The Rodney L. White Center for Financial Research

CEO Wage Dynamics: Evidence from a Learning Model

Lucian Taylor

04-10

The Wharton School
University of Pennsylvania

CEO Wage Dynamics:
Evidence from a Learning Model

Lucian A. Taylor*

September 14, 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.

* The Wharton School, University of Pennsylvania, luket@wharton.upenn.edu. I am grateful for comments from Chris Armstrong, Philip Bond, Alex Edmans, Itay Goldstein, Wayne Guay, Evgeny Lyandres (discussant), Bang Nguyen, Kasper Nielsen, Gordon Phillips (discussant), Michael Roberts, Motohiro Yogo, and seminar participants at Duke University, the FIRS Conference 2010, the Rothschild Caesarea Center Conference 2010, University of Amsterdam, University of Pennsylvania, University of Utah, and Yale University. I am grateful for financial support from the Rodney L. White Center for Financial Research.

I. Introduction

A common view in the popular press and academic literature is that CEOs have considerable power over their own compensation, due to weak governance[1]. This view has implications for how a CEO's pay changes over time as agents learn about his or her ability. For instance, a CEO may negotiate for a raise after strong profits reveal good news about his or her ability. If the CEO has enough bargaining power with the board of directors, the CEO may capture the entire surplus created by the good news, which offsets the benefits to shareholders. In contrast, if low profits reveal bad news about the CEO's ability, the board may want to negotiate for a lower level of CEO pay. If the CEO has enough bargaining power— for instance, due to a long-term contract— then the CEO may block any negotiation and keep the level of compensation unchanged. The result is that shareholders bear the full negative surplus from bad news. Harris and Holmstrom (1982) formalize this story and show that, when the worker has all the bargaining power, shareholders suffer more from bad news about a worker's ability than they benefit from good news.

There are many reasons why CEOs may have less power over their compensation than this story suggests. If the CEO's human capital is entirely firm specific, then the CEO cannot make a credible threat to leave the firm following high profits, so he cannot bid up his pay. Renegotiating long-term contracts may be costly. If the CEO's outside option is to work at a slightly smaller firm, then the current firm may get away with paying the CEO slightly less than he contributes to profits.

This paper's goal is to measure CEOs' power over their own compensation, and then examine the implications for shareholder value. My approach is to estimate a dynamic learning model. In the model, a firm's profits depend 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. The firm's stock returns depend endogenously on beliefs about CEO ability. Separate parameters control how the surplus from good and bad news about CEO ability gets split between the CEO (in the form of higher or lower pay) and shareholders (in the form of a higher or lower share price). The model nests the predictions of other dynamic learning models, including Jovanovic (1979), Harris and Hölmstrom (1982), Gibbons and Murphy (1992), and Hölmstrom (1999).

[1]E.g., Bebchuk, Fried, and Walker (2002) and Bertrand and Mullainanthan (2001). Also, "Why we need to limit executive compensation," *Business Week* November 4, 2008.

I highlight two model predictions that will help explain how I estimate the model. For certain parameter values, the model predicts that changes in expected CEO pay are positively correlated with both positive and negative lagged excess stock returns. (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.) I obtain this prediction because abnormally high (low) profits generate a high (low) abnormal return and also an increase (decrease) in the CEO's perceived ability, which in turn causes expected CEO pay to increase (decrease) in the next year. I find strong empirical support for this prediction, both for good and bad news about CEO ability. This evidence 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 about the CEO's ability. I find strong empirical support for this prediction, which was originally documented by Clayton, Hartzell, and Rosenberg (2005).

I use the general method of moments (GMM) to estimate the model's four parameters: the CEO's share of the surplus from good and bad news, the volatility in profitability, and prior uncertainty about CEO ability. Estimation uses data on excess stock returns and CEO compensation from Execucomp between 1992 and 2007. I estimate the four parameters using 12 empirical 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 expected 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. The CEO's share and prior uncertainty have a similar, positive effect on the 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 change in beliefs, and higher prior uncertainty makes beliefs move more in response to a given signal. The CEO's share and prior uncertainty have opposite effects on the drop in return volatility with CEO tenure. Intuitively, uncertainty about CEO ability translates into more uncertainty about dividends when there is either more initial uncertainty, or when shareholders capture more (and hence CEOs capture less) of this uncertainty. Since these two moments depend in different ways on these two parameters, we separately measure the two parameters.

The estimated model fits several features of the data. It exactly fits the positive relation 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. The model produces a drop in return volatility with CEO tenure, although the predicted drop is smaller than the empirical one. One shortcoming of the model is that it predicts a drop in return volatility leading up to successions, whereas in the data we see volatility rise. When I extend the model to allow endogenous CEO firings, the model produces both an increase in return volatility leading up to successions and then a drop in volatility after.

Parameter estimates indicate that CEOs have some power over their compensation, but not much. Depending on compensation measure I use, the CEO captures an estimated 20–32% of the surplus from good news about his or her ability. Shareholders capture the remaining portion. In other words, the CEO's expected pay rises just 0.20–0.32 for 1 with increases in the CEO's expected contribution to firm profits. These estimates are far from the benchmark in which CEOs have enough power to capture 100% of a positive surplus, and they are also far from the benchmark in which CEOs have no bargaining power and capture none of the positive surplus. Shareholders appear to benefit more than the CEO does from good news about the CEO's ability. The CEO bears an estimated 8–20% of the negative surplus resulting from bad news about his or her ability. In other words, the level of CEO pay drops 0.08–0.20 for 1 with decreases in the CEO's expected contribution to firm profits. This result is inconsistent with the view that CEOs have enough power to block cuts in their pay following bad news. However, these estimates are also inconsistent with the view that CEOs have no power, because shareholders rather than CEOs appear to bear most of the costs from bad news about CEO ability. While I find that CEOs benefit more from good news than they suffer from bad news (i.e. the 20% share of positive surplus exceeds the 8% share of negative surplus), this difference in not statistically significant.

These results have implications for the question, does CEO ability matters for shareholders? The model predicts that returns are more sensitive to news about CEO ability when shareholders capture more of the surplus from this news. However, this effect appears small in the estimated model: the sensitivity rises only from 69.6 to 73.5 when I increase shareholders' share of the surplus from 0 to 100%. Another way to approach the question is to ask what happens to stock prices when a CEO dies unexpectedly in office, which causes an exogenous change in the acting CEO's perceived ability. The model predicts no change in stock price if the deceased CEO has captured his entire surplus, because changes in the CEO's pay have exactly offset changes in the CEO's perceived ability. If the CEO does not capture his entire surplus, as my parameter estimates suggest, then the model predicts an increase in return volatility around the CEO's death. This directional prediction is consistent with the empirical findings of Johnson, Magee, Nagarajan, and Newman (1985), who examine a sample of 53 executives who died unexpectedly in office. However, the observed increase in return volatility around deaths is smaller than the one predicted by the estimated model. Consistent with my model's predictions and parameter estimates, Johnson et al. report an average event return of zero, and find that event returns are decreasing in a proxy for the deceased CEO's perceived ability.

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. This extension mainly affects the estimated volatility of profitability, and does not necessarily affect the CEO's estimated share of the surplus.

The paper is structured as follows. Section 2 describes related literature. Section 3 presents the learning model's assumptions and predictions. Section 4 describes the data, identification, and the GMM estimator. Section 5 presents parameter estimates and statistics on model fit. Section 6 discusses implications for shareholder value and unanticipated CEO deaths. Section 7 describes the robustness exercise, and section 8 concludes.

II. Literature

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 these models explain the magnitudes in data on executive pay. Jovanovic (1979), Hölmstrom (1999), and Gibbons and Murphy (19982) assume the worker captures 100% of the surplus from both good and bad news about ability. Harris and Hölmstrom (1982) predict the worker captures 100% of the surplus from good news about ability and none of the surplus from bad news. 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. Boschen and Smith (1995) examine the dynamics of CEO pay and find empirical results consistent with my model's predictions, as I discuss later. Edmans, Gabaix, Sadzik, and Sannikov (2009) also model the dynamics of executive pay and also find that pay should be sensitive to lagged performance. Their story is about consumption smoothing, whereas mine is about learning. Milbourn (2003) develops a model of pay and learning about ability, but his goal is to explain cross-sectional variation in stock-based compensation.

Gabaix and Landier (2008) also study how CEOs and shareholders split the surplus from the CEO's ability. Their evidence comes from the cross section, whereas this paper's evidence comes from the time series. Specifically, Gabaix and Landier examine the matching between CEO ability and firm size, whereas this paper focuses on learning about CEO ability over time. Gabaix and Landier find that CEOs capture only 2% of the surplus they create, which is even smaller than my estimates of 8–33%. We reach the same conclusion: shareholders capture the majority of the surplus. Unlike Gabaix and Landier (2008), I find that large differences in ability across CEOs are needed to fit the data. Alder (2009) applies different functional forms in a model similar to Gabaix and Landier (2008), and he finds a higher capture rate for CEOs.

Several others authors have looked at variants of the question, do CEOs matter? For instance, Bennedsen, Perez-Gonzalez, and Wolfenzen (2008) ask whether that firm performance deteriorates when CEOs are distracted by deaths or illness in their family. Bertrand and Schoar (2003) ask whether CEOs have styles that they carry from one firm to another. This paper asks whether shareholders receive any benefits from CEO ability, net of CEO pay. Murphy and Zabojnik (2007) examine this question using a model in which CEOs captures all rents from general human capital and only a fraction of rents from firm-specific human capital. Their goal is to explain why CEO pay has increased over time, and why firms increasing hire outsiders. Unlike Murphy and Zabojnik, I focus on learning and the dynamics rather than the level of CEO pay, I allow frictions besides firm-specific human capital, and I estimate the model in order to measure how rents are split between the CEO and shareholders. Malmendier and Tate (2009) find that CEOs extract higher pay after their perceived ability rises, as proxied by the CEO receiving a prestigious business award. My results are consistent with CEOs extracting benefits after an increase in perceived ability, but my results suggest that shareholders extract even more benefits than the CEO does.

III. 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 firm-specific profitability. No one can observe a CEO's ability. Rather, CEOs and shareholders alike learn about CEO ability over time by observing the firm's realized profits. When a CEO's perceived ability changes, so does his 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.

A. Assumptions

The model features firms index by i. Firms employ CEOs, indexed by j. Firms live for an infinite number of discrete periods t, which I interpret as years.

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

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

η_{ij} is the unobservable ability of CEO j in firm i. a_i reflects the contribution of non-CEO factors in firm i. I assume for now that a_i is known; for robustness later I include learning about a_i. v_t is a shock with conditional mean equal to 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. ε_{it} is an unobservable iid 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_{ij} and ε_{it} are not.

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. I define τ_{jt} as the number of years completed by CEO j as of the end of year t. For simplicity I drop the subscripts on τ.

Assumption 2: Agents start with common, normally distributed prior beliefs about the ability of CEO j in firm i: $\eta_{ij} \sim \mathcal{N}(m_{0i}, \sigma_0^2)$

Different firms i can hire from different CEO talent pools, so the prior mean skill 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 assets B_i; my model allows this type of dependence. The prior mean m_{0i} eventually drops out of the analysis. For now, I assume the CEOs all share the same prior uncertainty σ_0.

Assumption 3: Investors use Bayes' Rule to update beliefs about η_{ij} after each year. They update their beliefs by observing the firm's profitability Y_{it}.

Assumption 4: Realized pay for CEO j in firm i and year t equals

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

r_{it} is the firm's industry-adjusted stock return, which is endogenous in the model. Expected pay $E_t[w_{ijt}]$ and the pay-performance sensitivity b_{ijt} are both known at the beginning of period t. The pay-performance sensitivity b_{ijt} depends on the CEO's holdings of stock and option grants, and on the annual bonus, which are not modeled. As I interpret the model, the CEO's labor contract sets the sensitivity b_{ijt} to induce the CEO to exert an optimal amount of effort in year t. The level of pay, $E_t[w_{ijt}]$ is set so the CEO agrees to work at firm i rather than some other firm, and agrees to bear the risk from the pay-performance sensitivity.

The model makes predictions about changes over time in a CEO's expected pay, and I use these predictions to estimate the model. The model makes no predictions about the level of expected pay. Also, the model makes no predictions about the pay-performance sensitivity, b_{ijt}, which is exogenous. In other words, I do not include the underlying economic structure that determines the pay-performance sensitivity. However, correctly estimating the model requires correctly measuring b_{ijt}. I model this sensitivity, i.e. b_{ijt}, in a reduced-form manner. The reduced-form approach for b_{ijt} simplifies the model and especially the estimation, and allows the pay-performance sensitivity to depend on firm and CEO characteristics in a flexible way. For robustness, in the Appendix[2] I solve an extension where sensitivity b_{ijt} is determined endogenously. This extension adds endogenous CEO effort to the production function and introduces CEO risk aversion. I show that this model extension has only a small effect on predictions about changes in expected CEO pay, which form the basis of my estimation procedure. Therefore, in the Appendix I argue that estimation results will not change much if I make the pay-performance sensitivity endogenous.

The next assumption allows me to derive the firm's stock returns.

Assumption 5: Investors use exogenous discount factor β to discount future dividends. The firm immediately pays out any cash flows, including potential negative cash flows, as dividends.

[2] The Appendix is on the author's website.

This strong assumption makes the math easier, since it implies that book assets are constant over time. (For this reason I sometimes drop the t subscript on B_{it}.) This assumption has little effect on the estimation results, since identification does not rely on changes in firm size (Section 4.1), and since I use data on stock returns rather than earnings. Stock returns properly adjust for dividend payouts and retained earnings.

The next assumption is about expected CEO pay. Here I make a reduced-form assumption, which I motivate in the following subsection.

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

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

$$= \theta_t B_{it} (E_t [\eta_{ij}] - E_{t-1} [\eta_{ij}]) \tag{4}$$

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

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

By equation (1), the CEO's expected contribution to firm profits in period t is $B_{it} E_t [\eta_{ij}]$. The change in this expected contribution, due to learning about CEO ability η_{ij}, is $B_{it}(E_t[\eta_{ij}] - E_{t-1}[\eta_{ij}])$. 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, the CEO captures a fraction θ^{up} of this surplus, and when beliefs decrease the CEO captures a fraction θ^{down}. The CEO captures the entire surplus in the special case where $\theta^{up} = \theta^{down} = 1$. If $\theta^{up} < 1$ and $\theta^{down} < 1$, then the CEO and shareholders split the surplus resulting from a change in the CEO's perceived ability. Measuring θ^{up} and θ^{down} is a main goal of this paper.

B. Interpreting the specification for expected CEO pay

Assumption 6 about changes in expected pay is admittedly ad hoc. In this section I sketch out three rationales for this assumption. They relate to firm-specific human capital, heterogeneity in firm size, and long-term contracts. I present only one of these rationales as formal models. For the remaining two I provide intuition only. The purpose here is to illustrate that there are many economic factors that affect how the CEO and shareholders split the surplus generated by news about CEO ability. I choose the reduced-form assumption for expected pay, for two reasons. First, it is impossible to include all these factors in an economic model that can be estimated, and by choosing only certain factors I may omit other potentially important ones. Second, the goal of this project is not to compare and contrast the various economic factors affecting CEOs' bargaining power, but rather to measure their total effect, which I can do using my simple reduced-form approach.

The first rationale is about firm-specific human capital. Here I sketch a model that produces the specification for expected pay (Assumption 6) as a prediction rather than an assumption. The model here is similar to the one by Murphy and Zabojnik (2007). Details on this model and formal results are in the Appendix.

This model uses all the assumptions from the main model, except I remove Assumption 6 and add the following assumptions. The ability of CEO j running any other firm i' with the same production function as firm i is assumed equal to

$$\eta_{i'j} = \eta_{ij}\theta. \tag{7}$$

This assumption says that a fraction $1-\theta$ of CEO j's human capital is firm-specific, meaning it is useless outside his current firm. The remaining fraction θ of his human capital is general, meaning it transfers perfectly across firms. There is free entry by firms with the same production function as in equation (1), and also the same scale B_i. At no cost to himself, the CEO can move to one of these new firms. CEOs can also take a job outside the sector, where they earn a reservation wage $\underline{w}(E_t[\eta_{ij}]) \equiv B_i\theta E_t[\eta_{ij}] - \psi_i$. Firms offer a profit-maximizing take-it-or-leave-it contract to the CEO each year. The firm incurs a succession cost if it fires the CEO or the CEO quits, and this cost is sufficiently high so that it is never optimal for the firm to fire the CEO or give him an incentive to quit.

In the Appendix, I show that the firm optimally offers a contract each year with expected pay equal to the reservation wage:

$$E_t[w_{ijt}] = \theta B_{it} E_t[\eta_{ij}] - \psi_i, \tag{8}$$

so the change in expected pay is

$$\Delta E_t[w_{ijt}] = \theta B_{it}\left(E_t[\eta_{ij}] - E_{t-1}[\eta_{ij}]\right), \tag{9}$$

which is Assumption 6 in the special case where $\theta^{up} = \theta^{down} = \theta$. Intuitively, the CEO can capture no rents from his firm-specific human capital, because he cannot make a credible threat to take his firm-specific capital to a different firm. Also, I show that it is never optimal for the firm to fire the CEO (which is basically an assumption of this model), and the CEO never chooses to leave the firm (which is not an assumption). CEOs' predicted expected pay in this setup nests the predictions from models by Jovanovic (1979) and Gibbons and Murphy (1996). Those authors assume CEO ability is perfectly transferable ($\theta^{up} = \theta^{down} = 1$), so their models predict that the CEO captures the entire surplus from his ability.

The second rationale relates to differences in firm size, and is in the spirit of matching models such as Edmans, Gabaix, and Landier (2008). In these models, CEOs with higher ability match to larger firms. If CEO j leaves firm i, he can work for the next smallest firm, i', which has assets

$$B_{i'} = B_i \theta. \tag{10}$$

If $\theta < 1$ then the CEO's outside option firm is smaller than his current firm. If ability transfers perfectly across firms, then the CEO's expected contribution to profits in his outside option firm i' is

$$E_t \left[B_{i'} \eta_{ij} \right] = \theta E_t \left[B_i \eta_{ij} \right]. \tag{11}$$

The CEO's outside option i' is only willing to pay him up to $\theta E_t \left[B_i \eta_{ij} \right]$, so the CEO's current firm i can get away with paying him an amount depending on $\theta E_t \left[B_i \eta_{ij} \right]$. Intuitively, there is no perfect substitute for firm i, which allows firm i's shareholders to capture the surplus $(1-\theta) E_t \left[B_i \eta_{ij} \right]$ resulting from its size advantage over the next-smallest firm.

The last rationale relates to long-term contracts. Risk-averse CEOs may want to insure against future changes in their pay, which requires insuring against changes in their perceived ability. In other words, CEOs may want their expected pay to change less than one-for-one with their perceived contribution to firm profits, which corresponds to values of θ^{up} and θ^{down} less than 1 in my model. The learning model of Harris and Hölmstrom (1982) makes a prediction along these lines. Their model allows the firm to offer the worker a long-term contract. They show that the firm optimally insures the worker against bad news about his ability, so the worker's pay never drops; this prediction corresponds to $\theta^{down} = 0$ in my model. Harris and Hölmstrom assume the worker cannot pledge his human capital to the firm, so the worker can threaten to leave and renegotiate his labor contract following good news about his ability. Since the worker's ability is transferable across firms, the worker captures 100% of the surplus from good news about his ability; this prediction corresponds to $\theta^{up} = 1$ in my model. In sum, my model makes predictions similar to those in Harris and Hölmstrom (1982) in the special case where $\theta^{down} = 0$ and $\theta^{up} = 1$.

C. 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

$$\eta_{ij} \sim N \left(m_{ijt}, \sigma_\tau^2 \right), \tag{12}$$

where t denotes calendar time and τ_{ijt} is the number of years CEO j has completed in office at the end of year t. (I usually drop the subscripts on τ for convenience.) Agents update their beliefs about CEO ability by observing the mean-zero surprise from profits:

$$\widetilde{Y}_{ijt} \equiv Y_{ijt} - a_i - \left(\frac{M_{it}}{B_{it}}\right) v_t - m_{t-1} = \eta_{ij} + \varepsilon_{it} - m_{ijt-1} \tag{13}$$

Applying Bayes' rule, the posterior variance follows

$$\sigma_\tau^2 = \sigma_0^2 \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \tau\sigma_0^2}\right), \tag{14}$$

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

$$m_{ijt} = m_{ijt-1} + \frac{\sigma_\tau^2}{\sigma_\varepsilon^2}\widetilde{Y}_{it} \tag{15}$$

The weight $\sigma_\tau^2/\sigma_\varepsilon^2$ on the signal declines with tenure and equal zero in the limit where tenure goes to infinity.

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

$$\Delta E_t\left[w_{ijt}\right] = \theta_t B_{it}\left(m_{ijt-1} - m_{ijt-2}\right). \tag{16}$$

Substituting in equation (15) yields

$$\Delta E_t\left[w_{ijt}\right] = \theta_t B_{it}\frac{\sigma_{\tau-1}^2}{\sigma_\varepsilon^2}\widetilde{Y}_{it-1}. \tag{17}$$

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

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

$$D_{it} = B_{it}Y_{it} - w_{ijt}, \tag{18}$$

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

$$M_{it} = E_t\left[\sum_{s=0}^{\infty}\beta^{s+1}D_{it+s}\right]. \tag{19}$$

From these equations I derive an expression for the firm's excess stock return. I show that the excess return is proportional to the earnings surprise $\widetilde{Y}_{it}$, so we can use equation (17) to relate returns and changes in expected CEO pay.

D. 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 Appendix.

Stock returns

Prediction 1: The excess stock return (firm minus industry) in year t equals[3]

$$r_{it} \approx \frac{B_{it}}{M_{it}}\left[1+\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\frac{\sigma_\tau^2}{\sigma_\varepsilon^2}\left(1-\theta_{t+1}\right)\right]\widetilde{Y}_{it}-median\left(r_{it}\right) \quad (20)$$

$$r_{it} \approx \frac{B_{it}}{M_{it}}\left[\frac{\sigma_\varepsilon^2}{\sigma_\tau^2}+\beta\left(\frac{1-\beta^{T-\tau}}{1-\beta}\right)\left(1-\theta_{t+1}\right)\right]\left(m_{ijt}-m_{ijt-1}\right)-median\left(r_{it}\right) \quad (21)$$

The expected excess return is zero by construction. Excess returns are skewed when and only when $\theta^{up} \neq \theta^{down}$, in which case θ_{t+1} is correlated with unexpected profitability $\widetilde{Y}_{it}$.[4] The Appendix contains an expression for the predicted median excess return, $median\left(r_{it}\right)$, which is a function of current and final tenure, depends on all model parameters, and equals zero when $\theta^{up} = \theta^{down}$. The equations above show that excess returns depend on unexpected profitability $\widetilde{Y}_{it}$ or, equivalently, the change in beliefs about CEO ability $\left(m_{ijt} - m_{ijt-1}\right)$. In the Appendix I show that upward (downward) revisions in perceived CEO ability coincide with excess returns r_{it} above (below) their predicted median. This prediction holds even when the CEO captures his entire surplus, i.e. when $\theta^{up} = \theta^{down} = 1$.

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. To see why, imagine the CEO's share θ is below 1. Good news about CEO ability this year coincides with a high dividend this year but also higher 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 where 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 coincides with a high dividend this year but not a higher market value at the end of the year.

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

Prediction 2:

1. In the special case with no learning, i.e., $\sigma_0^2 = 0$, or in the limit when tenure goes to

[3]The relation is approximate because I assume $b_{ijt} << M_{it}$, as explained in the previous section.

[4]Note $\theta_{t+1} = \theta^{up}$ when $\widetilde{Y}_{it} \geq 0$ and $\theta_{t+1} = \theta^{down}$ when $\widetilde{Y}_{it} < 0$.

infinity, then the variance of excess stock returns equals

$$var_t\left(r_{it}\right) = \left(\frac{B_{it}}{M_{it}}\right)^2 \sigma_\varepsilon^2. \tag{22}$$

In the case with no learning, return volatility is constant over time.

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

$$var_t\left(r_{it}\right) = \left(\frac{B_{it}}{M_{it}}\right)^2 \left(\sigma_{\tau-1}^2 + \sigma_\varepsilon^2\right), \tag{23}$$

where $\sigma_{\tau-1}^2$ is the uncertainty about CEO ability at the beginning of year t, given in equation (14).

3. If $\sigma_0^2 > 0$, $\theta^{up} \leq 1$, and $\theta^{down} \leq 1$, then the variance of stock returns decreases with CEO tenure, increases with prior uncertainty σ_0, and decreases with θ^{up} and θ^{down}.

Higher uncertainty about CEO ability leads to more uncertainty about dividends and hence higher return volatility. Higher uncertainty occurs when there is more prior uncertainty σ_0^2 or when fewer years of learning have occurred. Consistent with this prediction, empirical evidence from Clayton, Hartzell, and Rosenberg (2005) indicates that stock return volatility increases following CEO succession. 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 from shocks to cash flows, which have volatility σ_ε^2. In the special case where $\theta^{up} = \theta^{down} = 1$, uncertainty about CEO ability affects only the riskiness of this period's net profits, because the CEO pay in future periods will exactly offset changes in expected gross profits resulting from changes in the CEO's perceived ability. Return volatility is decreasing in 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.

CEO pay

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

$$\frac{\Delta E_t\left[w_{ijt}\right]}{M_{it-1}} \approx r_{it-1}\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t\right) + g\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t, \frac{B_{it-1}}{M_{it-1}}\right) \tag{24}$$

$$\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t\right) = \frac{\sigma_{\tau-1}^2\theta_t}{\sigma_\varepsilon^2 + \sigma_{\tau-1}^2\beta\left(\frac{1-\beta^{T-\tau+1}}{1-\beta}\right)\left(1-\theta_t\right)}. \tag{25}$$

These equations use the approximation that the pay-performance sensitivity b_{ijt} is much less than the firm's market value, which I confirm empirically. The scaled change in expected CEO pay equals a function g (expression in the Appendix) plus the lagged excess stock return times the sensitivity function γ. The sensitivity γ depends on the current and final tenure, θ_t (which equals θ^{up} if beliefs increased in period $t-1$ and equals θ^{down} otherwise), and other model parameters.

Equation (24) predicts that the change in expected CEO pay depends on the firm's lagged excess return. I examine several special cases to help explain the intuition.

The first case assumes the CEO receives the entire surplus from changes in beliefs, so $\theta^{up} = \theta^{down} = 1$. In this special case the equations above simplify to

$$\frac{\Delta E_t\left[w_{ijt}\right]}{M_{it-1}} = r_{it-1}\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t = 1\right) \tag{26}$$

$$\gamma\left(\tau, T; \beta, \sigma_\varepsilon, \sigma_0, \theta_t = 1\right) = \frac{\sigma^2_{\tau-1}}{\sigma^2_\varepsilon} \tag{27}$$

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 $t-1$. The profits have two effects: (1) positive excess stock returns in year $t-1$; and (2) an increase in the CEO's perceived ability, which causes expected CEO pay to rise in for period t, which means $\Delta E_t\left[w_{ijt}\right]$ is positive. Therefore, we have a positive correlation between change in expected pay and last period's excess stock return r_{it-1}.

Another special case is where 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 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 receives none of the surplus from changes in his perceived ability. In this case we obtain $\gamma = 0$, because the CEO's expected pay does not change over time (from Assumption 6).

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.

I also obtain the following more general result:

Corollary to Prediction 3: If $\sigma_0^2 > 0$ and $\theta_t > 0$, then the sensitivity $\gamma(\cdot)$ is increasing in prior uncertainty σ_0, decreasing in profit volatility σ_ε, increasing in the CEO's share θ_t, decreasing in tenure τ, independent of firm size, and approaches zero in the limit where tenure τ goes to infinity.

The intuition for these comparative static results is similar to the intuition above. CEO pay moves more with lagged stock returns when there is more uncertainty, because higher uncertainty causes beliefs to move more in response to any given signal. Therefore the slope γ is higher when there is more initial uncertainty (σ_0) and when fewer periods of learning have occurred (lower tenure). In the limit where CEO tenure goes to infinity, there is zero uncertainty about ability, so beliefs do not change at all, and hence expected pay does not change either. Expected CEO pay moves less with lagged stock returns when cash flows are 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 (24). Of course, if model parameters like θ^{up} and σ_0 depend on firm size, then this last prediction may change.

IV. Estimation

First I estimate the pay-performance sensitivity b_{ijt} using a simple OLS regression. Then I estimate the four remaining parameters using GMM. These parameters include the volatility of profitability shocks σ_ε, 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. In this draft I allow heterogeneity in b_{ijt} but assume other parameters are constant across firms, CEOs, and years. In future drafts I will allow these parameters to depend on observable characteristics like firm size and whether the CEO is an insider or outsider. Estimation uses annual data on realized CEO pay, stock returns, and stock return volatility. Before describing the data and estimator in detail, I provide intuition for how the model is identified.

A. Identification

To see how the pay-performance sensitivity b_{ijt} is identified, recall assumption 4:

$$w_{ijt} = E_t [w_{ijt}] + b_{ijt} r_{it}. \tag{28}$$

Since excess return r_{it} is orthogonal to expected pay $E_t [w_{ijt}]$, we can estimate b_{ijt} from a simple regression of CEO pay w_{ijt} on the contemporaneous excess return r_{it} and information known at the beginning of period t. Using the estimate $\widehat{b}_{ijt}$, we can estimate the level of pay as

$$\widehat{E}_t [w_{ijt}] = w_{ijt} - \widehat{b}_{ijt} r_{it}. \tag{29}$$

The volatility of profitability, σ_ε, is identified off of excess stock return volatility for long-tenured CEOs. As explained in the previous section, return volatility drops with CEO tenure and approaches a limit that depends on only on parameter σ_ε.

Disentangling 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 where $\theta^{up} = \theta^{down} = \theta$, then I explain how I separately estimate θ^{up} and θ^{down}. The drop in return volatility during a CEO's tenure depends mainly on σ_0 and θ, but not on the other parameters. Figure 1 plots all the combinations of θ and σ_0 that allow the model to match the empirical change in return volatility (the difference between return volatility in tenure year zero and in years 10+). (To produce the figure I hold constant σ_ε constant at its estimated.) There are an infinite number of pairs $\{\theta, \sigma_0\}$ that match this particular empirical moment. For instance, the model can match the moment if there is high uncertainty σ_0 (which increases the drop in volatility), but a large surplus 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 this empirical moment.

INSERT FIGURE 1 HERE

We need at least one additional empirical moment to uniquely identify θ and σ_0. An obvious candidate is γ, the sensitivity of changes in expected CEO pay to lagged stock returns. From equation (25), the sensitivity 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 (comparative statics results in previous section). Since the parameters drive these moments in different directions, we can use the two moments to uniquely pin down the two parameters. Figure 1 plots the combinations of σ_0 and θ that allow the model to match the empirical value of the sensitivity γ. The model matches this moment either by choosing high uncertainty σ_0 and low θ (so beliefs change a lot over time, but the changes do not result in large changes in CEO pay), or by choosing low uncertainty and high θ (so beliefs change less over time, but most of the changes get passed on to the CEO). The lines in the figure cross at a unique point. In other words, 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 estimated beliefs increased or decreased (details below). In the subsample where beliefs decreased, the sensitivity of expected pay to lagged returns (γ) depends on ϕ^{down} but not ϕ^{up} (equation (25) above), and vice versa for the subsample where beliefs increased. By measuring the sensitivity γ in these two subsample, I can separately measure ϕ^{up} and ϕ^{down}.

In theory, we could estimate the model using data on firm profitability[5]. I choose to estimate the model using stock returns for two reasons. First, doing so makes results less sensitive to the assumed earnings specification (equation (1)). Second, the model assumes investors learn about CEO ability from realized earnings only, whereas in reality investors also learn from additional signals, e.g., about growth opportunities. These additional signals show up in stock returns, so using stock return data mitigates this shortcoming of the model[6].

B. Data

Data come from Execucomp, CRSP, Compustat, and Kenneth French's website. The sample includes CEOs in the Execucomp database from 1992-2007. Execucomp includes firms in the S&P 1500 as well as firms removed from the index that are still trading, and some client requests. I do not eliminate CEOs who had not left office by the end of 2007 or CEOs whose terms began before 1992. The Appendix provides details on how I clean the data.

Annual stock returns are computed from monthly CRSP returns and information on firms' fiscal calendars. Industry annual returns are computed from Kenneth French's 49 equal-weighted industry portfolios, and are computed from monthly data in order 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 annual CEO pay, w_{ijt}. It is now common to measure

[5] Within-CEO time series volatility in earnings could pin down σ_ε. Cross-CEO variation in earnings could pin down σ_0. The sensitivity of pay to lagged firm profitability would pin down θ.

[6] Taylor (2009) takes an alternate approach, allowing agents to learn from both earnings and an orthogonal, latent signal. Taylor (2009) estimates the model, including the latent signal's precision, using earnings data.

pay as total compensation plus changes in CEO wealth (e.g. Core, Guay, and Verrecchia (2003)). This measure is ideal for studying CEO incentives. Since this paper studies the CEO labor market rather than CEO incentives, I argue that we need a different measure of pay. A plausible interpretation of the model is that the firm and CEO renegotiate the labor contract at the beginning of each year. The contract sets expected pay in the coming year to the level that induces the CEO to remain at the firm and work throughout year t. Clearly, salary and bonus should be included in w_{ijt}. 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 we include stock and option grants in the year they were awarded or the year when they vested? Rather than taking a stand, I use both methods. Next I provide details and additional motivation for both measures.

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 [7]. One justification for including stock and options in the year granted rather than year vested is that, although a CEO loses unvested shares if he leaves the firm, there is anecdotal evidence that the CEO's new firm will pay a hiring bonus that compensates the CEO for lost, unvested shares and options[8]. In other words, if the new firm will make the CEO whole, then from the point of view of the CEO's wealth, grants effectively vest immediately.

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

[7] Salary is Execucomp "salary," bonus is "bonus," value of options granted is "opt_awards_blk_value," value of restricted stock granted is "rstkgrnt," other incentive compensation is long term incentive plan payouts ("ltip") before 2006 and non-equity incentive plan compensation ("noneq_incent") in 2006 and after. Other annual compensation is "othann"+"othcomp" before 2006 and "defer_rpt_as_comp_tot"+"othcomp" in 2006 and after.

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

were granted. The Appendix contains details on constructing this measure, which to my knowledge is new to the literature. This proxy better measures annual flows to the CEO if he/she truly loses unvested shares upon leaving the firm. This measure is also smoother over time. To see why smoothness is desirable, consider a firm that, every fourth year, offers the CEO a large option grant that vests 25% each year. This second proxy ($w^{(vest)}$) will be fairly smooth over time, whereas the first proxy ($w^{(grant)}$) will register a large spike in pay every 4th year. $w^{(grant)}$ is therefore somewhat undesirable, because the model would interpret this spike in pay as a spike in the CEO's perceived ability every fourth year, whereas the spike is really just an artifact of the firm's timing convention for option grants.

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. Industry-adjusted stock returns equal the firm return minus the return on the corresponding equal-weighted Fama-French 49 industry portfolio. 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 the sample) then I use the actual value, which amounts to assuming everyone knew from the beginning when the CEO would leave office. If T_j is not known (i.e. CEO's last year is not in sample), then I forecast it using the CEO's age and tenure from his last observation in the database; details are in the Appendix.

Estimation uses data on the change in expected pay, scaled by lagged market value. Following equation (29), I estimate expected pay $\widehat{E}_t[w_{ijt}]$ by subtracting the unexpected portion ($\widehat{b}_{ijt}r_{it}$) from realized pay (w_{ijt}). As I explain above, b_{ijt} is identified in a regression of realized pay on contemporaneous returns and information at the beginning of the year t. I parameterize b_{ijt} as

$$b_{ijt} = a_0 + a_1 \log(M_{it}),$$

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

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

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

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

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

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

I winsorize $\Delta\widehat{E}_t\left[w_{ijt}\right]/M_{it-1}$ at the 1st and 99th percentiles, and I subtract off the yearly median, 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 cleaned database contains 20,499 firm/year observations and 4,513 CEOs. There are fewer observations of the change in expected pay, since these measures cannot be computed in CEOs' first year in office. Mean realized pay is around \$6 million using the $w^{(grant)}$ 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 $w^{(grant)}$, the standard deviation of changes in expected pay is around \$7 million or, scaled by lagged market cap, 0.5%. Using $w^{(vest)}$, the standard deviation of changes in expected pay is lower (around \$4 million or 0.4% as a percent of lagged market cap). $RETVAR_{it}$, the annualized variance of weekly returns within a firm/year, has median 0.12, which corresponds to annualized return volatility of $\sqrt{0.12} = 35\%$. This volatility is below the full sample standard deviation of annual excess returns (47%), possibly because of differences in average returns across stocks. 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. There is considerable variation in firm size in the sample.

C. GMM Estimator

I estimate the four model parameters in $\Theta = \left[\begin{array}{cccc} \sigma_\varepsilon^2 & \sigma_0^2 & \theta^{up} & \theta^{down} \end{array}\right]$ using the general method of moments (GMM). The GMM estimator $\widehat{\Theta}$ is

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

I estimate the four parameters in using 12 moments. M is a 12×1 vector of empirical moments, and $m(\Theta)$ is the corresponding vector of moments predicted from the model using parameter values Θ. I set $\mathbf{W}$ equal to the efficient weighting matrix, which is the inverse of the estimated covariance of moments M.

The first 10 moments in M and m are means, and the last two are regression slopes. The first 10 moments are the average variance of excess returns for CEOs in their 1st, 2nd, ..., 9th, and 10+ year in office. The 11th [12th] moment is the slope from an OLS regression of $\Delta\widehat{E}_t\left[w_{ijt}\right]/M_{it-1}$ (the scaled change in expected CEO pay) on r_{it-1} (lagged excess return), in the subsample in which perceived CEO ability increased [decreased]. One way to create these subsamples is to condition on the sign of the estimated change in expected pay, $\Delta\widehat{E}_t\left[w_{ijt}\right]$. A problem with this approach is that $\Delta\widehat{E}_t\left[w_{ijt}\right]$ is measured with error. Another approach is to condition on whether the realized excess return r_{it-1} is above or below its median. A value of r_{it-1} above (below) the median indicates that beliefs increased (decreased) during period $t-1$. I choose this latter approach since realized returns are measured with less error. To compute the predicted regression slopes, I first use each empirical observation $\{r_{it-1}, \tau_{jt}, T_j\}$ to compute the predicted value of $\Delta E_t\left[w_{ijt}\right]/M_{it-1}$ using equation (24); I then regress the model's predicted values of $\Delta E_t\left[w_{ijt}\right]/M_{it-1}$ on the empirical values of r_{it-1}, conditioning on whether the empirical r_{it-1} exceeds the median.

The estimation procedure accommodates certain types of measurement error in CEO pay. If I introduce an error term δ_{ijt} into equation (24), then we have

$$\frac{\Delta E_t\left[w_{ijt}\right]}{M_{it-1}} = r_{it-1}\gamma\left(\cdot\right) + g\left(\cdot\right) + \delta_{ijt}. \tag{30}$$

As long as measurement error δ_{ijt} is uncorrelated with r_{it-1}, then none of the moments used in GMM estimation will depend on δ_{ijt}, so the parameter estimates will not change.

V. Estimation results

Before presenting parameter estimates, I describe how the model fits the data. Figure 2 plots the variance of excess stock returns against CEO tenure. Variance in years 1-10 are the first 10 moments used in the GMM 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 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 very 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 rises to 39% for CEOs in years 10+. The model generates a drop in return variance due to learning, as expected. Return variance drops rapidly and then levels off in the data, whereas the model generates a gradual drop in return volatility. The predicted drop is gradual, because agents are learning very slowly, which in turn occurs because the agents' signal is estimated to be extremely noisy (estimates below). The model predicts a smaller drop in volatility than we see in the data. The reason is that the GMM procedure chooses to match age categories 2 through 10+ fairly well, at the expense of fitting age category 1 very poorly. If I estimate the model with only 4 moments (the 2 regression slopes, and return variance in age categories 1 and 10+ only), the model can exactly fit these 4 moments. However, the model fits return variance in age categories 2 through 9 quite poorly. Altering the model so that it better fits the drop in return volatility after successions is a priority for future work.

INSERT FIGURE 2 HERE

In the data we see a gradual increase in return volatility before the previous CEO leaves office, while the model predicts a drop in volatility. The model predicts a drop in volatility because successions are exogenous (when a CEO is hired, his succession date is known with certainty), uncertainty about ability drops with tenure, news about the incumbent's ability impacts fewer future years as we approach succession, and there is no uncertainty about the replacement CEO's prior mean ability. 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_{jt} drops below an exogenous threshold, $\underline{\mu}$.[9] Taylor (2009) 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 (2009). I simulate this altered model and plot average return volatility versus tenure in Figure 3. The model now produces a rise and fall in volatility around successions. Predicted volatility rises before successions, because some successions are due to firings, and firings must be preceded by abnormally low earnings in order to pull the posterior mean below the firing threshold. As a result, the altered model produces negative average abnormal stock returns leading up to successions (not shown), whereas the main model produces no average abnormal returns. These abnormal returns contribute to higher return volatility before successions. This alternate model still fails in two ways: the predicted run-up in volatility is smaller than the one observed in the data, and volatility peaks in year 0 (the previous CEO's last year) in the model but in year 1 in the data.

INSERT FIGURE 3 HERE

Although this model with endogenous firings fits the data better, I focus on estimation results from the simpler model, for three reasons. First, while pre-succession volatility looks quite different in the altered model, post-succession volatility does not (compare Figures 2

[9] I assume endogenous firings do not change the expression for stock returns, in order to keep this exercise simple. This is a poor assumption, since pre-firing stock prices may reflect a positive probability of a firing. Correcting for this requires solving a dynamic program, as in Taylor (2009). I have chosen to keep this exercise simple since it is not a main result of the paper.

and 3). Since the GMM 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, I do not have a closed-form solution for the altered model, and hence GMM estimation is not feasible[10].

Next I examine how well the model fits the relation between expected pay and lagged excess returns. I present results using $w^{(grant)}$; results are similar using the other pay measure. Figure 4 plots the empirical scaled change in expected CEO pay versus the lagged excess return. Each point is a firm/year observation. The solid vertical line splits the sample by the median lagged return. According to the model, beliefs about CEO ability decreased in the left subsample and increased in the right subsample. The figure shows the best fit line in each subsample. The slopes of these two lines are the last two moments used in GMM estimation. The slope is significantly positive in both subsamples, indicating that expected pay rises more (or falls less) when the signal of CEO ability is higher. The slope is significantly higher in the subsample with rising beliefs (0.0028 versus 0.0012), suggesting expected pay is more sensitive to good news than bad news.

INSERT FIGURE 4 HERE

Figure 5 shows the same plot, but with model predictions instead of real data. The model matches the empirical slopes exactly. In the model as in the data, expected pay is less sensitive to lagged returns in the subsample with decreasing rather than increasing beliefs. In the model we see a much tighter relation between expected pay and lagged return than we see in the real data. Even the predicted relation does not have an R^2 of 1, because the predicted sensitivity depends on current and final tenure, which vary across observations. In the model we never see observations with both a positive change expected pay and a lagged return below its median (and vice versa), but there are many such observations in the real data. These empirical observations may be due to measurement error in expected pay, which the estimation procedure can accommodate (equation (30)).

INSERT FIGURE 5 HERE

Table 2 summarizes the 12 moments used in GMM estimation. p-values test whether the model-implied moment matches the empirical moment. Using both pay measures, equality is rejected at the 5% level for 4 of 12 moments, and at the 1% level for 3 of 12 moments.

[10] Taylor (2009) solves this problem by using the simulated method of moments (SMM) estimator instead of GMM. That approach is also possible here, but it adds complexity to the estimation procedure and only buys us the new estimated parameter, $\underline{\mu}$.

The table also provides a χ^2 statistic that jointly tests whether the model matches all 12 empirical moments. The p-value rejects equality at the 1% confidence level. 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 had used only 4 moments to estimate the 4 parameters, model fit would have appeared perfect.

INSERT TABLE 2 HERE

Table 3 presents the GMM parameter estimates using both measures of CEO pay. The estimated standard deviation of profitability, σ_ε, is roughly 35% using both pay measures. The model needs this high value to match the high level of stock return volatility. This high volatility implies that agents learn extremely slowly about CEO ability. To see this, recall from Bayes' Rule that change in the posterior mean ability after the CEO's first year in office equals

$$\Delta m_{ijt} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2} \widetilde{Y}_{it}. \tag{31}$$

Plugging in parameter estimates for σ_0 and σ_ε, we have $\Delta m_{ijt} = 0.016\ \widetilde{Y}_{it}$. In other words, the sensitivity of beliefs to the signal is just 0.016, so beliefs change very slowly over time.

INSERT TABLE 3 HERE

The estimated standard deviation of prior beliefs about CEO ability (σ_0) is 4.5% using $w^{(grant)}$ and 4.7% using $w^{(vest)}$. Parameter σ_0 is measured in units of percent return on assets (ROA) per year. Using this estimate, the difference in *true, unobservable* ability between a CEO at the 5th and 95th percentiles is $2 \times 1.65 \times \sigma_0 = 15\%$ of ROA per year, which is extremely large. The difference in *perceived* ability (i.e. the posterior mean) is much smaller, because agents are learning very slowly. For instance, after 5 and 10 years of learning have occurred, the standard deviation of perceived ability across CEOs is just 1.2% and 1.7%, respectively[11]. Even these amounts are huge. The average sample firm has \$11 billion in assets, so a 1.7% difference in ability across CEOs amounts to a difference of \$11 billion $\times$ 1.7% = \$185 million in annual profits.

Why does the model need such high uncertainty to fit the data? The drop in stock return volatility is quite large. To fit this large drop, the model needs either a very high value of σ_0 or a low value of θ, as discussed in Section 4.1. As we will see, it turns out the

[11] According to Bayes' rule, the variance in the posterior mean after τ periods of learning have occurred equals

$$var(m_{jt}|\tau) = \sigma_0^2 \left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2/\tau} \right).$$

Using $\tau = 5$ and $\tau = 10$ and the parameter values in Table 3, we obtain $stdev(m_{jt}|\tau) = 1.2\%$ and 1.7%.

model needs both.

For $w^{(grant)}$, the estimates of θ^{up} and θ^{down} are 0.198 and 0.084, respectively. These estimates imply that CEOs capture 19.8% of the surplus from an improvement in their perceived contribution to firm profits, and they bear 8.4% of the negative surplus resulting from a decrease in their perceived contribution. In other words, the level of CEO pay moves 0.198 (0.084) for 1 with the CEO's perceived contribution to firm profits when good (bad) news arrives. Both values of θ are significantly positive and are much closer to 0 (CEO receives none of his surplus) than 1 (CEO receives entire surplus). Point estimates indicate expected pay is more sensitive to good news than bad news, i.e. $\widehat{\theta}^{up} > \widehat{\theta}^{down}$, but the difference between these estimates is not statistically significant (difference has t-statistic 1.5). Results are qualitatively similar using $w^{(vest)}$, the other pay measure: the estimates of θ^{up} and $\theta^{(down)}$ are 0.332 and 0.202, respectively, and are not statistically different from each other.

The model needs these low values of θ^{up} and θ^{down} to fit the data, because the dollar changes in expected CEO pay are small compared to the size of their firms and the CEO's estimated importance. Imagine a CEO's perceived ability changes by 0.5 percentage points, which is not unreasonable given the learning results above. If the CEO works at a firm with \$11 billion in assets (the sample average), then the CEO's expected contribution to profits has increased by \$11 billion $\times$ 0.5% = \$55 million. If the CEO captured his entire surplus, then his expected pay would rise by \$55 million. We do not observe changes in expected pay this large in the data; the sample standard deviation equals \$7.4 million and \$4.0 million using the two pay measures. To fit the small observed changes in expected CEO pay, the model infers that the CEO captures a small fraction of the \$55 million surplus.

The results imply that CEOs have some power over their own compensation, but not much. The estimates of θ^{up} and θ^{down} reject the extreme view that CEOs have all the bargaining power, and they also reject the extreme view that CEOs have no bargaining power. If CEOs have total bargaining power, as they do in the model of Harris and Hölmstrom (1982), then we expect them to capture the entire surplus from any good news about their ability (i.e. $\theta^{up} = 1$), and we expect them to avoid any cuts in their expected pay when their perceived ability drops (i.e. $\theta^{(down)} = 1$). Using $w^{(grant)}$, the estimates above reject the hypothesis that $\theta^{up} = 1$ with a t-statistic of 11.5, and reject the hypothesis that $\theta^{(down)} = 0$ with a t-statistic of 2.7. CEOs with considerable bargaining power should gain more from good news than they lose from bad news (i.e. $\theta^{up} > \theta^{down}$), but the data cannot reject the hypothesis that $\theta^{up} = \theta^{down}$ ($t = 1.5$). If CEOs have no bargaining power, then CEOs' expected pay should never rise ($\theta^{up} = 0$), and CEOs should bear the entire negative surplus from bad news ($\theta^{down} = 1$). Using $w^{(grant)}$, the estimates above reject the hypothesis that $\theta^{up} = 0$ with $t = 2.8$, and reject the hypothesis that $\theta^{down} = 1$ with $t = 29.1$. Results are qualitatively similar using the other pay measure, $w^{(vest)}$.

The estimated model's predictions are quite different from the predictions of counterfactual benchmark models with total CEO power ($\theta^{up} = 1$, $\theta^{down} = 0$) or no CEO power ($\theta^{up} = 0$, $\theta^{down} = 1$). Table 4, Panel B compares the models in terms of their predicted sensitivity of scaled changes in expected CEO pay to lagged excess returns. The benchmark model with total CEO power predicts a sensitivity of 0.0147 for good news and 0 for bad news. In contrast, the empirical and predicted model's sensitivities are 0.0028 for good news and 0.0012 for bad news. These sensitivities are also quite different from those in the benchmark model with no CEO power (0 for good news and 0.0149 for bad news). These comparisons reinforce that neither extreme benchmark (total CEO power or no CEO power) accurately describes the data.

VI. Implications for Shareholder Value

The parameter estimates have implications for the question, does CEO ability matter for shareholders? If CEOs capture less of their surplus, then shareholders benefit more from good news about CEO ability and, conversely, suffer more from bad news about CEO ability. The estimated model helps to quantify these costs and benefits. I perform two exercises. The first examines how stock prices respond to news about CEO ability, and how this sensitivity depends on the CEO's share of the surplus. The second examines predictions and data on abnormal stock returns around unanticipated CEO deaths.

As explained in Section 3, the model predicts that good news about CEO ability coincides with higher stock returns, and the relation between the two is stronger when CEOs capture less of their surplus. I use equation (21) to compute the predicted slope of excess returns on changes in posterior mean CEO ability, assuming the CEO has been in office 5 years and is expected to leave after 10 years. All parameters start at their estimated values using $w^{(grant)}$. Predicted slopes are in Panel C of Table 4. When I increase the CEO's surplus θ^{up} from 0 to 0.198 (the estimated value) to 1, the predicted slopes decrease from 73.5 to 72.7 to 69.6. A slope of 69.6 means that a 0.1 percentage point increase in perceived CEO ability[12] coincides with a 6.96% higher excess return. As expected, the sensitivities decrease in the

[12] A 0.1 percentage point increase in perceived ability is not atypical for CEOs who have been in office 5 years. Recall from above that after 5 and 10 years of learning have occurred, the standard deviation of perceived ability across CEOs is just 1.2% and 1.7%, respectively.

CEO's surplus, but the change in sensitivities is fairly small. According to this exercise, CEO ability matters only slightly more to shareholders when CEOs capture less of their surplus.

Unanticipated CEO deaths provide a natural experiment with implications for the effect of CEO ability on shareholders. The more CEO ability matters for shareholders, the more we expect stock prices to change around unanticipated CEOs' deaths. If we assume[13] $\theta^{up} = \theta^{down} = \theta$, then the model makes several predictions about stock returns around unanticipated CEO deaths. Surprisingly, these predictions depend strongly on the CEO surplus parameter θ.

Prediction 4: Under the assumptions of the main model, the stock return in response to an unanticipated death of CEO j at the beginning of period t equals

$$R_{i,t}^{event} = \frac{B_{it}}{M_{it}} \left(m_{i0} - m_{ijt-1} \right) (1 - \theta) \beta \left(\frac{1 - \beta^{T_j - \tau + 1}}{1 - \beta} \right). \tag{32}$$

Proof in 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 is higher around unanticipated CEO deaths.

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

Here is the intuition for these predictions. Averaging across CEOs, the average incumbent's perceived ability m_{ijt-1} equals the prior mean m_{i0}, so the average event return is zero (H1). In other words, the average CEO is neither better nor worse than his/her replacement, so on average a CEO death is neither good nor bad news for shareholders. If the CEO captures his entire surplus ($\theta = 1$), then the deceased CEO's expected pay exactly offset his expected contribution to profits. In this special case, shareholders are indifferent between the current CEO and his replacement, so we should see no change in stock price when the CEO unexpectedly dies (H2). If $\theta < 1$, meaning shareholders receive some positive surplus,

[13] This assumption simplifies the predictions considerably. As I explain in the previous section, parameter estimates cannot reject the hypothesis that $\theta^{up} = \theta^{down}$.

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 Johnson, Magee, Nagarajan, and Newman (1985). They identify 53 unanticipated deaths of senior executives at publicly traded U.S. firms from 1971–1982. While their sample includes some non-CEO executives, the majority of observations (36 out of 53) are CEOs. Their sample only includes deaths not attributed to prolonged illness, complications from surgery, or unknown cause. In their final sample, 26 out of 53 deaths are from heart attacks, 12 are due to accidents or suicides, and so on.

Consistent with H1, Johnson et al. find that the average return around the death is indistinguishable from zero. H2 is easy to reject simply due to noise in the data, and indeed they find that not all event returns equal zero.

Johnson et al. find that stock return volatility almost doubles in the days around the death announcement, which is consistent with H3 and $\theta < 1$. In addition to testing this directional prediction, we can compare empirical and predicted magnitudes. The Appendix contains a formula for the predicted increase in return variance around unanticipated CEO deaths. Panel D of Table 4 compares the empirical and predicted increase in percent return variance. I derive the empirical variance from Table 4 in Johnson et al. (1985). To obtain predicted values, I set the estimated value of θ to the average of $\widehat{\theta}^{up}$ and $\widehat{\theta}^{down}$. The empirical increase is 1.3 percent squared, whereas the estimated model predicts an increase of 22.3. The counterfactual model in which the CEO captures none of the entire surplus misses even more, predicting an increase in return variance of 30.2. The counterfactual model in which the CEO captures the entire surplus predicts no increase in return variance, which comes closest to matching the data.

To test H4 we need proxies for the perceived ability of the deceased and replacement executives (m_{ijt-1} and m_{i0}, respectively). Assumption 6 in the model tells us that the level of pay is a good proxy for an executive's perceived ability. Identifying the executive's replacement is a challenge. Johnson et al. (1985) argue that the highest paid executive (excluding the deceased) is a likely replacement for the deceased executive. Following this argument, a proxy for $m_{ijt-1} - m_{0i}$ is the incumbent's direct compensation in the year preceding his/her death, divided by the direct compensation of most highly paid executive (excluding the deceased). In cross-sectional regressions, Johnson et al. show that the death announcement return has a negative slope on this compensation ratio, even after including several control variables. In other words, stock prices drop when highly paid CEOs dies, and vice versa. This negative correlation is consistent with H4 and $\theta < 1$. In sum, the results from Johnson et al. are directionally consistent the model and conclusion that $\theta < 1$, meaning CEOs do not capture their entire 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_i is unobservable and fluctuates over time, and agents learn about its value at the same time they learn about CEO ability. I call uncertainty about a_i "firm uncertainty." Profitability still follows

$$Y_{it} = a_{it} + \eta_{ij} + v_t \left(\frac{M_{it}}{B_{it}} \right) + \varepsilon_{it} \tag{33}$$

but firm quality evolves over time according to

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

Variables η_{ij}, a_{it}, ε_{it}, and u_{it} are all unobsevable. All other parameters are known. Agents learn about η_{ij} (CEO ability) and a_{it} (firm quality) from realized profitability according to Bayes Rule or, equivalently, the Kalman filter. For instance, high realized profitability will increase the posterior mean of both CEO ability and firm quality. I assume shocks are distributed as

$$\begin{pmatrix} \varepsilon_{it} \\ u_{it} \end{pmatrix} \sim N \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}. \tag{35}$$

When a new CEO takes office at the beginning of period t, prior beliefs about a_{it-1} and η_{ij} are reset to

$$\begin{pmatrix} a_{it-1} \\ \eta_{ij} \end{pmatrix} \sim N \left(\begin{pmatrix} \widehat{a}_{t-1|t-1} \\ m_{i0} \end{pmatrix}, \begin{pmatrix} \Sigma_{a_{t-1}|t-1} & 0 \\ 0 & \sigma_0^2 \end{pmatrix} \right), \tag{36}$$

where $\widehat{a}_{t-1|t-1}$ is the posterior from the previous period, and $\Sigma_{a_t|s}$ indicates the variance of beliefs about a_{it} at the end of period s. All other model assumptions are the same as before, although I assume $\theta^{up} = \theta^{down} = \theta$, for simplicity. The main model is the special case of this model with $\sigma_u = 0$ and $\rho = 0$ (so $a_{it} = \overline{a}_i$), which implies no uncertainty about firm quality ($\Sigma_{a_{t-1}|t-1} = 0$).

I solve this extension numerically. Details are in the Appendix. Figure 6 compares predictions from models with and without learning about firm quality. The solid line shows the base case with no uncertainty or learning about firm quality; this case assumes σ_u is (almost) zero and other parameters equal their estimated values from Table 3. The solid line keeps those other parameters the same but raises the value of σ_u to 0.06, so that agents are learning about firm quality. Starting in the top left of Figure 6, raising σ_u raises the amount of firm uncertainty, as expected. Firm uncertainty exhibits almost no variation with CEO tenure. The reason is that uncertainty about a_{it} eventually converges to a steady state level. The simulated firm is typically at this steady state before the new CEO takes office, and bringing in a new CEO with uncertain ability has little effect on firm uncertainty. On the top right, uncertainty about CEO ability drops slower when there is more firm uncertainty. When agents observe high realized profitability, they do not know whether it is due to good luck (the ε shock), high CEO ability, or high firm quality. This last, new source of uncertainty makes profitability a less precise signal about CEO ability, so agents learn slower about CEO ability. In the bottom left, we see that increasing firm uncertainty shifts return volatility upwards. This is because we are adding more uncertainty to the model. Given the result in the upper right, I expected return volatility to drop more with tenure when there is less firm uncertainty. However, in the lower left we see almost no effect on the decline in return volatility. This difference may simply be too small to detect. Finally, in the bottom right we see that 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 raise beliefs about firm quality, which in turn raises expected future profits.

INSERT FIGURE 6 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. For the sake of this exercise, I interpret the solid lines in Figure 6 as both the empirical moments and the predicted moments from the baseline model with no firm uncertainty; in other words, I assume the simple model with no firm uncertainty was fitting the data well. 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. The model no longer fits the empirical moments— the dashed line is far from the solid line. How can we make the model fit the empirical moments again? Keeping σ_u at its high level of 0.06, I choose other parameter values by hand to make the model fit the "data," i.e, the solid line. I plot the new predicted moments as dots and label them "hand fit." The dots are close to the solid line, although the fit is not perfect. I changed two parameters to re-fit the model. First, I reduced profit volatility (σ_ε) from 0.349 to 0.29 in order to bring the level of return volatility back down (bottom left). Less volatile profits makes agents learn faster. To offset this effect, I reduced prior uncertainty about CEO ability (σ_0) from 0.045 to 0.04. Surprisingly, these two changes bring the sensitivity of expected pay to lagged returns back to its empirical level, without requiring any change in θ (bottom right). The main reason is that reducing σ_ϵ increases the predicted sensitivity, 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 necessary change the estimated CEO surplus parameter θ.

VIII. Conclusion

I solve and estimate a model in which CEO ability affects profits, agents learn gradually about CEO ability, and the CEO and shareholders split the surplus resulting from a change in the CEO's perceived ability. Parameter estimates tell a mixed story about CEO pay and CEO power. The level of CEO pay rises following good news about CEO ability, but CEOs capture only an estimated 20–33% of the positive surplus. The level of pay falls following bad news about CEO ability, but CEOs only bear an estimated 8–20% of the negative surplus. These estimates contrast the extreme view that CEOs have enough power to capture all the benefits from good news while avoiding any costs from bad news. The results also contrast the opposite extreme view, which is that CEOs have so little power that they receive no benefits from good news yet bear all the costs of bad news.

This paper measures the relative bargaining power of CEOs and shareholders, but it does not explain which economic factors determine CEOs' bargaining power. Possible factors relate to firm-specific human capital, governance, long-term contracts with renegotiation costs, heterogeneity in firm size, and other frictions in the CEO labor market. I find that these factors, taken together, are of first-order importance for understanding the CEO labor market. Understanding which factors matter most is an important area for future work.

REFERENCES

Alder, Simeon, 2009, In the wrong hands: Complementarities, resource allocation, and aggregate TFP, Working paper, UCLA.

Bebchuk, Lucian A., Jesse M. Fried, and David I. Walker, 2002, Managerial power and rent extraction in the design of executive compensation, *The University of Chicago Law Review* 69(3), 751–847.

Bennedsen, Morton, Francisco Perez-Gonzalez, and Daniel Wolfenzon, 2008, Do CEOs matter?, Working paper, Stanford University.

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.

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.

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

Edmans, Alex, Xavier Gabaix, and Augustin Landier, A multiplicative model of optimal CEO incentives in market equilibrium, *Review of Financial Studies,* forthcoming.

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.

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

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

Jensen, Michael C. and Kevin J. Murphy, 1990, Performance pay and top-management incentives, *Journal of Political Economy* 98(2): 225–264.

Johnson, W. Bruce, Robert P. Magee, Nandu J. Nagarajan, and Harry A. Newman, 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*, Forthcoming.

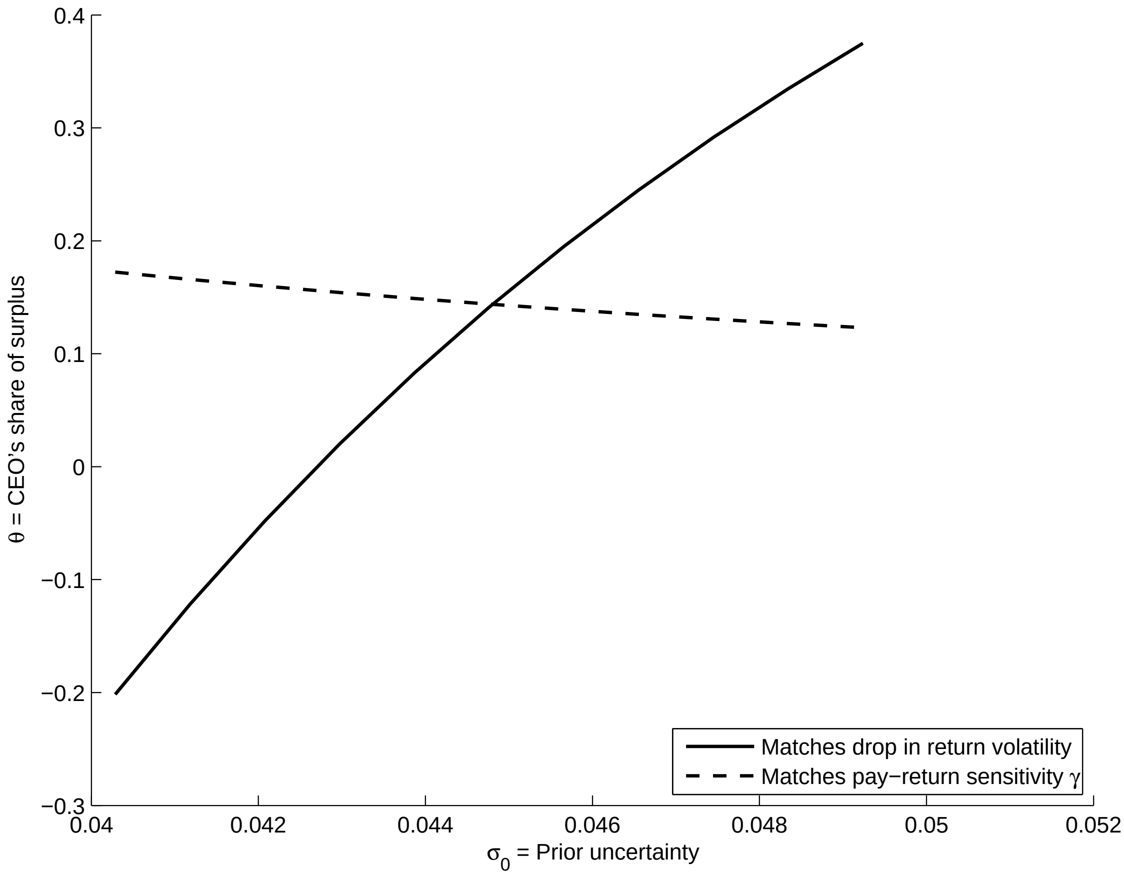
Milbourn, Todd T., 2003, CEO reputation and stock-based compensation, *Journal of Financial Economics* 68: 233–262.

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

Murphy, Kevin J., 1999, Executive compensation, in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Vol. 3b Elsevier Science North Holland, Chapter 38: 2485–2563.

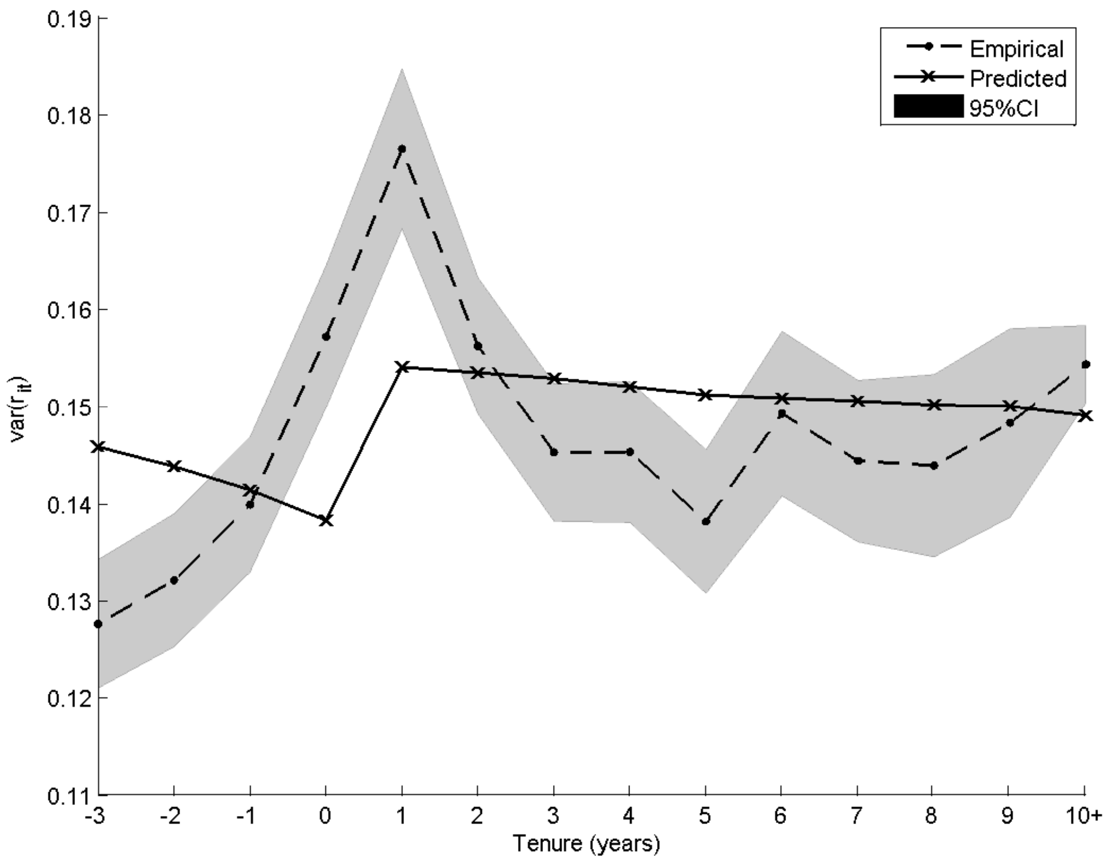
Taylor, Lucian A., 2009, Why are CEOs rarely fired? Evidence from structural estimation, Working paper, University of Pennsylvania.

Figure 1: **Disentangling Uncertainty and the CEO's Surplus**



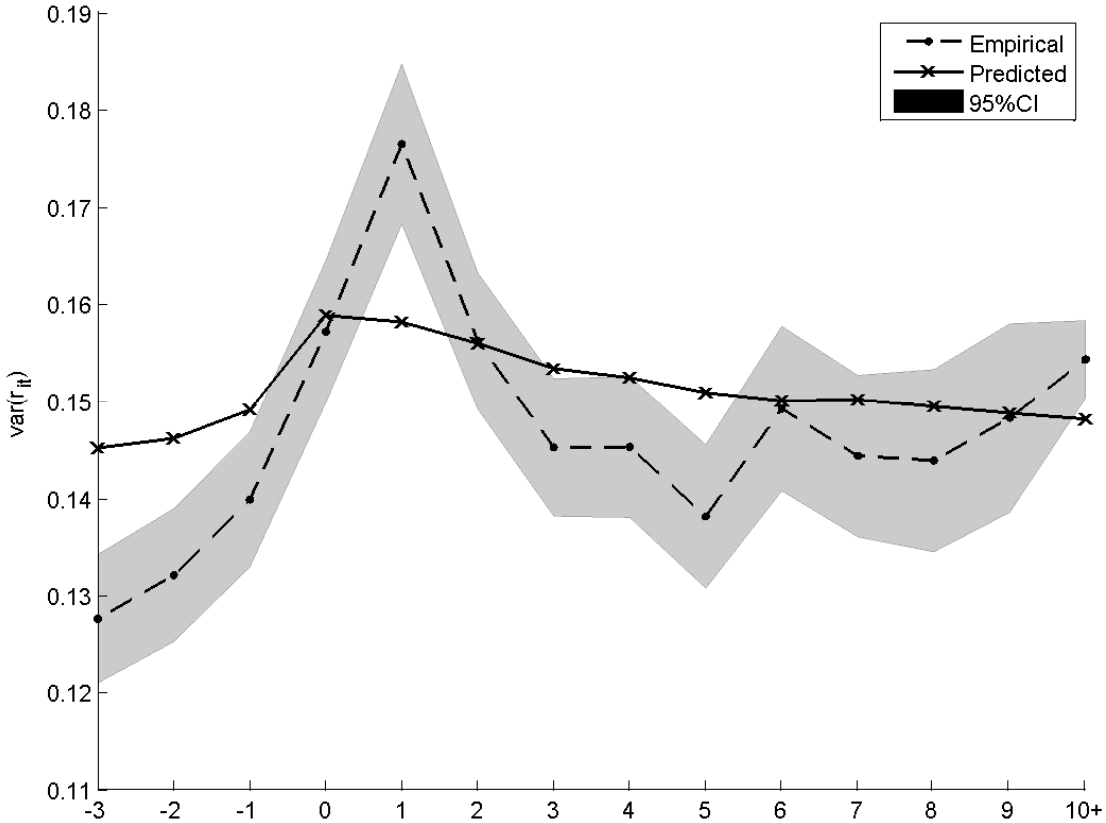
The solid line plots the combinations of parameters $\theta = \theta^{up} = \theta^{down}$ (the CEO's fraction of the surplus) and σ_0 (prior uncertainty about CEO ability) that allow the model to match the empirical drop in return volatility. The drop in return volatility is computed as the variance of weekly returns (annualized) for CEOs in their first year in office, minus the corresponding value for CEOs in years 10+ in office. The empirical return variances are in Table 2. The dashed line plots the combinations of parameters θ and σ_0 that allow the model to match the slope from a regression of scaled changes in the expected CEO pay on lagged excess stock returns. These empirical slopes are from Table 2.

Figure 2: **Stock Return Volatility and CEO Tenure**



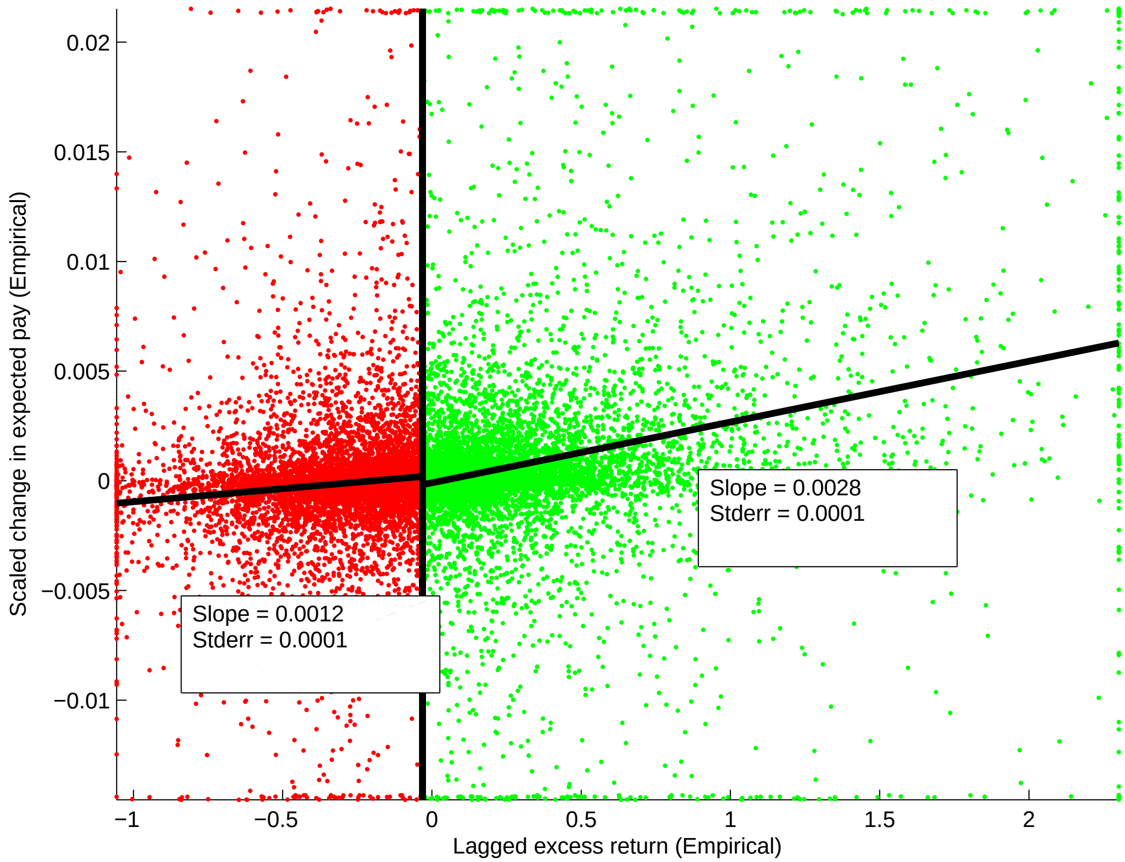
This figure plots the variance of annual excess stock returns ($var_t(r_{it})$) at various CEO tenure levels. Year 1 is the CEO's first year in office; year zero is the previous CEO's last year in office. The dashed line with its corresponding 95% confidence interval is computed from the empirical sample.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

Figure 3: **Stock Return Volatility and CEO Tenure, Allowing CEO Firings**



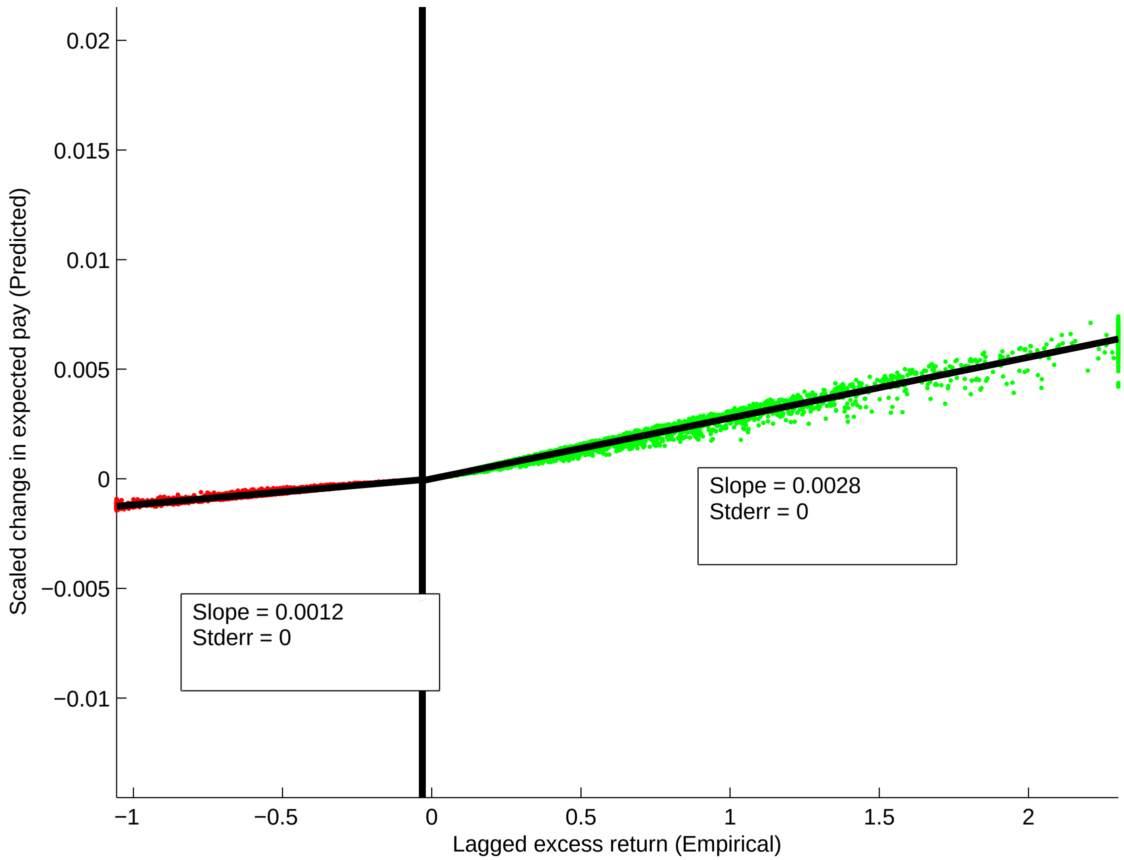
This figure plots the variance of annual excess stock returns ($var_t(r_{it})$) at various CEO tenure levels. Year 1 is the CEO's first year in office; year zero is the previous CEO's last year in office. The dashed line with its corresponding 95% confidence interval is computed from the empirical sample. 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 parameter values in Table 3, with w^{grant}. This version of the model assumes the CEO is fired if his posterior mean ability m_t drops below $\underline{mu} = -0.03$. All other features of the model are the same as before. This model predicts 2.8% of CEOs are fired per year, 25% of CEOs are eventually fired, and median CEO tenure is 7 years.

Figure 4: **Empirical Sensitivity of Pay to Lagged Returns**



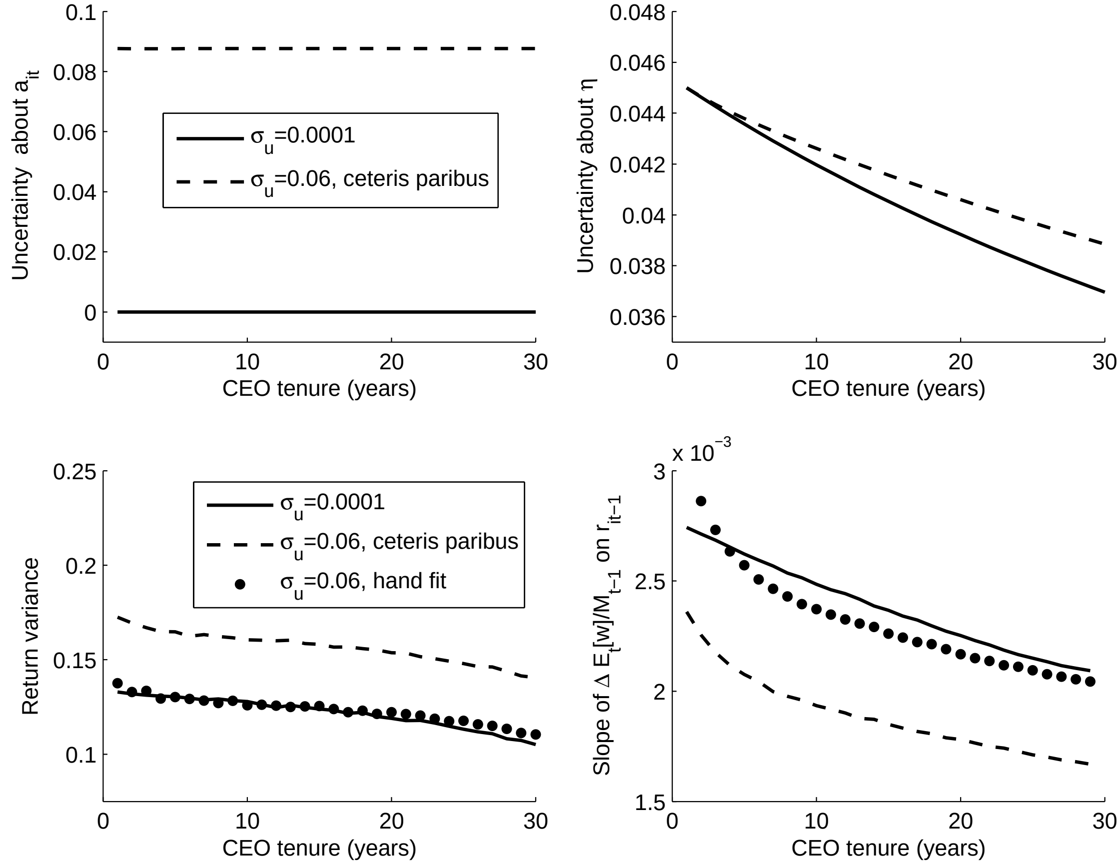
This figure plots the empirical scaled change in expected CEO pay $= \Delta E_t[w_{ijt}]/M_{it-1}$ versus the empirical lagged excess returns $= r_{it-1}$. Each point is a single observation. The solid vertical line denotes the median lagged excess return. The other solid lines are the best-fit lines from two OLS regressions, one from each subsamples formed by the median excess return. Textboxes show the estimated slope and its standard error from these two OLS regressions.

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



This figure plots the predicted scaled change in expected CEO pay $= \Delta E_t[w_{ijt}]/M_{it-1}$ versus the empirical lagged excess returns $= r_{it-1}$. Each point is a single observation. The predicted scaled change in expected pay is computed using equation () from the model, which requires the lagged excess return, current tenure, and final tenure as inputs. The solid vertical line denotes the median lagged excess return. The other solid lines are the best-fit lines from two OLS regressions, one from each subsamples formed by the median excess return. Textboxes show the estimated slope and its standard error from these two OLS regressions.

Figure 6: **Model Predictions with Learning about Firm Quality**



This figure shows predicted moments from the model in Section 7.1, 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_{ij}" is the standard deviation of beliefs about η_{ij}, CEO ability. $cov(\Delta w_{ijt}, r_{it-1})$ is the covariance of dollar changes in CEO pay with stock returns from the previous year. All results are from model simulations described in the Appendix. The solid line uses $\sigma_u = 0.001$, $\rho = 0.75$, and other parameter values taken from Table 3, using w^{grant}. The dashed line uses $\sigma_u = 0.06$, $\rho = 0.75$, and other parameter values taken from Table 3. The dots use $\sigma_u = 0.06$, $\rho = 0.75$, $\theta = 0.169$, $\sigma_\epsilon = 0.29$, and $\sigma_0 = 0.04$.

Table 1: **Summary Statistics**

This table shows summary statistics for the sample of 4513 CEOs. Data on CEO pay come from Execucomp, 1992-2007. w_t^{vest} is CEO pay, including stock and option grants in the year when they vest. w_t^{grant} is CEO pay, including stock and option grants in the year when they were 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 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.

Variable	Notation	N	Mean	Std	25th pctl	Median	75th pctl
Realized pay ($M)	w_t^{vest}	16235	6.01	10.11	1.16	2.51	6.05
Expected pay ($M)	$E_t[w_t^{vest})]$	16235	6.06	10.03	1.26	2.73	6.27
Change in E[pay] ($M)	$\Delta E_t[w_t^{vest}]$	12310	0.56	7.38	-1.28	0.00	1.70
Scaled Δ in E[pay] (%)	$\Delta E_t[w_t^{vest)}/M_{it-1}$	12310	0.078	0.547	-0.086	0.000	0.137
Realized pay ($M)	w_t^{grant}	20294	4.13	5.60	1.02	2.10	4.68
Expected pay ($M)	$E_t[w_t^{grant}]$	20294	4.13	5.60	1.06	2.12	4.68
Change in E[pay] ($M)	$\Delta E_t[w_t^{vest}]$	17276	0.175	4.030	-0.673	0.000	0.906
Scaled Δ in E[pay] (%)	$\Delta E_t[w_t^{vest}]/M_{it-1}$	17276	0.037	0.402	-0.047	0.000	0.076
Excess annual return	r_{it}	20499	0.014	0.472	-0.245	-0.034	0.192
Annual return	R_{it}	20499	0.185	0.517	-0.113	0.116	0.366
Variance of returns	$RETVAR_{it}$	20499	0.152	0.164	0.050	0.124	0.207
Current tenure (years)	τ	20499	7.9	7.4	3.0	6.0	10.0
Final tenure (years)	T	20499	13.5	8.3	8.0	12.0	17.0
Market cap ($B)	M	20499	6.08	20.94	0.43	1.17	3.78
Assets ($B)	B	20499	10.96	53.85	0.39	1.27	4.86

Table 2: **Moments from GMM Estimation**

Panel A shows the 12 moments used in the GMM estimation. The data moments are computed from the empirical sample of 4376 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 annualized variance of weekly excess stock returns during the CEO's t-th year in office. The empirical measure is the annualized variance of weekly excess returns. "Slope, beliefs up (down)" is the slope from an OLS regression of scaled changes in expected pay $= \Delta E_t[w_{ijt}]/M_{it-1}$ on lagged excess returns $= r_{it-1}$, in the subsample where lagged excess returns are above (below) the median. To compute the predicted slope, I compute a predicted value of $\Delta E_t[w_{ijt}]/M_{it-1}$ for each empirical value of r_{it-1}, and then regress predicted $\Delta E_t[w_{ijt}]/M_{it-1}$ on empirical r_{it-1}. The p-values in Panel A test the hypothesis that the data moment equals the model moment. Panel B shows results from GMM's test of over-identifying restrictions. Panel B's p-value jointly tests whether all 12 empirical moments equal their corresponding moment from the model.

Panel A: Moments from GMM 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	0.154	0.000	0.176	0.154	0.000
$var(r_{it})$,	tenure=2	0.156	0.153	0.444	0.156	0.153	0.444
$var(r_{it})$,	tenure=3	0.145	0.153	0.035	0.145	0.153	0.035
$var(r_{it})$,	tenure=4	0.145	0.152	0.069	0.145	0.152	0.069
$var(r_{it})$,	tenure=5	0.138	0.151	0.001	0.138	0.151	0.001
$var(r_{it})$,	tenure=6	0.149	0.151	0.712	0.149	0.151	0.712
$var(r_{it})$,	tenure=7	0.144	0.150	0.145	0.144	0.150	0.145
$var(r_{it})$,	tenure=8	0.144	0.150	0.193	0.144	0.150	0.193
$var(r_{it})$,	tenure=9	0.148	0.150	0.728	0.148	0.150	0.728
$var(r_{it})$,	tenure=10+	0.154	0.149	0.009	0.154	0.149	0.009
Slope,	beliefs up	0.0028	0.0028	0.998	0.0050	0.0050	0.998
Slope,	beliefs down	0.0012	0.0012	0.997	0.0030	0.0030	0.998

Panel B: Test of Over-identifying Restrictions

Results using w^{grant}		Results using w^{vest}	
χ^2=	59.6	χ^2=	59.9
p-value=	0.000	p-value=	0.000

Table 3: **GMM Parameter Estimates**

This table presents the GMM estimates of the model's parameters. The first row contains results from estimating the model using data on w^{grant}, which is CEO pay, including stock and option grants in the year when they vest. The second row contains results using data on w_t^{grant}, which is CEO pay, including stock and option grants in the year when they were granted. 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 receives a fraction θ^{up} (θ^{down}) of the resulting positive (negative) surplus.

	Prior uncertainty σ_0	Volatility of profitability σ_ϵ	CEO's share of (+) surplus θ^{up}	CEO's share of (−) surplus θ^{down}
Using w^{grant}:	0.045	0.349	0.198	0.084
	(0.008)	(0.008)	(0.070)	(0.031)
Using w^{vest}:	0.047	0.350	0.332	0.202
	(0.008)	(0.007)	(0.103)	(0.064)

Table 4: **Estimated Model versus Counterfactual Benchmark Models**

This table compares data, predictions from the estimated model, and predictions from alternate models. Models differ only by their assumed values of θ^{up} and θ^{down}, which are in Panel A. Other parameter values are set to their estimated values in Table 3, using $w^{(grant)}$. In Panel B, "Slope, beliefs up (down)" is the slope from an OLS regression of scaled changes in expected pay $= \Delta E_t[w_{ijt}]/M_{it-1}$ on lagged excess returns $= r_{it-1}$, in the subsample where lagged excess returns are above (below) the median. Standard errors are in parentheses. In Panel C, the slope of excess stock returns on changes in perceived ability is from equation (21). In Panel D, the formula for the predicted increase in % return variance around unanticipated CEO deaths is in the Appendix. I assume θ is the average of the estimates of θ^{up} and θ^{down} from Panel A. The empirical value in Panel D is derived from Table 4 in Johnson et al. (1985), and equals the variance of returns in the announcement period minus the average variance in each of 10 previous pre-announcement days. Predictions in Panel C and D assume the CEO's current and final tenures equal 5 and 10 years, respectively.

Panel A: Model assumptions							
			Estimated	All surplus	No surplus	Powerful CEO	Powerless CEO
Assumed	θ^{up}:		0.198	1	0	1	0
Assumed	θ^{down}:		0.084	1	0	0	1
Panel B: Slope of scaled changes in expected pay on lagged stock returns							
		Data	Models				
Slope, beliefs up		0.0028	0.0028	0.0147	0.0000	0.0147	0.0000
		(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
R^2		0.07	0.99	0.99	1.00	0.99	1.00
Slope, beliefs down		0.0012	0.0012	0.0149	0.0000	0.0000	0.0149
		(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
R^2		0.01	0.98	0.98	1.00	1.00	0.98
Panel C: Slope of excess returns on changes in perceived CEO ability							
			Models				
Slope, good news			72.7	69.6	73.5	69.6	73.5
Slope, bad news			73.2	69.6	73.5	73.5	69.6
Panel D: Increase in % return variance around unanticipated CEO deaths							
		Data	Models				
		1.3	22.3	0.0	30.2	N/A	N/A

The Rodney L. White Center for Financial Research

The Wharton School
University of Pennsylvania
3254 Steinberg Hall-Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104-6367

(215) 898-7616
(215) 573-8084 Fax
http://finance.wharton.upenn.edu/~rlwctr

The Rodney L. White Center for Financial Research is one of the oldest financial research centers in the country. It was founded in 1969 through a grant from Oppenheimer & Company in honor of its late partner, Rodney L. White. The Center receives support from its endowment and from annual contributions from its Members.

The Center sponsors a wide range of financial research. It publishes a working paper series and a reprint series. It holds an annual seminar, which for the last several years has focused on household financial decision making.

The Members of the Center gain the opportunity to participate in innovative research to break new ground in the field of finance. Through their membership, they also gain access to the Wharton School's faculty and enjoy other special benefits.

Members of the Center
2009 – 2010

Directing Members

Aronson + Johnson + Ortiz, LP
Ballyshannon Partners
Goldman Sachs & Company
The Nasdaq OMX Educational Foundation
The Terker Family
The Vanguard Group

Founding Members

Ford Motor Company Fund
Merrill Lynch, Pierce, Fenner & Smith, Inc.
Oppenheimer & Company
Philadelphia National Bank
Salomon Brothers
Weiss, Peck and Greer