	Median	75th pctl
Realized pay ($M)	$w_t^{(vest)}$	16344	6.01	10.11	1.17	2.51	6.05
Expected pay ($M)	$E_t[w_t^{(vest)}]$	16344	6.06	10.03	1.26	2.73	6.28
Change in E[pay] ($M)	$\Delta E_t[w_t^{(vest)}]$	12389	0.56	7.41	-1.29	0.00	1.71
Scaled Δ in E[pay]	$\frac{\Delta E_t[w_t^{(vest)}]}{M_{it-1}}$	12389	0.0008	0.0055	-0.0009	0.0000	0.0014
Realized pay ($M)	$w_t^{(grant)}$	20431	4.13	5.60	1.03	2.10	4.68
Expected pay ($M)	$E_t[w_t^{(grant)}]$	20431	4.13	5.60	1.06	2.11	4.68
Change in E[pay] ($M)	$\Delta E_t[w_t^{(grant)}]$	15760	0.119	3.863	-0.646	0.000	0.854
Scaled Δ in E[pay]	$\frac{\Delta E_t[w_t^{(grant)}]}{M_{it-1}}$	15760	0.0003	0.0040	-0.0005	0.0000	0.0007
Excess annual return	r_{it}	20636	0.014	0.472	-0.245	-0.034	0.192
Annual return	R_{it}	20636	0.185	0.516	-0.113	0.116	0.366
Variance of returns	$RVAR_{it}$	20636	0.152	0.164	0.050	0.124	0.207
Current tenure (years)	τ	20636	7.9	7.4	3.0	6.0	11.0
Final tenure (years)	T	20636	13.4	8.3	8.0	12.0	17.0
Market cap ($B)	M	20636	6.06	20.88	0.43	1.17	3.76
Assets ($B)	B	20636	10.95	53.71	0.40	1.27	4.86
CEO years inside company		11541	8.6	10.4	0.0	3.5	15.8
Fraction of CEOs insiders		20631	0.59	0.13	0.50	0.61	0.73
Industry homogeneity		19419	0.265	0.082	0.210	0.250	0.330
Number of similar firms		20525	13	11	5	10	18
Outside directorships		12034	0.54	0.86	0.00	0.00	1.00
Fraction of pay unvested		16060	0.91	1.10	0.06	0.49	1.29
CEO has explicit contract		3489	0.42	0.49	0	0	1
Institutional ownership		20630	0.435	0.311	0.112	0.463	0.693
CEO age in first year		20446	48	8	43	49	54
Ln (firm age, in years)		20636	2.8	0.9	2.2	2.9	3.5

Table 2: **Moments from SMM Estimation**

Panel A shows the 12 moments used in SMM estimation. The data moments are computed from the empirical sample of 4540 CEOs from Execucomp. The model moments are computed from the model, using the parameter estimates in Table 3. $var(r_{it})$, tenure=t is the variance of annual excess stock returns during the CEO's t-th year in office. The empirical measure is the annualized variance of weekly excess returns. The predicted value is computed from equation (IA.55) in the Internet Appendix. "Slope, beliefs up (down)" is the slope on lagged excess return times an indicator for whether the lagged return is greater than (less than or equal to) its median; the dependent variable in this regression is the scaled changes in expected pay = $\Delta E_t[w_{it}]/M_{it-1}$; the regression also includes as regressors an indicator for whether the lagged return was above or below its median. To compute the predicted slope, I simulate data from the model and estimate the same OLS regression. The p-values in Panel A test the hypothesis that the data moment equals the model moment. Panel B shows results from the test of over-identifying restrictions. Panel B's p-values jointly test whether all 12 empirical moments equal their corresponding predicted moment.

Panel A: Moments from SMM Estimation

		Results using $w^{(grant)}$			Results using $w^{(vest)}$		
	Moment	Data moment	Model moment	p-value	Data moment	Model moment	p-value
$var(r_{it})$,	tenure=1	0.176	0.176	0.984	0.176	0.178	0.984
$var(r_{it})$,	tenure=2	0.156	0.152	0.198	0.156	0.153	0.198
$var(r_{it})$,	tenure=3	0.145	0.150	0.211	0.145	0.150	0.211
$var(r_{it})$,	tenure=4	0.145	0.149	0.278	0.145	0.149	0.278
$var(r_{it})$,	tenure=5	0.139	0.149	0.011	0.139	0.148	0.011
$var(r_{it})$,	tenure=6	0.149	0.149	0.932	0.149	0.148	0.932
$var(r_{it})$,	tenure=7	0.144	0.149	0.345	0.144	0.148	0.345
$var(r_{it})$,	tenure=8	0.144	0.149	0.397	0.144	0.148	0.397
$var(r_{it})$,	tenure=9	0.148	0.149	0.975	0.148	0.148	0.975
$var(r_{it})$,	tenure=10+	0.154	0.148	0.009	0.154	0.147	0.009
Slope,	beliefs up	0.00270	0.00270	0.999	0.0050	0.0049	1.000
Slope,	beliefs down	0.00137	0.00137	0.999	0.0030	0.0031	0.999

Panel B: Test of Over-identifying Restrictions

Results using $w^{(grant)}$		Results using $w^{(vest)}$	
χ^2=	19.3	χ^2=	16.2
p-value=	0.007	p-value=	0.023

Table 3: **Parameter Estimates**

This table shows estimates of the model's parameters. The first (second) row contains results from estimating the model using data on $w^{(grant)}$ ($w^{(vest)}$), which measures CEO pay by including stock and option grants in the year when they were granted (vested). σ_0 is the standard deviation of prior beliefs about CEO ability. σ_ϵ is the volatility of shocks to firm profitability. Following good (bad) news about the CEO's ability, the CEO captures a fraction θ^{up} (θ^{down}) of the resulting positive (negative) surplus. σ_z is the volatility of the additional signal about CEO ability.

	Prior uncertainty σ_0	Volatility of profitability σ_ϵ	CEO's share, good news θ^{up}	CEO's share, bad news θ^{down}	Volatility of z signal σ_z
Using $w^{(grant)}$	0.042	0.364	0.536	0.067	0.018
	(0.003)	(0.002)	(0.097)	(0.048)	(0.013)
Using $w^{(vest)}$	0.055	0.363	0.936	0.060	0.033
	(0.006)	(0.002)	(0.058)	(0.088)	(0.015)

Table 4: **Sensitivity of Parameters to Firm, CEO, and Industry Characteristics** Panel A shows the parameter estimates from Table 3, using the CEO pay measure $w^{(grant)}$. 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 characteristic, holding other characteristics constant. t-statistics are in parentheses. These sensitivities are computed using equation (30), 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 the specification used to compute their sensitivity, so 8 specifications control for 8 characteristics, and 3 specifications control for 9 characteristics.

Panel A: Parameter Estimates (from Table 3)

	Prior uncertainty σ_0	Volatility of profitability σ_ϵ	CEO's share, good news θ^{up}	CEO's share, bad news θ^{down}	Volatility of z signal σ_z
Estimate	0.042	0.364	0.536	0.067	0.018
Standard error	(0.003)	(0.002)	(0.097)	(0.048)	(0.013)

Panel B: Sensitivity of Parameter Values to Characteristics

Characteristic	Mean (Stdev)	Change in parameter value associated with a one standard deviation increase in characteristic (t-statistic)				
CEO yrs inside company	8.64	-0.009	-0.016	0.512	0.030	-0.006
	(10.37)	(-1.92)	(-6.16)	(1.39)	(0.37)	(-0.50)
Fraction CEOs insiders	0.593	-0.009	-0.044	0.586	0.016	-0.015
	(0.133)	(-2.56)	(-21.7)	(2.27)	(0.28)	(-1.48)
Industry homogeneity	0.265	0.001	-0.001	-0.062	-0.027	0.013
	(0.082)	(0.34)	(-0.67)	(-0.25)	(-0.50)	(1.42)
Number of similar firms	12.9	0.002	0.023	0.324	-0.099	0.011
	(11.4)	(0.69)	(11.6)	(1.27)	(-1.83)	(1.16)
Outside directorships	0.541	0.000	-0.007	-0.059	-0.027	0.010
	(0.860)	(-0.03)	(-3.42)	(-0.14)	(-0.38)	(0.84)
Fraction pay unvested	0.905	0.001	0.004	0.021	-0.026	-0.017
	(1.10)	(0.36)	(2.25)	(0.07)	(-0.41)	(-1.75)
CEO has explicit contract	0.420	0.003	-0.007	0.070	-0.229	0.034
	(0.494)	(0.39)	(-2.70)	(0.13)	(-2.72)	(2.21)
Ln(company assets)	7.32	-0.013	-0.037	0.274	-0.061	0.013
	(1.83)	(-3.57)	(-17.0)	(0.98)	(-0.94)	(1.21)
Institutional ownership	0.435	-0.003	0.001	0.293	-0.051	-0.008
	(0.311)	(-1.09)	(0.85)	(1.27)	(-1.00)	(-0.97)
CEO age in 1st year	47.9	-0.004	-0.013	-0.025	0.005	0.008
	(7.9)	(-1.06)	(-6.43)	(-0.09)	(0.10)	(0.79)
Ln(firm age, in yrs)	2.80	-0.002	-0.033	0.004	0.035	0.004
	(0.90)	(-0.56)	(-15.0)	(0.01)	(0.56)	(0.43)

Table 5: **Relation between Death Returns and Abnormal CEO Pay**

This table shows the predicted slope in a regression of CEO death announcement returns on the deceased CEO's abnormal pay. The predicted slope is given by equation (33). An asterisk denotes the predicted is slope is within the empirical range estimated by Nguyen and Nielsen (2010). The table shows predicted values for different values of years left in office ($T - \tau$) and the CEO's share of the surplus (θ). Discount factor β is set to 0.9, and the book-to-market is set to its sample median, 1.06.

θ= CEO's share	Years left in office ($T - \tau$)				
of surplus	1	3	5	7	9
0.1	-8.1	-22.0	-33.2	-42.3	-49.6
0.2	-3.6	-9.8	-14.7	-18.8	-22.1
0.3	-2.1*	-5.7	-8.6	-11.0	-12.9
0.4	-1.4	-3.7	-5.5	-7.0	-8.3
0.5	-0.9	-2.4*	-3.7	-4.7	-5.5
0.6	-0.6	-1.6	-2.5*	-3.1	-3.7
0.7	-0.4	-1.0	-1.6	-2.0*	-2.4*
0.8	-0.2	-0.6	-0.9	-1.2	-1.4
0.9	-0.1	-0.3	-0.4	-0.5	-0.6
1.0	0.0	0.0	0.0	0.0	0.0

Internet Appendix for "CEO Wage Dynamics: Evidence from a Learning Model"

October 1, 2010

Contents

1 Model Solution

1.1 Excess returns (Prediction 1)

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:

$$\lim_{N_i\to\infty}\frac{1}{N_i}\sum_{k=1}^{N_i}\left(Y_{kt}-a_k-m_{kt-1}\right) = v_t+\lim_{N_i\to\infty}\frac{1}{N_i}\sum_{k=1}^{N_i}\left(\eta_k-m_{kt-1}+\varepsilon_{kt}\right) \tag{IA.1}$$
$$= v_t. \tag{IA.2}$$

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.

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 firm's expected return is $E_t\left[R_t\right]=(1-\beta)/\beta$. The unexpected return is

$$R_t-E_t\left[R_t\right]=M_t^{-1}\left(D_t-E_t\left[D_t\right]+M_{t+1}-E_t\left[M_{t+1}\right]\right) \tag{IA.3}$$

The unexpected dividend is

$$D_t-E_t\left[D_t\right] = B\left(\eta-m_{t-1}+\left(\frac{M_t}{B}\right)v_t+\varepsilon_t\right)+w_t-E_t\left[w_t\right] \tag{IA.4}$$
$$= M_t v_t+B_t\widetilde{Y}_t-b_t r_t, \tag{IA.5}$$

since (as I show later) the expected excess return r_t equals zero, implying $w_t-E_t\left[w_t\right]=b_t r_t$. Recalling from the learning results that

$$m_t=m_{t-1}+\frac{\sigma_\tau^2}{\sigma_\varepsilon^2}\widetilde{Y}_t+\frac{\sigma_\tau^2}{\sigma_z^2}\widetilde{z}_t \tag{IA.6}$$

we have

$$D_t-E_t\left[D_t\right]=M_t v_t+B\frac{\sigma_\varepsilon^2}{\sigma_\tau^2}\left(m_t-m_{t-1}\right)-B\frac{\sigma_\varepsilon^2}{\sigma_z^2}\widetilde{z}_t-b_t r_t. \tag{IA.7}$$

The surprise in future market value is

$$M_{t+1}-E_t\left[M_{t+1}\right] = E_{t+1}-E_t\left[\sum_{s=0}^{\infty}\beta^{s+1}D_{t+1+s}\right] \tag{IA.8}$$
$$= E_{t+1}-E_t\left[\sum_{s=0}^{\infty}\beta^{s+1}\left(B\eta-w_{it+1+s}\right)\right], \tag{IA.9}$$

because firm fixed effect a_i is known (hence no change in its expected value), and shocks ε and v have conditional mean zero. 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

$$M_{t+1} - E_t\left[M_{t+1}\right] = E_{t+1} - E_t\left[\sum_{s=0}^{T-\tau_t-1} \beta^{s+1}\left(B\eta - w_{t+1+s}\right)\right]. \tag{IA.10}$$

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

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

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

$$E_{t+1} - E_t\left[w_{t+1}\right] = E_{t+1}\left[w_{t+1}\right] - E_t\left[E_{t+1}\left[w_{t+1}\right]\right]. \tag{IA.11}$$

From the model we know

$$E_{t+1}\left[w_{it+1}\right] = E_t\left[w_{it}\right] + \theta_{it+1} B\left(E_{t+1}\left[\eta\right] - E_t\left[\eta\right]\right),$$

so

$$E_{t+1} - E_t\left[w_{t+1}\right] = E_t\left[w_{it}\right] + \theta_{t+1} B\left(E_{t+1}\left[\eta\right] - E_t\left[\eta\right]\right) \tag{IA.12}$$

$$-E_t\left[E_t\left[w_t\right] + \theta_{t+1} B\left(E_{t+1}\left[\eta\right] - E_t\left[\eta\right]\right)\right]. \tag{IA.13}$$

Recall that θ_{t+1} is known at the beginning of period $t+1$ but not at the beginning of t. We therefore need to treat θ_{t+1} as a random variable at time t.

$$E_{t+1} - E_t\left[w_{t+1}\right] = \theta_{t+1} B\left(m_t - m_{t-1}\right) - B E_t\left[\theta_{t+1}\left(m_t - m_{t-1}\right)\right] \tag{IA.14}$$

$$\begin{aligned} &= \theta_{t+1} B\left(m_t - m_{t-1}\right) \\ &\quad - B\theta^{up} E\left[m_t - m_{t-1} | m_t - m_{t-1} > 0\right] \Pr\left\{m_t - m_{t-1} > 0\right\} \\ &\quad - B\theta^{down} E_t\left[m_t - m_{t-1} | m_t - m_{t-1} < 0\right] \Pr\left\{m_t - m_{t-1} < 0\right\}. \end{aligned} \tag{IA.15}$$

Since the change in beliefs is normally distributed with mean zero, we have

$$\begin{aligned} E_{t+1} - E_t\left[w_{t+1}\right] &= \theta_{t+1} B\left(m_t - m_{t-1}\right) \\ &\quad - B_i\left(\frac{\theta^{up}}{2} E_t\left[m_t - m_{t-1} | m_t - m_{t-1} > 0\right]\right) \\ &\quad - B_i\left(\frac{\theta^{down}}{2} E_t\left[m_t - m_{t-1} | m_t - m_{t-1} < 0\right]\right) \end{aligned} \tag{IA.16}$$

Using the symmetry of the normal distribution,

$$\begin{aligned} E_{t+1} - E_t\left[w_{t+1}\right] &= \theta_{t+1} B\left(m_t - m_{t-1}\right) + \\ & BE_t\left[m_t - m_{t-1} | m_t - m_{t-1} > 0\right]\left(\frac{\theta^{down} - \theta^{up}}{2}\right). \end{aligned} \tag{IA.17}$$

Next I solve for the distribution of $m_t - m_{t-1}$, conditional on any time before the end of period t. Using results from the learning section, the change in beliefs $m_t - m_{t-1}$ is normally distributed with mean zero and variance

$$Var_t\left(m_t - m_{t-1}\right) = \frac{\sigma_\tau^4}{\sigma_\varepsilon^4}\left(\sigma_{\tau-1}^2 + \sigma_\varepsilon^2\right) + \frac{\sigma_\tau^4}{\sigma_z^4}\left(\sigma_{\tau-1}^2 + \sigma_z^2\right) + 2\frac{\sigma_\tau^2}{\sigma_\varepsilon^2}\frac{\sigma_\tau^2}{\sigma_z^2}\sigma_{\tau-1}^2. \tag{IA.18}$$

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

$$\kappa(\tau) \equiv E_t\left[m_t - m_{t-1} | m_t - m_{t-1} > 0\right]\left(\frac{\theta^{down} - \theta^{up}}{2}\right) \tag{IA.19}$$

$$\kappa(\tau) = \left(Var_t\left(m_{ijt} - m_{ijt-1}\right)\right)^{1/2}\phi(0)\left(\theta^{down} - \theta^{up}\right), \tag{IA.20}$$

which is deterministic, has the same sign as $\theta^{down} - \theta^{up}$, and equals zero when $\theta^{up} = \theta^{down}$. Substituting in this result, we have

$$E_{t+1} - E_t\left[w_{t+1}\right] = \theta_{t+1} B\left(m_t - m_{t-1}\right) + B\kappa(\tau). \tag{IA.21}$$

Repeating the previous steps for a general $s > 0$, it is possible to show that

$$E_{t+1} - E_t\left[w_{t+1+s}\right] = E_{t+1} - E_t\left[w_{t+s}\right] + B\left[E_{t+1} - E_t\left\{B\kappa(\tau + s),\right\}\right] \tag{IA.22}$$

$$= E_{t+1} - E_t\left[w_{t+s}\right], \tag{IA.23}$$

since $\kappa(\tau + s)$ is deterministic. Using backwards induction, it follows that

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

Plugging this result in, we have

$$M_{t+1} - E_t\left[M_{t+1}\right] = B\left(m_t - m_{t-1}\right)\beta\left(\frac{1 - \beta^{T-\tau}}{1 - \beta}\right) - \left(E_{t+1} - E_t\left[w_{t+1}\right]\right)\sum_{s=0}^{T-\tau-1}\beta^{s+1} \tag{IA.24}$$

$$= \beta\left(\frac{1 - \beta^{T-\tau}}{1 - \beta}\right) B\left[\left(1 - \theta_{t+1}\right)\left(m_t - m_{t-1}\right) - \kappa(\tau)\right], \tag{IA.25}$$

and the firm's stock return is

$$\begin{aligned}
R_t &= E[R_t] + M_t^{-1}\left(M_t v_t + B\frac{\sigma_\varepsilon^2}{\sigma_\tau^2}(m_t - m_{t-1}) - B_t\frac{\sigma_\varepsilon^2}{\sigma_z^2}(z_t - m_{t-1})\right) \qquad \text{(IA.26)}\\
&\quad + M_t^{-1}\left(-b_t r_t + \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)B\left[(1-\theta_{t+1})(m_t - m_{t-1}) - \kappa(\tau)\right]\right)\\
R_t &= E[R_t] + v_t - \frac{b_t r_t}{M_t} \qquad \text{(IA.27)}\\
&\quad + \frac{B}{M_t}\left((m_t - m_{t-1})\left[\frac{\sigma_\varepsilon^2}{\sigma_\tau^2} + \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})\right]\right)\\
&\quad + \frac{B}{M_t}\left(-\frac{\sigma_\varepsilon^2}{\sigma_z^2}(z_t - m_{t-1}) - \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau)\right)
\end{aligned}$$

The average of excess returns r_t across industry firms goes to zero in the limit as the number of industry firms becomes infinite. To see this, note

$$E_t\left[-\theta_{t+1}(m_t - m_{t-1})\right] = \kappa(\tau). \qquad \text{(IA.28)}$$

Since all firms in the industry have the same assumed expected return $E[R]$, then the average realized industry return $\overline{R}_t$ equals

$$\overline{R}_t = E[R_t] + v_t, \qquad \text{(IA.29)}$$

and the return in excess of the industry return is

$$\begin{aligned}
r_t &\equiv R_t - \overline{R}_t\\
r_t &= -\frac{b_t r_t}{M_{it}} + \frac{B}{M_t}\left(\left[\frac{\sigma_\varepsilon^2}{\sigma_\tau^2} + \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})\right](m_t - m_{t-1})\right) \qquad \text{(IA.30)}\\
&\quad + \frac{B}{M_t}\left(-\frac{\sigma_\varepsilon^2}{\sigma_z^2}(z_t - m_{t-1}) - \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau)\right)
\end{aligned}$$

Using the approximation $b_{ijt} << M_{it}$, we have the approximation

$$\begin{aligned}
r_t &\approx \frac{B}{M_t}\left(\left[\frac{\sigma_\varepsilon^2}{\sigma_\tau^2} + \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})\right](m_t - m_{t-1})\right) \qquad \text{(IA.31)}\\
&\quad - \frac{B}{M_t}\left(\frac{\sigma_\varepsilon^2}{\sigma_z^2}(\widetilde{z}_t - \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau)\right).
\end{aligned}$$

We can also write the excess return as

$$r_t \approx \frac{B}{M_t}\left(\left[\frac{\sigma_\varepsilon^2}{\sigma_\tau^2}+\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})\right]\frac{\sigma_\tau^2}{\sigma_\varepsilon^2}\widetilde{Y}_t\right) \tag{IA.32}$$

$$+\frac{B_{it}}{M_{it}}\left(\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})\frac{\sigma_\tau^2}{\sigma_z^2}\widetilde{z}_t-\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau)\right)$$

$$r_t \approx \frac{B}{M_t}\left(\left[1+\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\frac{\sigma_\tau^2}{\sigma_\varepsilon^2}(1-\theta_{t+1})\right]\widetilde{Y}_t\right) \tag{IA.33}$$

$$+\frac{B}{M_t}\left(\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})\frac{\sigma_\tau^2}{\sigma_z^2}\widetilde{z}_t-\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau)\right) \tag{IA.34}$$

$$\approx \frac{B}{M_t}\left(\widetilde{Y}_t+\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1})(m_t-m_{t-1})-\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau)\right) \tag{IA.35}$$

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

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

$$median\left(r_{it}|\tau,T,\frac{B}{M_t};\beta,\sigma_\varepsilon,\sigma_0,\sigma_z,\theta^{down},\theta^{up}\right)=-\frac{B}{M_t}\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\kappa(\tau), \tag{IA.36}$$

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

1.2 Return volatility (Prediction 2)

The proof of prediction 2 uses the following result on asymmetric random normal variables. If X is distributed as $N(0,\sigma)$ and

$$Y = \theta_+ X \text{ if } X \geq 0 \tag{IA.37}$$

$$= \theta_- X \text{ if } X < 0, \tag{IA.38}$$

then the variance of Y equals

$$Var(Y)=\sigma^2\left[\frac{\theta_+^2+\theta_-^2}{2}-\phi(0)^2(\theta_+-\theta_-)^2\right]. \tag{IA.39}$$

I use equation (IA.35) to compute return volatility. Recall from the learning section that

$$m_t - m_{t-1} = \widetilde{Y}_t \frac{\sigma_\tau^2}{\sigma_\epsilon^2} + \widetilde{z}_t \frac{\sigma_\tau^2}{\sigma_z^2} \tag{IA.40}$$

$$\sim N\left(0, \frac{\sigma_\tau^4}{\sigma_\epsilon^4}(\sigma_{\tau-1}^2 + \sigma_\epsilon^2) + \frac{\sigma_\tau^4}{\sigma_z^4}\left(\sigma_{\tau-1}^2 + \sigma_z^2\right) + 2\frac{\sigma_\tau^4}{\sigma_\epsilon^2\sigma_z^2}\sigma_{\tau-1}^2\right). \tag{IA.41}$$

Random variable θ_{t+1} depends on the sign of $(m_t - m_{t-1})$ so the variance of the second term can be determined using the result from the previous section, as follows. I introduce notation that will come in handy soon:

$$\theta_{+/-} = \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta_{t+1}) \tag{IA.42}$$

$$\theta_{+} = \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta^{up}) \tag{IA.43}$$

$$\theta_{-} = \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)(1-\theta^{down}), \tag{IA.44}$$

$$\lambda\left(T-\tau;\theta^{up},\theta^{down},\beta\right) \equiv \frac{\theta_+^2+\theta_-^2}{2} - \phi\left(0\right)^2\left(\theta_+ - \theta_-\right)^2 \tag{IA.45}$$

The variance of the second term in equation (IA.35) is therefore

$$\left(\frac{B}{M_t}\right)^2\left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}(\sigma_{\tau-1}^2+\sigma_\epsilon^2) + \frac{\sigma_\tau^4}{\sigma_z^4}\left(\sigma_{\tau-1}^2+\sigma_z^2\right) + 2\frac{\sigma_\tau^4}{\sigma_\epsilon^2\sigma_z^2}\sigma_{\tau-1}^2\right)\lambda\left(T-\tau;\theta^{up},\theta^{down},\beta\right) \tag{IA.46}$$

The variance of the first term in equation (IA.35) above is

$$\left(\frac{B_{it}}{M_{it}}\right)^2\left(\sigma_{\tau-1}^2+\sigma_\epsilon^2\right).$$

The variance of returns therefore equals

$$\begin{aligned} var_t\left(r_{it}\right) = {} & \left(\frac{B}{M_t}\right)^2\left(\sigma_{\tau-1}^2+\sigma_\epsilon^2\right) + \\ & \left(\frac{B}{M_t}\right)^2\left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}(\sigma_{\tau-1}^2+\sigma_\epsilon^2)\right)\lambda\left(T-\tau;\theta^{up},\theta^{down},\beta\right) + \\ & \left(\frac{B}{M_t}\right)^2\left(\frac{\sigma_\tau^4}{\sigma_z^4}\left(\sigma_{\tau-1}^2+\sigma_z^2\right) + 2\frac{\sigma_\tau^4}{\sigma_\epsilon^2\sigma_z^2}\sigma_{\tau-1}^2\right)\lambda\left(T-\tau;\theta^{up},\theta^{down},\beta\right) \\ & 2\frac{B}{M_t}\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)cov_t\left(\tilde{Y}_t,\left(1-\theta_{t+1}\right)\left(m_t-m_{t-1}\right)\right). \end{aligned} \tag{IA.47}$$

Most of the rest of this section is spent computing the covariance above. Let

$$Z_t = \tilde{Y}_t \frac{\sigma_\tau^2}{\sigma_\epsilon^2} + \widetilde{z}_t \frac{\sigma_\tau^2}{\sigma_z^2}.$$

We have

$$\begin{aligned}
& cov\left(\tilde{Y}_t, \theta_{+/-} Z_t\right) = E\left[\tilde{Y}_t \theta_{+/-} Z_t\right] && \text{(IA.48)} \\
= \; & E\left[\theta_+ \tilde{Y}_t Z_t \mid Z_t \geq 0\right] Pr\left\{Z_t \geq 0\right\} && \text{(IA.49)} \\
& + E\left[\theta_- \tilde{Y}_t \left(Z_t\right) \mid Z_t < 0\right] Pr\left\{Z_t < 0\right\} \\
= \; & \frac{1}{2}\theta_+ E\left[\tilde{Y}_t Z_t \mid Z_t \geq 0\right] + \frac{1}{2}\theta_- E\left[\tilde{Y}_t Z_t \mid Z_t < 0\right]
\end{aligned}$$

It is possible to show that if $(Y, \delta) \sim N(0, \Sigma)$, then

$$E[Y(a_1 Y + a_2 \delta)|a_1 Y + a_2 \delta > 0] = E[Y(a_1 Y + a_2 \delta)|a_1 Y + a_2 \delta < 0]. \tag{IA.50}$$

Substituting in this result, we have

$$cov\left(\tilde{Y}_t, \theta_{+/-}\left(Z_t\right)\right) = \frac{\theta_+ + \theta_-}{2} E\left[\tilde{Y}_t Z_t \mid Z_t \geq 0\right]$$

It is also possible to show that if $(Y, \delta) \sim N(0, \Sigma)$ then

$$\begin{aligned}
E\left[Y\delta | aY + b\delta \geq 0\right] &= E\left[Y\delta\right] = cov\left(Y, \delta\right) && \text{(IA.51)} \\
E\left[Y^2 | aY + b\delta \geq 0\right] &= E\left[Y^2\right] = var\left(Y\right) && \text{(IA.52)}
\end{aligned}$$

where $a, b > 0$. Using this result, we have

$$\begin{aligned}
& cov_t\left(\tilde{Y}_t, \theta_{+/-}\left(Z_t\right)\right) = \frac{\theta_+ + \theta_-}{2} E_t\left[\tilde{Y}_{ijt} Z_t \mid Z_t \geq 0\right] \\
= \; & \frac{\theta_+ + \theta_-}{2}\left[\frac{\sigma_\tau^2}{\sigma_\epsilon^2} Var_t\left(\tilde{Y}_t\right) + \frac{\sigma_\tau^2}{\sigma_z^2} Cov_t\left(\tilde{Y}_t, \widetilde{z}_t\right)\right] && \text{(IA.53)} \\
= \; & \frac{\theta_+ + \theta_-}{2}\left[\frac{\sigma_\tau^2}{\sigma_\epsilon^2}\left(\sigma_\varepsilon^2 + \sigma_{\tau-1}^2\right) + \frac{\sigma_\tau^2}{\sigma_z^2}\sigma_{\tau-1}^2\right], && \text{(IA.54)}
\end{aligned}$$

and return volatility equals

$$\begin{aligned}
var_t\left(r_{it}\right) = \; & \left(\frac{B}{M_t}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right) + && \text{(IA.55)} \\
& 2\left(\frac{B}{M_t}\right)^2 \left(\frac{\theta_+ + \theta_-}{2}\left[\frac{\sigma_\tau^2}{\sigma_\epsilon^2}\left(\sigma_\varepsilon^2 + \sigma_{\tau-1}^2\right) + \frac{\sigma_\tau^2}{\sigma_z^2}\sigma_{\tau-1}^2\right]\right) + \\
& \left(\frac{B}{M_t}\right)^2 \left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}(\sigma_{\tau-1}^2 + \sigma_\epsilon^2) + \frac{\sigma_\tau^4}{\sigma_z^4}\left(\sigma_{\tau-1}^2 + \sigma_z^2\right) + 2\frac{\sigma_\tau^4}{\sigma_\epsilon^2\sigma_z^2}\sigma_{\tau-1}^2\right) \lambda\left(T - \tau; \theta^{up}, \theta^{down}, \beta\right).
\end{aligned}$$

Special cases:

1. In the limit in which $\sigma_z \to \infty$ (i.e. there is effectively no additional signal):

$$\begin{aligned}
\lim_{\sigma_z \to \infty} var_t\left(r_{it}\right) &= \left(\frac{B}{M_t}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right)\left(1 + 2\frac{\theta_+ + \theta_-}{2}\frac{\sigma_\tau^2}{\sigma_\epsilon^2}\right) + && \text{(IA.56)}\\
&\quad \left(\frac{B}{M_t}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right)\left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}\lambda\left(T-\tau;\theta^{up},\theta^{down},\beta\right)\right) && \text{(IA.57)}\\
\lim_{\sigma_z \to \infty} var_t\left(r_{it}\right) &= \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right)\left(1 + 2\frac{\theta_+ + \theta_-}{2}\frac{\sigma_\tau^2}{\sigma_\epsilon^2}\right) + && \text{(IA.58)}\\
&\quad \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right)\left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}\left[\frac{\theta_+^2 + \theta_-^2}{2}\right]\right) - && \text{(IA.59)}\\
&\quad \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right)\left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}\left[\phi\left(0\right)^2\left(\theta_+ - \theta_-\right)^2\right]\right). && \text{(IA.60)}
\end{aligned}$$

2. If $\theta^{up} = \theta^{down} = \theta$, then

$$\begin{aligned}
\theta_+ &= \theta_- = \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta\right) && \text{(IA.61)}\\
\lambda\left(T-\tau;\theta^{up},\theta^{down},\beta\right) &= \theta_{+/-}^2 = \left[\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta\right)\right]^2 && \text{(IA.62)}
\end{aligned}$$

$$\begin{aligned}
var_t\left(r_{it}\right) &= \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right) + && \text{(IA.63)}\\
&\quad 2\left(\frac{B_{it}}{M_{it}}\right)^2 \beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta\right)\left[\frac{\sigma_\tau^2}{\sigma_\epsilon^2}\left(\sigma_\varepsilon^2 + \sigma_{\tau-1}^2\right) + \frac{\sigma_\tau^2}{\sigma_z^2}\sigma_{\tau-1}^2\right] + \\
&\quad \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\frac{\sigma_\tau^4}{\sigma_\epsilon^4}(\sigma_{\tau-1}^2 + \sigma_\epsilon^2)\right)\left[\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta\right)\right]^2 + \\
&\quad \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\frac{\sigma_\tau^4}{\sigma_z^4}\left(\sigma_{\tau-1}^2 + \sigma_z^2\right)\right)\left[\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta\right)\right]^2 + && \text{(IA.64)}\\
&\quad \left(\frac{B_{it}}{M_{it}}\right)^2 \left(2\frac{\sigma_\tau^4}{\sigma_\epsilon^2\sigma_z^2}\sigma_{\tau-1}^2\right)\left[\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta\right)\right]^2. && \text{(IA.65)}
\end{aligned}$$

3. If $\theta^{up} = \theta^{down} = 1$, then

$$\begin{aligned}
\theta_+ &= \theta_- = 0 && \text{(IA.66)}\\
\lambda\left(T-\tau;\theta^{up},\theta^{down}\right) &= 0 && \text{(IA.67)}\\
var_t\left(r_{it}\right) &= \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\epsilon^2\right) && \text{(IA.68)}
\end{aligned}$$

4. If $\sigma_0^2 = 0$

$$var(r_{it}) = \left(\frac{B_{it}}{M_{it}}\right)^2 \sigma_\epsilon^2 \tag{IA.69}$$

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

$$\lim_{\tau\to\infty} var_t\left(r_{it}\right) = \left(\frac{B_{it}}{M_t}\right)^2 \sigma_\varepsilon^2, \tag{IA.70}$$

Follows by noting

$$\lim_{\tau\to\infty} \sigma_{\tau-1}^2 = 0. \tag{IA.71}$$

Claim: If $\theta^{up} = \theta^{down} = \theta$, $\sigma_0^2 > 0$, and $0 \le \theta \le 1$, then $var_{it}\left(r_{it}\right)$ decreases in tenure.

Proof: Posterior variance σ_τ^2 is strictly positive and decreasing in τ if $\sigma_0^2 > 0$. The quantity $\left(1-\beta^{T-\tau}\right)$ is also decreasing in τ. The result follows by inspecting the expression for return variance, noting that quantities decreasing in τ are multiplied by weakly positive quantities. Even if $\theta = 1$, the first term in equation (IA.55) still decreases in τ.

Claim: If $\theta^{up} = \theta^{down} = \theta$, and $0 \le \theta \le 1$, $var_t\left(r_{it}\right)$ is increasing in σ_0^2.

Proof: Posterior variance σ_τ^2 is increasing in σ_0^2. Terms multiplying σ_τ 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 θ.

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

1.3 CEO pay (Prediction 3)

Using the last model assumption, we have

$$\frac{\Delta E_t\left[w_t\right]}{M_{t-1}} = \theta_t \frac{B}{M_{t-1}}\left(m_{t-1} - m_{t-2}\right)$$

Rearranging equation (IA.31) yields

$$\left(m_{t-1} - m_{t-2}\right) \approx \frac{r_{it-1}\frac{M_{t-1}}{B} + \beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\kappa\left(\tau-1\right) + \frac{\sigma_\varepsilon^2}{\sigma_z^2}\widetilde{z}_{t-1}}{\frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2} + \beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)},$$

so

$$\Delta E_t\left[w_t\right] \approx \theta_t B \frac{r_{t-1}\frac{M_{it-1}}{B_{it-1}} + \beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\kappa\left(\tau-1\right) + \frac{\sigma_\varepsilon^2}{\sigma_z^2}\widetilde{z}_{ijt-1}}{\frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2} + \beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)} \qquad \text{(IA.72)}$$

$$\frac{\Delta E_t\left[w_{ijt}\right]}{M_{it-1}} \approx \left(r_{it-1} + \widetilde{z}_{t-1}\frac{B}{M_{t-1}}\left(\frac{\sigma_\varepsilon^2}{\sigma_z^2}\right)\right)\left[\frac{\theta_t}{\frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2} + \beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)}\right] + \frac{B}{M_{t-1}}\frac{\theta_t\kappa\left(\tau-1\right)\beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)}{\frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2} + \beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)}$$

$$\frac{\Delta E_t\left[w_t\right]}{M_{t-1}} \approx \gamma r_{t-1} + \gamma\frac{B}{M_{t-1}}\left(\frac{\sigma_\varepsilon^2}{\sigma_z^2}\right)\widetilde{z}_{t-1} + g\left(\cdot\right) \qquad \text{(IA.73)}$$

$$\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t\right) = \frac{\sigma_{\tau-1}^2\theta_t}{\sigma_\varepsilon^2 + \sigma_{\tau-1}^2\beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)} \qquad \text{(IA.74)}$$

$$g\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t, \frac{B}{M_{t-1}}\right) = \frac{B}{M_{t-1}}\kappa\left(\tau-1\right)\beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t\right) \qquad \text{(IA.75)}$$

If we condition on $\Delta E_t\left[w_t\right] \geq 0$, then the randomness in θ_t is resolved: $\theta_t = \theta^{up}$. If we condition on $\Delta E_t\left[w_t\right] < 0$ then $\theta_t = \theta^{down}$. In both cases, g becomes a purely deterministic function that is uncorrelated with the return r_{t-1}.

Comparative statics for γ : If there is no learning, i.e., $\sigma_0 = 0$, then $\gamma = 0$ since $\sigma_{\tau-1}^2 = 0$. Also, in limit where τ goes to infinity then we have $\sigma_{\tau-1}^2 = 0$ and hence $\gamma = 0$. By inspection, for $\sigma_0 > 0$, slope γ is increasing in θ_t, independent of firm size M_{t-1} or B, decreasing in signal noise σ_ε^2, increasing in initial uncertainty σ_0^2, and independent of the additional signal's precision $1/\sigma_z$. Next I show that γ is also decreasing in tenure.

Claim: Assuming $\sigma_0^2 > 0$ and $\theta > 0$, then $\frac{\partial\gamma}{\partial\tau} < 0$.

Proof: Not clear from inspection, since $\frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2}$ increases with τ, and since $\left(1-\beta^{T-\tau+1}\right)$ decreases with τ. If $\theta = 1$ then γ decreases with τ by inspection. Examine denominator of

γ :

$$A \equiv \frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2} + (1-\theta)\,\frac{\beta\left(1-\beta^{T-\tau+1}\right)}{1-\beta} \tag{IA.76}$$

$$= \frac{\sigma_\varepsilon^2}{\sigma_{\tau-1}^2} + (1-\theta)\,Q \tag{IA.77}$$

$$= \frac{\sigma_\varepsilon^2\left(1+(\tau-1)\,\sigma_0^2/\sigma_\varepsilon^2\right)}{\sigma_0^2} + (1-\theta)\,Q \tag{IA.78}$$

$$\frac{\partial A}{\partial \tau} = 1 + (1-\theta)\,\frac{\partial Q}{\partial \tau} \tag{IA.79}$$

$$\frac{\partial Q}{\partial \tau} = \frac{\beta}{1-\beta}\beta^{T-\tau+1}\ln\beta \leq 0 \tag{IA.80}$$

We have $\frac{\partial \gamma}{\partial \tau} < 0$ iff

$$\frac{\partial A}{\partial \tau} > 0 \tag{IA.81}$$

$$1-(1-\theta)\,\frac{\partial Q}{\partial \tau} > 0 \tag{IA.82}$$

$$1+(1-\theta)\,\frac{\beta}{1-\beta}\beta^{T-\tau+1}\ln\beta > 0 \tag{IA.83}$$

$$(1-\theta)\,\frac{\beta}{1-\beta}\beta^{T-\tau+1}\ln\beta > -1 \tag{IA.84}$$

Can show that $0 < \beta < 1$ implies

$$-1 < \frac{\beta}{1-\beta}\beta^{T-\tau+1}\ln\beta < 0, \tag{IA.85}$$

so $\theta > 0$ implies $\frac{\partial \gamma}{\partial \tau} < 0$.

1.4 Stock returns around unanticipated CEO deaths (Prediction 4)

Assume at the beginning of year t there is a CEO in office in firm i with posterior ability m_{t-1}, expected to complete a total of T years in office before leaving, and has just completed his τ th year in office. The CEO unexpectedly dies at the beginning of year t. The firm will choose a new CEO with prior mean ability m_0.

I assume $\theta^{up} = \theta^{down} = \theta$. In this special case, it is easy to show that

$$E_t\left[w_t\right] - E_0\left[w_0\right] = B\theta\left(m_{t-1} - m_0\right), \tag{IA.86}$$

where $E_0\left[w_t\right]$ is the expected pay of a new CEO.

Immediately before the CEO's death, the firm expects to have a CEO with ability m_{t-1} in office for $T-\tau$ more years, then a CEO with expected ability m_0 for all the years after that. When the CEO dies, the firm now expects to have a CEO with expected ability m_0 in office for all future periods. The change in expected dividends (after death minus before death) therefore equals

$$\begin{aligned}
\Delta E_t\left[D_{t+s}\right] &= B\left(m_0-m_{t-1}\right)-E_t\left[w_0-w_{t+s}\right], \qquad s=0,...,T_j-\tau-1 && \text{(IA.87)}\\
&= B\left(m_0-m_{t-1}\right)-E_t\left[E_{t+s}\left[w_0-w_{t+s}\right]\right] && \text{(IA.88)}\\
&= B\left(m_0-m_{t-1}\right)-B\theta E_t\left[m_0-m_{t+s-1}\right] && \text{(IA.89)}\\
&= B\left(m_0-m_{t-1}\right)\left(1-\theta\right).
\end{aligned}$$

When the CEO dies, the market value immediately adjusts by an amount equal to the present value of the changes in expected future dividends:

$$\begin{aligned}
\Delta M_t^{death} &= \beta\sum_{s=0}^{T_j-\tau-1}\beta^s\Delta E_t\left[D_{t+s}\right] && \text{(IA.90)}\\
&= B\left(m_0-m_{t-1}\right)\left(1-\theta\right)\beta\frac{1-\beta^{T-\tau}}{1-\beta}, && \text{(IA.91)}
\end{aligned}$$

so the unanticipated return around the death equals

$$R^{death}=\frac{B}{M_t}\left(m_0-m_{t-1}\right)\left(1-\theta\right)\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right). \tag{IA.92}$$

Next I compute the variance in these returns across deceased CEOs, assuming they all share the same current and final tenure:

$$var\left(R^{death}\right)=var\left(m_{t-1}|\tau\right)\left[\frac{B}{M_t}\left(1-\theta\right)\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\right]^2. \tag{IA.93}$$

Applying Bayes' Rule, we have

$$var\left(m_{t-1}|\tau\right)=\sigma_0^2\left(\frac{\sigma_0^2\left(1/\sigma_\varepsilon^2+1/\sigma_z^2\right)}{1/\tau+\sigma_0^2\left(1/\sigma_\varepsilon^2+1/\sigma_z^2\right)}\right), \tag{IA.94}$$

so

$$var\left(R^{death}\right)=\sigma_0^2\left(\frac{\sigma_0^2\left(1/\sigma_\varepsilon^2+1/\sigma_z^2\right)}{1/\tau+\sigma_0^2\left(1/\sigma_\varepsilon^2+1/\sigma_z^2\right)}\right)\left[\frac{B}{M_t}\left(1-\theta\right)\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\right]^2. \tag{IA.95}$$

For the prediction regarding abnormal compensation, note that the model's last assumption implies

$$\frac{E_{t-1}\left[w_{t-1}\right]-E_t\left[w_0\right]}{M_t}=\theta\frac{B}{M_t}\left(m_{t-2}-m_0\right). \tag{IA.96}$$

If the CEO dies at the beginning of year t, then we do not observe his expected pay in year t. We do observe the level of pay in his last year in office, $E_{t-1}[w_{t-1}]$. Using the formulas above, we have

$$\frac{E_{t-1}[w_{t-1}] - E_t[w_0]}{M_{it}} = \theta\frac{B}{M_t}(m_{t-1} - m_0) - \theta\frac{B}{M_t}(m_{t-1} - m_{t-2}) \tag{IA.97}$$

$$\frac{B}{M_t}(m_{t-1} - m_0) = \frac{1}{\theta}\frac{E_{t-1}[w_{t-1}] - E_t[w_0]}{M_t} + \frac{B}{M_t}(m_{t-1} - m_{t-2}). \tag{IA.98}$$

Substituting into the equation for the death announcement return and labelling $E_{t-1}[w_{t-1}] - E_t[w_0] = Abnormal\ pay_{t-1}$, then we have

$$R^{death} = \lambda\frac{Abnormal\ pay_{t-1}}{M_t} + u_t \tag{IA.99}$$

$$\lambda = -\frac{1-\theta}{\theta}\beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right) \tag{IA.100}$$

$$u_t = -\frac{B}{M_t}(m_{t-1} - m_{t-2})(1-\theta)\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right). \tag{IA.101}$$

The change in beliefs in the CEO's last year in office, $(m_{t-1} - m_{t-2})$, is independent of beliefs at the beginning of the last year in office, so u_t is independent of $Abnormal\ pay_{t-1}$.

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 ($pvest_t$):

$$vovest_t = novest_t \times povest_t. \tag{IA.102}$$

A similar formula applies to shares vesting:

$$vsvest_t = nsvest_t \times psvest_t. \tag{IA.103}$$

The number of options vesting during the year is

$$novest_t = opt\_unex\_exer\_num_t \frac{ajex_t}{ajex_{t-1}} - opt\_unex\_exer\_num_{t-1} + opt\_exer\_num_t. \tag{IA.104}$$

$opt\_unex\_exer\_num_t$ 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 $opt\_unex\_exer\_num_{t-1}$ that have already vested but have not yet been exercised. An amount $novest_{jt}$ of new options vests, then the CEO gets rid of some of these options by exercising them ($opt\_exer\_num_t$), so the CEO is left with a supply $opt\_unex\_exer\_num_t$ 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 $novest$ equal to zero for fewer than 5% of observations that are negative.

The number of shares vesting during the year is given by

$$nsvest_t = stock\_unvest\_num_{t-1} - stock\_unvest\_num_t \frac{ajex_t}{ajex_{t-1}} + new\_granted\_num_t \tag{IA.105}$$

$stock\_unvest\_num_t$ 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 $\overline{S}_t$, 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):

$$K_{t-1} = S_t - \frac{opt\_unex\_exer\_est\_val_{t-1}}{opt\_unex\_exer\_num_{t-1}}. \tag{IA.106}$$

$opt\_unex\_exer\_est\_val_{t-1}$ 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 Model Extension: Learning about Firm Quality

I make the following changes in notation. For convenience I drop subscripts on several variables. $\widehat{a}_{t|s}$ and $\widehat{\eta}_{|s}$ denotes the posterior mean of a_{it} and η_i, respectively, at the end of period s. Therefore, $\widehat{\eta}_{|s} = m_{is}$ from the original notation. $\Sigma_{a_t|s}$ and $\Sigma_{\eta|s}$ are the posterior variance of beliefs about a_{it} and η_i, 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 $x_t \equiv \left[\begin{array}{cc} a_t & \eta \end{array} \right]$ follows (as long as CEO stays in office)

$$x_t = \Phi x_{t-1} + (I - \Phi) \begin{pmatrix} \overline{a} \\ 0 \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \end{pmatrix} \tag{IA.107}$$

$$\Phi = \begin{pmatrix} \rho & 0 \\ 0 & 1 \end{pmatrix} \tag{IA.108}$$

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

$$\mu_{t|t-1} = \Phi \mu_{t-1|t-1} + (I - \Phi) \begin{pmatrix} \overline{a} \\ 0 \end{pmatrix} \tag{IA.109}$$

$$\Omega_{t|t-1} = \Phi' \Omega_{t-1|t-1} \Phi + \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & 0 \end{pmatrix}. \tag{IA.110}$$

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 η_i to σ_0^2. I rewrite profitability as

$$X_t \equiv Y_t - \left(\frac{M_t}{B}\right) v_t = \mathbf{1}' x_t + \varepsilon_t \tag{IA.111}$$

where $\mathbf{1} = \left[\begin{array}{cc} 1 & 1 \end{array} \right]'$. Bayes rule states that

$$\mu_{t|t} = \Omega_{t|t}\left(\Omega_{t|t-1}^{-1}\mu_{t|t-1} + \mathbf{1}\sigma_{\varepsilon}^{-2}X_t\right). \tag{IA.112}$$

$$\Omega_{t|t} = \left[\Omega_{t|t-1}^{-1} + \mathbf{1}\sigma_{\varepsilon}^{-2}\mathbf{1}'\right]^{-1}. \tag{IA.113}$$

Before describing the numerical procedure, I provide an expression for the firm's market value. Since $\theta^{up} = \theta^{down}$, expected CEO pay can be written $E[w_t] = \theta B m_{t-1} - \psi$, where ψ is a firm-specific constant that determines the level of pay. The firm's future dividends equal

$$D_{t+s} = BY_{t+s} - w_{t+s} \tag{IA.114}$$

$$D_{t+s} = B_{t+s}\left(a_{t+s} + \eta + \left(\frac{M_{it+s}}{B_{it+s}}\right)v_{t+s} + \varepsilon_{t+s}\right) - \left(\theta B m_{t+s-1} - \psi + b_{t+s}r_{it+s}\right) \tag{IA.115}$$

To get expected dividends we need expected values of a_{it+s} :

$$a_{t+s} = \rho^{s+1}a_{t-1} + (1-\rho)\,\overline{a}\sum_{\tau=0}^{s}\rho^{\tau} + \sum_{\tau=0}^{s}u_{t+s}\rho^{s-\tau} \tag{IA.116}$$

$$E_t[a_{t+s}] = \rho^{s+1}\widehat{a}_{t-1|t-1} + \overline{a}_i\left(1-\rho^{s+1}\right) \tag{IA.117}$$

Expected future dividends therefore equal

$$E_t[D_{t+s}] = B\left(\rho^{s+1}\widehat{a}_{t-1|t-1} + \left(1-\rho^{s+1}\right)\overline{a}\right) + \psi + Bm_{t-1}(1-\theta), \quad s = 0, ..., T-\tau \tag{IA.118}$$

$$E_t[D_{t+s}] = B\left(\rho^{s+1}\widehat{a}_{t-1|t-1} + (1-\rho^{s+1})\,\overline{a}\right) + \psi + Bm_0(1-\theta), \quad s > T-\tau.$$

$$E_t[D_{t+s}] = B\left(\rho^{s+1}\widehat{a}_{t-1|t-1} + \left(1-\rho^{s+1}\right)\overline{a}\right) + \psi + Bm_0(1-\theta), \quad s > T-\tau.$$

The market-to-book ratio at the beginning of year t equals

$$\frac{M_t}{B} = \beta\sum_{s=0}^{\infty}\beta^s\left(\rho^{s+1}\widehat{a}_{t-1|t-1} + \left(1-\rho^{s+1}\right)\overline{a} + \psi/B\right) + \beta\sum_{s=0}^{T-\tau}\beta^s m_{t-1}(1-\theta) + \beta\sum_{s=T-\tau+1}^{\infty}\beta^s m_0(1-\theta) \tag{IA.119}$$

$$= \beta\left[\frac{\overline{a}+\psi/B}{1-\beta} + \rho\frac{\widehat{a}_{t-1|t-1}-\overline{a}}{1-\beta\rho}\right] + \tag{IA.120}$$

$$\beta\left[(1-\theta)\left(m_{t-1}\frac{1-\beta^{T-\tau+1}}{1-\beta} + m_0\frac{\beta^{T-\tau+1}}{1-\beta}\right)\right] \tag{IA.121}$$

I simulate 100,000 successive CEOs, assuming each spends 30 years in office. Simulation steps are as follows:

1. I start the full simulations with $a_{it} = \overline{a}_i$ and $\Sigma_{a_{it}|t-1} = \sigma_u^2$, and I throw away the first 100 CEOs in order to allow uncertainty about a_{it} to reach a steady state.

2. Simulate the true, unobservable value of a_{it} according to equation (34), drawing shocks u_{it} from its assumed distribution .

3. When a new CEO takes office, draw the CEO's true ability η_i from the prior distribution, set the off-diagonal elements of Ω to zero, and set the diagonal element of Ω corresponding to η_i equal to σ_0^2.

4. Simulate X_t according to equation (IA.111), using the true values of a_{it}, η_i, and simulated shocks ε_{it}.

5. Update the posterior mean and variance according to equations (IA.112) and (IA.113) above.

6. Compute the firm's dividends and market values according to equation (IA.114) and (IA.120). Compute returns.

4 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:

$$\log\left(1 + YearsLeft_{jt}\right) = \log a_0 + b_1 \log Age_{jt} + b_2 \log Tenure_{jt} + \varepsilon_{jt}. \qquad \text{(IA.122)}$$

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

$$\begin{aligned} \widehat{T}_j &= Tenure_{jt^*} + \widehat{a}_0 Age_{jt^*}^{\widehat{b}_1} Tenure_{jt^*}^{\widehat{b}_2} - 1 && \text{(IA.123)} \\ &= Tenure_{jt^*} + e^{12.5} Age_{jt^*}^{-2.75} Tenure_{jt^*}^{0.114} - 1, && \text{(IA.124)} \end{aligned}$$

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

5 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. Combining these missing CEO firm/years with non-missing CEO firm/years, I start with 26,812 CEO firm/years. 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 1,760 firm/year observations where BECAMECEO is missing. Next I exclude 168 CEOs (1,180 firm/years) 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 347 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 $w^{(vest)}$ 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 expected $w^{(vest)}$ in a CEO's first two years in Execucomp. I cannot compute the change in expected $w^{(grant)}$ in a CEO's first year in office. In these years when change in expected 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 1,939 firm/year records where change in expected pay measure is missing. I exclude 31 firm/years where I cannot observe or forecast the CEO's total tenure T_j. Next I throw out 876 firm/years where I cannot find the firm's lagged market cap in CRSP, and then I eliminate 43 firm/years in which $RETVAR$ is missing.

6 Estimating the Contemporaneous Pay-Performance Sensitivity

Table 1 explains how I estimate b_{it}, the sensitivity of CEO pay to contemporaneous excess stock returns.

7 CEO Tenure, Return Volatility, and the Variance of Profitability

Figure 3 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 ε_{it} using the following panel model:

$$ROA_{it} = \beta_0 + \beta_1 ROA_{it-1} + \beta_2 \log\left(Assets_{it-1}\right) + \beta_i + \beta_t + \beta_\tau + \varepsilon_{it}, \tag{IA.125}$$

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

$$\widehat{\varepsilon}_{it}^2 = \gamma_0 + \gamma_1 \log\left(Assets_{it-1}\right) + \gamma_i + \gamma_t + \gamma_\tau + u_{it}, \tag{IA.126}$$

where $\widehat{\varepsilon}_{it}^2$ is estimated from regression (IA.125), γ_i is a firm fixed effect, γ_t is a year fixed effect, and γ_τ is a CEO tenure fixed effect.

Table 2 shows the estimated tenure fixed effects γ_τ for the conditional variance of ROA. The fixed effect for tenure = 10+ years is normalized to zero. None of the tenure fixed effects is signficantly 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 ($\widehat{\varepsilon}_{it}^2$). Results are in Table 2. 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: **Estimated Pay-Performance Sensitivity** b_{it}

This table contains regression estimates used to compute b_{it}, the sensitivity of CEO pay to contemporaneous excess stock returns. I assume $b_{it} = a_0 + a_1 \log(M_{it})$. I estimate coefficients a_0 and a_1 in the following OLS regression:

$$w_{it} = c_1 w_{it-1} + (c_2 + c_3 \log(M_{it}))\, r_{it-1} + (a_0 + a_1 \log(M_{it}))\, r_{it} + u_{it}.$$

w_{it} is realized annual CEO pay. I estimate the regression using two different measures of CEO pay. $w^{(vest)}$ includes stock and option grants at the time they vest, and $w^{(grant)}$ includes them at the time they are granted. The table below presents coefficients from the OLS regression, their standard errors in parentheses, and the regression R^2 and number of observations used.

Dependent variable	Intercept	c_1	c_2	c_3	a_0	a_1	R^2	N
$w_t^{(grant)}$	3.965	0.067	-1.296	0.289	-0.413	0.164	0.068	16,214
	(0.044)	(0.002)	(0.405)	(0.056)	(0.382)	(0.056)		
$w_t^{(vest)}$	5.717	0.113	-7.656	1.514	-4.621	1.086	0.152	12,815
	(0.090)	(0.003)	(0.848)	(0.116)	(0.815)	(0.117)		

Table 2: **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.

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