OPTIMAL INCENTIVE CONTRACTS IN THE PRESENCE OF CAREER CONCERNS: THEORY AND EVIDENCE

Robert Gibbons
Kevin J. Murphy

Number 563 August 1990
OPTIMAL INCENTIVE CONTRACTS IN THE
PRESENCE OF CAREER CONCERNS:
THEORY AND EVIDENCE

Robert Gibbons
Kevin J. Murphy

Number 563 August 1990
MEXICO: REAL PER CAPITA INCOME
(INDEX 1980=100)

MEXICO: THE REAL WAGE IN MANUFACTURING
(INDEX 1980=100)
OPTIMAL INCENTIVE CONTRACTS IN THE PRESENCE OF CAREER CONCERNS: THEORY AND EVIDENCE*

Robert Gibbons  
*MIT and NBER*

Kevin J. Murphy  
*University of Rochester*

First Draft: July, 1989  
This Revision: August, 1990

Abstract

This paper studies career concerns—concerns about the effects of current performance on future compensation—and describes how optimal incentive contracts are affected when career concerns are taken into account. Career concerns arise frequently: they occur whenever the market uses a worker's current output to update its belief about the worker's ability and competition then forces future wages (or wage contracts) to reflect these updated beliefs. Career concerns are stronger when a worker is further from retirement, because a longer prospective career increases the return to changing the market's belief.

In the presence of career concerns, the optimal compensation contract optimizes total incentives—the combination of the implicit incentives from career concerns and the explicit incentives from the compensation contract. Thus, the explicit incentives from the optimal compensation contract should be strongest when a worker is close to retirement. We find empirical support for this prediction in the relation between chief-executive compensation and stock-market performance.

* We are very grateful to Bengt Holmström, David Kreps, Meg Meyer, and Jean Tirole for substantial advice and encouragement during this paper's long gestation. We are also grateful for helpful comments from David Card, Henry Farber, Seonghoon Joon, Eugene Kandel, Alan Krueger, Bruce Meyer, Sherwin Rosen, Michael Wiesbach, and Jerold Zimmerman, and from seminar audiences at BU, Harvard (Kennedy School), Kentucky, Northwestern, NYU, Oxford, Princeton, Queen's, Rochester, Yale (S.O.M.), and the 1989 summer meetings of the Econometric Society and of the NBER Labor Studies Group. Richard Sloan and Kathleen Jones provided able research assistance. Financial support from the following sources is also gratefully acknowledged: the NSF (grants IST 86-09691 and SES 88-09200), and the Industrial Relations Section at Princeton University (Gibbons); and the John M. Olin Foundation and the Managerial Economics Research Center, University of Rochester (Murphy).
OPTIMAL INCENTIVE CONTRACTS IN THE PRESENCE OF CAREER CONCERNS: THEORY AND EVIDENCE

Robert Gibbons and Kevin J. Murphy

1. Introduction

This paper studies career concerns—concerns about the effects of current performance on future compensation—and describes how optimal incentive contracts are affected when career concerns must be taken into account. Career concerns arise frequently: they occur whenever the (internal or external) labor market uses a worker's current output to update its belief about the worker's ability and then bases future wages on these updated beliefs. In such a setting, the worker will want to take actions the market cannot observe, in an attempt to increase output and thus influence the market's belief; in equilibrium, however, the market will anticipate these actions and so draw the correct inference about ability from the observed output. Career concerns are stronger when a worker is further from retirement, because a longer prospective career increases the return to changing the market's belief. A worker who is further from retirement thus is willing to take more costly unobservable actions in an attempt to influence the market's belief.

Career concerns were first discussed by Fama (1980), who argued that incentive contracts are not necessary because managers are disciplined through the managerial labor market: superior performances will generate high wage offers; poor performances, low offers. Holmström (1982) showed that although such labor-market discipline can have substantial effects, it is not a perfect substitute for contracts: in the absence of contracts, managers typically work too hard in early years (while the market is still assessing the manager's ability)
and not hard enough in later years. We conclude from Fama's and Holmström's work that contracts are necessary to provide managers with optimal incentives.

In this paper, we add contracts to the Fama-Holmström model. We show that career concerns can still create important incentives, even in the presence of incentive contracts. Accordingly, in the presence of career concerns, the optimal compensation contract optimizes total incentives—the combination of the implicit incentives from career concerns and the explicit incentives from the compensation contract. Because the implicit incentives from career concerns are weakest for workers close to retirement, explicit incentives from the optimal compensation contract should be strongest for such workers, while for young workers it could be optimal for current pay to be completely independent of current performance.

Our formal model examines the career concerns that arise from competition among current and prospective employers in an external labor market, but career concerns also arise in internal labor markets, in two ways. First, competition among divisions within a firm (or even among supervisors within the same division) can mimic the competition we study in the external labor market. Second, even if there is only one supervisor, career concerns arise if promotions are based on a worker's assessed ability but the supervisor cannot perfectly distinguish between effort and ability. Although we do not model career concerns in internal labor markets, we expect analogous results to hold: Implicit incentives from promotion opportunities should be weakest for workers close to retirement and stronger where promotion opportunities are plentiful rather than scarce (as in expanding organizations). Current pay should be most sensitive to current performance for workers close to retirement and for workers with no promotion opportunities (such as workers at the top of the corporate hierarchy or other job ladder, and workers in declining organizations).¹

In addition to analyzing the characteristics of optimal incentive contracts in the presence of career concerns, this paper also examines the empirical support for the career-concerns

¹ Rosen (1986) makes a similar observation in the context of elimination tournaments or promotion ladders: the incentive to win an early round can be big even if the prize for winning that round is small, because winning also buys entry to subsequent rounds with larger prizes.
model in a particular setting: the relation between chief-executive compensation and stock-market performance. We study longitudinal pay and performance data on a large sample of chief executive officers (CEOs) and find empirical support for the model. We estimate, for example, that each 10% change in shareholder wealth corresponds to 1.5% changes in cash compensation for CEOs less than three years from retirement, but only .9% pay changes for CEOs more than three years from retirement.

In order to analyze both the implicit incentives from career concerns and the explicit incentives from compensation contracts, our model incorporates several of the fundamental issues associated with wage determination: incentives, learning, market forces, contracts, and risk aversion. Of these five elements of the model, the first three are necessary if career concerns are to arise, the fourth is necessary if explicit incentives are to be considered, and the fifth (or something like it) is necessary so that optimal contracts do not completely eliminate career concerns. Other models of career concerns—such as Fama (1980), Holmström (1982), and MacLeod and Malcomson (1988)—incorporate incentives, learning, and market forces, but not risk aversion or contracts. Other dynamic agency models—such as Lambert (1983), Rogerson (1985), Murphy (1986), Holmström and Milgrom (1987), and Fudenberg, Holmström, and Milgrom (1987)—incorporate incentives, contracts, and risk-aversion, but not learning or market forces. Finally, other dynamic insurance models—such as Harris and Holmström (1982)—incorporate risk-aversion, contracts, learning, and market forces, but not incentives.

---

2 Career concerns are related to the "ratchet effect" that arises in dynamic models of regulated firms, such as Freixas, Guesnerie, and Tirole (1985), Baron and Besanko (1987), and Laffont and Tirole (1988). These models ignore the market forces analyzed in this paper because there does not exist a market of prospective regulators, but the assumption that the regulator cannot commit to ignore information once it has been revealed has a similar effect. Lazear (1986) and Gibbons (1987) study the ratchet effect in models of the employment relationship, but also ignore market forces because the worker's private information concerns a firm-specific attribute, such as job difficulty. Aron (1987) and Kanemoto and MacLeod (1987) study the ratchet effect in models of the employment relationship that include market forces because the worker's private information concerns a worker-specific attribute, such as ability.

3 Holmström and Ricart i Costa (1986) extend the Harris-Holmström model by adding an observable action based on private information, as opposed to the unobservable action based on symmetric information considered here. MacDonald (1982) and Murphy (1986) analyze the effects of learning and market forces on job assignment, rather than on insurance contracts.
2. Theoretical Analysis

Consider a worker who works for T periods. In period t, the worker controls a stochastic production process in which output \( y_t \) is the sum of the worker's ability \( \eta \), the worker's non-negative effort \( a_t \), and noise \( \epsilon_t \):

\[
(2.1) \quad y_t = \eta + a_t + \epsilon_t .
\]

Before production begins, there is symmetric (but imperfect) information about the worker's ability: the worker and all prospective employers believe that \( \eta \) is Normally distributed with mean \( m_0 \) and variance \( \sigma_0^2 \). The error terms are Normally distributed with mean zero and variance \( \sigma_e^2 \), and are independent of each other and of ability, \( \eta \).

We assume that employers are risk-neutral but that the worker has the following exponential utility function:

\[
(2.2) \quad U(w_1, ..., w_T; a_1, ..., a_T) = -\exp \left( -r \left\{ \sum_{t=1}^{T} \delta^{t-1} [w_t - g(a_t)] \right\} \right) ,
\]

where \( w_t \) is the wage paid in period t and \( g(a_t) \) is a (monetary) measure of disutility of effort. We assume further that \( g(a_t) \) is convex and satisfies \( g'(0) = 0 \), \( g'(\infty) = \infty \), and \( g'' \geq 0 \). (The assumption that \( g'' \geq 0 \) is a sufficient condition for certain maximization problems to be concave and for several intuitive comparative-static results to hold.) Note that (2.2) is not the additively separable exponential utility function often encountered in the literature: (2.2) implies that the worker is indifferent among all deterministic wage streams with constant present value (computed using \( \delta \) as the discount factor), just as if the worker had access to a perfect capital market. Our results are similar in spirit but more complicated to express under the more conventional assumptions that the worker's exponential utility function is additively separable and that the worker has no access to credit; see the Appendix. Given our empirical focus on
CEO pay, however, it seems more reasonable to assume that the CEO needs no rather than has no access to credit.

To keep our theoretical analysis simple, we make two assumptions about contracting possibilities: (A1) short-term (i.e., one-period) contracts are linear in output; and (A2) long-term (i.e., multi-period) contracts are not feasible. Our assumption that short-term contracts are linear is obviously convenient: we wish to formalize the idea that contractual incentives should be strong when career-concern incentives are weak, and the strength of the contractual incentives is easily summarized by the slope of the linear contract. Furthermore, Holmström and Milgrom (1987) demonstrate that a model much like the single-period version of our model (but lacking the uncertainty about the agent's ability, $\eta$) can be reinterpreted so as to ensure that the optimal contract is linear. (In the Appendix, we discuss the extent to which the Holmström-Milgrom argument extends to our model.)

Our assumption that long-term contracts are not feasible can be reinterpreted. We show (again, in the Appendix) that this assumption is equivalent to assuming that long-term contracts exist but must be Pareto-efficient at each date; that is, long-term contracts must be renegotiation-proof. The idea that contracts should be renegotiation-proof has been widely accepted in recent work; see, for example, Dewatripont (1989) on contracting under adverse selection, Fudenberg and Tirole (1989) on contracting under moral hazard, and Hart and Moore (1988) on incomplete contracting under symmetric information (i.e., Nash implementation). In our model, renegotiation between the worker and a single employer mimics the effect of competition among the current and prospective employers.\(^4\)

We now derive the optimal compensation contracts for the two-period case and then state the results for the T-period case.

---

\(^4\) Another potential role for long-term contracts is consumption-smoothing; see Lambert (1983), Rogerson (1985), and Murphy (1986). Fudenberg, Holmström, and Milgrom (1987) show, however, that if the agent has perfect access to credit then the optimal long-term contract is equivalent to a sequence of optimal short-term contracts. In our model, the fact that the worker needs no access to credit, as described in connection with (2.2), eliminates this consumption-smoothing role for long-term contracts.
The Two-Period Model

The timing of the two-period model is as follows. At the beginning of the first period, prospective employers simultaneously offer a worker single-period linear wage contracts of the form \( w_1(y_1) = c_1 + b_1y_1 \). (Recall that information is symmetric, although imperfect, so there is no need for employers to offer menus of contracts in order to induce workers to self-select.) The worker chooses the most attractive contract and begins production. At the end of the first period, the firm (i.e., the first-period employer) and the market (i.e., prospective employers) observe \( y_1 \) and then simultaneously offer the worker single-period linear wage contracts of the form \( w_2(y_2) = c_2 + b_2y_2 \). These second-period contract offers will of course depend implicitly on \( y_1 \), because first-period output will convey information about the worker's ability, as described below. The force of our assumption (A2), however, is that second-period contracts can depend on first-period output only in this implicit manner, rather than explicitly through commitment at the beginning of the first period. Once again, the worker is free to choose the most attractive contract.

Given these compensation contracts (where, for now, the parameters \( c_1, b_1, c_2, \) and \( b_2 \) are arbitrary), the worker's expected utility (from the perspective of the first period) is the following function of first- and second-period effort, \( a_1 \) and \( a_2 \):

\[
(2.3) \quad - \mathbb{E} \left\{ \exp \left( -\delta (c_1 + b_1(\eta + a_1 + \epsilon_1) - g(a_1)) - \delta (c_2 + b_2(\eta + a_2 + \epsilon_2) - g(a_2)) \right) \right\}.
\]

From the perspective of the second period, however, after \( a_1 \) has been chosen and \( y_1 \) has been observed, the worker's expected utility is

\[
(2.4) \quad - \exp (-\delta^{-1} [c_1 + b_1y_1 - g(a_1)]) \cdot \mathbb{E} \left\{ \exp \left( -\delta (c_2 + b_2(\eta + a_2 + \epsilon_2) - g(a_2)) \right) \mid y_1 \right\},
\]

so the worker's second-period effort choice problem reduces to
The worker's optimal second-period effort, $a_2^*(b_2)$, therefore satisfies the first-order condition

$$g'(a_2) = b_2.$$  

Competition among prospective second-period employers implies that the contract the worker accepts for the second period must earn zero expected profits, so (after normalizing the price of output to unity), $c_2(b_2)$ satisfies

$$c_2(b_2) = (1-b_2) E\{y_2 | y_1\}.$$  

Using (2.1), the conditional expectation of the worker's second-period output given the worker's first-period output equals the sum of the conditional expectation of the worker's ability and the optimal second-period effort induced by the contract:

$$E\{y_2 | y_1\} = E\{\eta | y_1\} + a_2^*(b_2).$$  

To compute the conditional expectation of the worker's ability, the market must extract the relevant information about the worker's ability from the observed first-period output, and this requires a conjecture about the worker's first-period effort.

Suppose the market conjectures that the worker's first-period effort was $\hat{a}_1$. (As will become clear below, in equilibrium the market's conjecture about the worker's effort is correct. We restrict attention to pure-strategy equilibria.) Using well known formulas from DeGroot (1970), the conditional distribution of $\eta$ given the observed first-period output $y_1$ is then Normal with mean

$$m_1(y_1, \hat{a}_1) = \frac{\sigma_e^2 m_0 + \sigma_0^2 (y_1 - \hat{a}_1)}{\sigma_e^2 + \sigma_0^2}.$$  

(2.9)
and variance

\[
(2.10) \quad \sigma_i^2 = \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2}.
\]

We include \( \hat{a}_1 \) in the notation \( m_1(y_1, \hat{a}_1) \) for expositional clarity in what follows. It is also convenient to define \( \Sigma_2^2 = \sigma_i^2 + \sigma^2 \), the conditional variance of \( \eta + \varepsilon_2 \) given the observed first-period output \( y_1 \).

Substituting the appropriate expressions into (2.4) yields the worker's expected utility (from the perspective of the second period) for an arbitrary \( b_2 \), given first-period output \( y_1 \). The market thus believes that the optimal second-period slope, \( b_2^* \), maximizes

\[
(2.11) \quad -E \left\{ \exp \left( -r[c_2(b_2) + b_2(\eta + a_2^*(b_2) + \varepsilon_2) - g(a_2^*(b_2))] \right) \mid y_1 \right\}
\]

\[= - \exp \left( -r[m_1(y_1, \hat{a}_1) + a_2^*(b_2) - g(a_2^*(b_2)) - \frac{1}{2} rb_2^2 \Sigma_2^2] \right), \]

where the righthand side of (2.11) is derived using the observation that if \( x \) is Normally distributed with mean \( \mu \) and variance \( \sigma^2 \) then

\[
(2.12) \quad E \{ \exp(-kx) \} = \exp(-k\mu + \frac{1}{2} k^2 \sigma^2).
\]

Optimizing (2.11) and implicitly differentiating (2.6) yields the first-order condition for \( b_2^* \), the optimal second-period slope:\(^5\)

\[
(2.13) \quad b_2 = \frac{1}{1 + r \Sigma_2^2 g''[a_2^*(b_2)]}.
\]

\(^5\) Note that (2.13) is identical to the result given in footnote 8 in Lazear and Rosen (1981), who derive it using a second-order Taylor approximation to the more general (static) utility function \( U(w - g(a)) \). Thus, (2.13) is precisely optimal given (2.2) and approximately optimal more generally.
Using the assumption that $g''(x) \geq 0$, it is straightforward to show both that (2.11) is strictly concave, so (2.13) is sufficient, as well as that $b_2^*$ decreases with both risk-aversion ($r$) and uncertainty ($\Sigma^2_2$). Note also that $b_2^*$ is independent of $y_1$, which greatly simplifies the analysis of implicit incentives from career concerns, because it implies that the effect of first-period output on the second-period contract (i.e., the career-concern effect) is limited to the intercept given in (2.7).

Given the optimal second-period contract derived above, with its implicit dependence on $y_1$, the worker's first-period incentive problem is to choose $a_1$ to maximize

\[
(2.14) \quad -E\left\{ \exp \left( -r[c_1 + b_1(\eta + a_1 + \epsilon_1) - g(a_1)] - r\delta[c_2(b_2^*) + b_2(\eta + a_2^*(b_2^*) + \epsilon_2) - g(a_2^*(b_2^*))] \right) \right\}.
\]

Substituting (2.8) and (2.9) into (2.7) yields

\[
(2.15) \quad c_2(b_2^*) = (1-b_2^*) \left( \frac{\sigma^2_e m_0 + \sigma^2_0 (y_1-\hat{a}_1)}{\sigma^2_e + \sigma^2_0} + a_2^*(b_2^*) \right).
\]

The worker's optimal first-period effort, $a_1^*(b_1)$, therefore satisfies the first-order condition

\[
(2.16) \quad g'(a_1) = b_1 + \delta \frac{\partial c_2}{\partial a_1} = b_1 + \delta (1-b_2^*) \frac{\sigma^2_0}{\sigma^2_0 + \sigma^2_e} = B_1.
\]

The total incentive for first-period effort, denoted by $B_1$, is thus the sum of the explicit incentive from the first-period compensation contract, $b_1$, and the implicit incentive from career concerns, $\delta(1-b_2^*)\sigma^2_0/(\sigma^2_0 + \sigma^2_e)$. The career-concerns incentive is positive, since $0<b_2^*<1$ from (2.13), and increases (decreases) with the uncertainty about ability (production), $\sigma^2_e (\sigma^2_0)$.

Note that the expressions above take the market's second-period conjecture about first-period effort, $\hat{a}_1$, as given. Thus, (2.16) characterizes the worker's best response to this conjecture. In equilibrium, the market's conjecture must be correct, but the necessary fixed-
point computation is trivial because (2.16) does not depend on \( \hat{a}_1 \). Therefore, the equilibrium conjecture is

\[(2.17) \quad \hat{a}_1 = a_1^*(b_1) . \]

Competition ensures that firms earn zero expected profits. Because of assumption (A2), expected profits must be zero in each period. Hence,

\[(2.18) \quad c_1(b_1) = (1-b_1)E(y_1) = (1-b_1)[m_0+a_1^*(b_1)] . \]

Substituting \( a_1^*(b_1) \) and \( c_1(b_1) \) into (2.14) and applying (2.12) yields the worker's expected utility (from the perspective of the first period) for an arbitrary \( b_1 \):

\[(2.19) \quad - \exp \left( -r[m_0 + a_1^*(b_1) - g(a_1^*(b_1))] - r\delta[m_0 + a_2^*(b_2) - g(a_2^*(b_2))] \right) \cdot \exp \left( \frac{1}{2} r^2 [(B_1+\delta b_2^*)^2 \Sigma_1^2 - 2B_1 \delta b_2^* \delta \Sigma_1] \right) , \]

where \( B_1 \) reflects the sum of explicit and implicit incentives, as defined in (2.16). The optimal first-period slope, \( b_1^* \), thus satisfies the first-order condition

\[(2.20) \quad b_1 = \frac{1}{1 + r \Sigma_1^2 g''[a_1^*(b_1)]} - \delta (1-b_2^*) \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} - \frac{r \delta b_2^* \sigma_0^2 g''[a_1^*(b_1)]}{1 + r \Sigma_1^2 g''[a_1^*(b_1)]} , \]

where \( \Sigma_1^2 = \sigma_0^2 + \sigma_\epsilon^2 \) is the variance of \( \eta + \epsilon_1 \).

The main conclusion from this two-period model is that \( b_1^* < b_2^* \). This result is generalized in the T-period model, so we give only the intuition here. As is apparent from comparing the three terms in (2.20) to the single expression in (2.13), three effects contribute to the result. The first term in (2.20) reflects a noise-reduction effect: learning about the
worker’s ability causes the conditional variance of output to decline over time ($\Sigma^2_2 < \Sigma^2_1$), so the optimal tradeoff between insurance and incentives shifts towards the latter over time—$1 / \{1 + r \Sigma^3_1 g'[a^*_2(b^*_2)]\} < b^*_2$. The second term in (2.20) is the career-concerns effect, familiar from (2.16); it implies that optimal explicit incentives are adjusted to account for career-concerns incentives by imposing a lower pay-performance relation when career concerns are high. The third term in (2.20) reflects a human-capital-insurance effect: risk-averse workers with uncertain ability want insurance against low realizations of ability; in our model this insurance must take the form of a reduction in the slope of the first-period contract.

The worker’s demand for human-capital insurance can be quite strong. In fact, it is optimal for total first-period incentives to be negative ($B_1 < 0$) if the benefits from insuring the worker against low realizations of ability exceed the benefits of providing positive effort incentives.\(^6\) The career-concerns effect in (2.20) suggests that optimal contracts can have negative slopes ($b_1 < 0$) even when total incentives are positive, but this possibility creates no difficulties for our analysis. Negative total incentives, in contrast, require a reinterpretation of the first-order condition (2.16). In order to handle the possibility of negative total incentives in a simple way, we now assume that effort can be either positive or negative and that $g(-\infty) = -\infty$; we continue to assume that $g(a)$ is convex and satisfies $g'(0) = 0$, $g'(\infty) = \infty$, and $g'' \geq 0$. We interpret negative effort as hiding or stealing output; our assumptions on $g(a)$ imply that such activities are increasingly difficult on the margin, as may be realistic.

The $T$-Period Model

The three equations that summarize the workings of the competitive labor market can now be written as follows: First, the zero-expected-profit constraint in period $t$ is

\(^6\) Suppose, for example, that there is no noise in production (i.e., $\sigma^2_2 = 0$). Then first-period output perfectly reveals the worker’s ability, so $\sigma^2 = 0$. Since there is no uncertainty in the second period (i.e., $\Sigma^2_2 = \sigma^2_2 + \sigma^2_3 = 0$), the optimal slope is $b^*_2 = 1$, which imposes substantial human-capital risk on the worker: second-period pay moves one-for-one with the worker’s actual ability, and not at all with the market’s assessment of the worker’s ability based on first-period output (because $c^2(b^*_2)$ in (2.15) is zero). If $\delta = 1$ and $g(a) = [a^2]/2$ then the total first-period incentive is $B_1 = b_1 = (1 - r \sigma^2_3)/(1 + r \sigma^2_3)$, which clearly can be negative.
(2.21) \[ c_t = (1-b_t)E\{y_t|y_{1},...,y_{t-1}\}. \]

Second, the conditional expectation \( E\{y_t|y_{1},...,y_{t-1}\} \) is

(2.22) \[ E\{y_t|y_{1},...,y_{t-1}\} = m_{t-1} + \hat{a}_t, \]

where \( m_{t-1} \) is the market's expectation of the worker's ability as of the beginning of period \( t \), and \( \hat{a}_t \) is the market's conjecture about the effort the worker will supply in period \( t \). Finally, the market's conditional expectation of the worker's ability, given the observed history of prior outputs \( (y_1,...,y_{t-1}) \) and its conjectures about prior effort levels \( (\hat{a}_1,...,\hat{a}_{t-1}) \), is

(2.23) \[ m_{t-1}(y_1,...,y_{t-1}; \hat{a}_1,...,\hat{a}_{t-1}) = \frac{\sigma_x^2 m_0 + \sigma_y^2 \sum_{\tau=1}^{t-1} (y_{\tau} - \hat{a}_{\tau})}{\sigma_x^2 + (t-1)\sigma_y^2}. \]

Substituting these three equations into the worker's utility function (2.2) and applying (2.12) yields the worker's expected utility from a sequence of contracts with arbitrary slopes \( (b_1,...,b_T) \) and associated intercepts given by (2.21), given the market's effort conjectures \( (\hat{a}_1,...,\hat{a}_T) \) and the worker's effort choices \( (a_1,...,a_t) \). (See the Appendix for this computation, as well as the others omitted below.) The worker's optimal effort level in period \( t \), \( a_t^* \), then solves a first-order condition analogous to (2.16):

(2.24) \[ g'(a_t^*) = b_t + \sum_{\tau=t+1}^{T} \delta^{T-\tau} \frac{\partial c_{\tau}}{\partial a_t} = b_t + \sum_{\tau=t+1}^{T} \delta^{T-\tau} (1-b_{\tau}) \frac{\sigma_0^2}{\sigma_x^2 + (T-\tau-1)\sigma_y^2} = B_t. \]

Naturally, \( a_t^* \) depends on \( b_t \) and \( (b_{t+1},...,b_T) \), but not on \( (b_1,...,b_{t-1}) \).

In equilibrium, the market's conjectures are correct: \( (\hat{a}_1,...,\hat{a}_T) = (a_1^*,...,a_T^*) \). The worker's expected utility from the sequence of contracts with slopes \( (b_1,...,b_T) \) can then be written as
\begin{equation}
(2.25) \quad - \exp \left( -r \left[ \sum_{t=1}^{T} \delta^{t-1} \left( m_0 + a_t^* - g(a_t^*) \right) \right] + \frac{1}{2} r^2 \left[ \sum_{t=1}^{T} \delta^{t-1} B_t^2 \sigma^2_\varepsilon + \sum_{t=1}^{T} \delta^{t-1} B_t \sigma^2_\epsilon \right] \right).
\end{equation}

For now, take \((b_2, ..., b_T)\) as given. Recall that changes in \(b_1\) affect only \(a_1^*\). Since \(g'(a_1^*) = B_1\) from (2.24), it is convenient to optimize (2.25) with respect to the first-period total incentive, \(B_1^*(b_2, ..., b_T)\), which solves the first-order condition

\begin{equation}
(2.26) \quad B_1 = \frac{1}{1 + r \sum_{t=2}^{T} \delta^{t-1} B_t} - \frac{r \sigma^2_\varepsilon g''(a_1^*) \sum_{t=2}^{T} \delta^{t-1} B_t}{1 + r \sum_{t=2}^{T} \delta^{t-1} B_t}.
\end{equation}

Applying (2.24) then yields the optimal first-period contract slope \(b_1^*(b_2, ..., b_T)\),

\begin{equation}
(2.27) \quad b_1 = \frac{1}{1 + r \sum_{t=2}^{T} \delta^{t-1} (1-t_1) \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + (t-1)\sigma^2_\epsilon}} - \frac{r \sum_{t=2}^{T} \delta^{t-1} B_t}{1 + r \sum_{t=2}^{T} \delta^{t-1} B_t}.
\end{equation}

Finally, using (2.27) recursively yields the optimal slopes \((b_1^*, ..., b_T^*)\), beginning with \(b_T^*\) and working backwards to \(b_1^*(b_2^*, ..., b_T^*)\).

We now provide a partial characterization of the optimal slopes \((b_t^*: t = 1, ..., T)\): \(b_t^* < b_t^* \) for all \(t < T\); that is, contractual incentives are strongest for those about to retire. (See the Appendix for the derivation.) We suspect that the optimal slopes typically increase monotonically with \(t\) (in fact, we have been unable to construct a counter-example), but the result that \(b_t^* < b_T^* \) for all \(t < T\) is sufficient for our empirical work, in which we show that the sensitivity of pay to performance is greater for CEOs less than \(\tau\) years from retirement than for those more than \(\tau\) years from retirement, where \(\tau\) equals two, three, or four years.
3. Evidence on Career Concerns for Chief Executive Officers

Testing our model involves estimating the pay-performance relation between the agent's compensation (i.e., the wage, w, in our model's notation) and the principal's objective (i.e., output net of the wage, y-w) and detecting changes in the estimated pay-performance relation (i.e., the slope of the compensation contract, b) as the worker nears retirement. Obtaining the data necessary to analyze the model is difficult because existing longitudinal datasets for rank-and-file workers that contain data on wages seldom also contain data on performance (measured at the individual, divisional, or firm level).

We consider the agency relationship between the shareholders and the CEO of a publicly held corporation. We focus on this particular agency relationship for several reasons. First, the shareholder-CEO relationship is an archetypal principal-agent relationship: widely diffuse shareholders, through delegation of authority to the board of directors, hire managers to supply effort (i.e., take actions that increase shareholder wealth). Second, detailed longitudinal data on the CEO's compensation are publicly available in corporate proxy statements issued in conjunction with annual shareholders' meetings. Third, under the assumption that shareholders are approximately risk neutral, the principals have a common objective that is easily measured with available data: shareholders desire to maximize their wealth, as measured by the market value of the firm's common stock.\(^7\)\(^8\)

Although the managerial labor market is an attractive laboratory in which to investigate many of the characteristics and effects of incentive compensation policies, it may not be ideally suited for an investigation of career concerns. The career-concerns effects analyzed in our model require uncertainty about the worker's ability, but most CEOs are long-time employees

---

\(^7\) The assumption that well diversified shareholders are risk neutral with respect to the returns of any given firm is strictly correct only when the firm's returns are uncorrelated with market returns.

\(^8\) We ignore debt by focussing on shareholder wealth instead of the combined wealth of shareholders and bondholders; Smith and Warner (1979) argue that explicit bond covenants are designed to mitigate the agency problems between managers and bondholders.
before being appointed CEO, so shareholders (as well as potential alternative employers) may have precise information about the ability of a new CEO. In addition, the expected level of CEO compensation in our model is determined by a competitive market of prospective employers who continuously revise their bids for the CEO’s service as new information is revealed regarding the CEO’s managerial ability. The fact that CEOs rarely leave one to join another, coupled with the likely existence of large amounts of organization-specific human capital, suggests that a competitive market of the kind we assume may not exist.

We believe that these issues do not pose substantial problems for our empirical analysis, for the following reasons. First, shareholders are likely to be uncertain about the ability of a newly appointed CEO because the skills required to pilot a corporation are quite different from the skills required at lower levels in the organization; the performance of an individual as vice president, for example, is unlikely to yield precise information about the individual’s potential performance as CEO. Furthermore, this uncertainty may not be resolved quickly, both because it is difficult to isolate the CEO’s contribution from other factors that determine firm performance and because (as we discuss below) a change in the firm’s environment may renew the shareholders’ uncertainty about the CEO’s ability, even if his ability in the previous environment had become known.

Second, our assumption of a competitive market of prospective employers implies that a CEO receives the full benefit from an increase in the market’s estimate of his managerial ability (and bears the full cost from a decrease in the estimate). Our qualitative results are unchanged, however, if we assume instead that the CEO receives (bears) some, rather than all, of the increased (decreased) rent associated with changes in the market’s estimate of his ability. One model that would predict such rent sharing is a bilateral-monopoly model of the relationship between the CEO and his shareholders, where the rent to be divided is the return

---

9 The median CEO in our sample of nearly 3,000 CEOs described below worked in his firm 16 years before ascending to the top position.
10 Note, however, that Fama’s (1980) and Holmström’s (1982) early work on career concerns for top managers relies exclusively on just such a managerial labor market to provide managerial incentives.
on the CEO's organization-specific capital: in each period, the slope of the CEO's linear compensation contract under this new formulation of our model would be determined by the same kinds of (implicit and explicit) incentive considerations that arise in our original model, but the intercept would be determined by bargaining power rather than the zero-expected-profit constraint implied by a competitive market of potential employers.\footnote{If the sole objective of this paper were to deepen our understanding of CEO compensation then we would have developed this rent-sharing model rather than the competitive-market model developed above. We think it likely, however, that career concerns arise for many workers other than CEOs and that the competitive-market model applies more naturally than does the rent-sharing model for many of these other workers. Thus, in choosing to develop the competitive-market model we were in part attempting to convey this larger potential scope of the career-concerns model.}

A Longitudinal Sample of Chief Executive Officers

We test the implications of our career-concerns model using data constructed by following all chief executive officers listed in the Executive Compensation Surveys published in Forbes from 1971 to 1989. These surveys, derived from corporate proxy statements, include 2,972 executives serving in 1,493 of the nation's largest corporations during the fiscal years 1970-1988, or a total of 15,148 CEO-years of data. In order to distinguish CEOs close to retirement from CEOs with longer horizons, we initially restrict our analysis to CEOs who left their firms during the 1970–1988 sample period, using data from the 1990 Forbes survey to identify CEOs whose last fiscal year was 1988. This subsample includes 1,631 CEOs, representing 916 firms and 8,786 CEO-years.

Our model assumes that a CEO experiences something like the following career path: an executive is appointed CEO, is paid based on the performance of his firm, and remains in the CEO position until retirement, at which point his career ends. We begin our empirical analysis by presenting results that suggest that such a career path is typical. Figure 1 presents histograms describing the frequency distributions of tenure in the firm, tenure as CEO, and age, all measured at the date the CEO leaves office, for the 1,631 CEOs leaving their office during the 1970-1988 sample period. Panels A and B of Figure 1 show that, upon leaving
office, the average CEO has been in his firm for almost thirty years and has been in the CEO position for over ten years. Panel C shows that most executives leave their position near normal retirement age: 60% were between 60 and 66 when they left, and 31% were ages 64 or 65.

There is also evidence from other sources that when the typical CEO leaves office he either steps down into retirement or continues to serve the firm in a reduced, semi-retirement capacity. Vancil (1987) estimates that 75% of departing CEOs remain on their firms' boards of directors. This evidence seems inconsistent with the hypothesis that CEOs move from firm to firm, since CEOs joining rival firms are unlikely to maintain close relationships with their former employers. Vancil finds that an additional 6.4% die while in office; many others resign because of ill health. In our own sample, we find that only 36 of the 1,631 (2.2%) who left their firm during the sample period became CEO of another sample firm by the end of the period. In addition, we searched (the 1985 edition of) Who's Who in America for post-departure information on the subset (of about half) of our 1,631 departing CEOs who left their firms before 1985. We found that fewer than 6% of these CEOs joined another corporation after leaving their firm; another 1.4% joined law firms or universities. We were unable to find post-departure Who's Who information on 40% of the departed CEOs (most of whom appeared in Who's Who prior to their departure); we believe that the most likely cause for the absence of these data is retirement or death.

Figure 1 also presents a histogram describing the frequency distribution of years remaining as CEO for the 8,786 CEO-year observations from the subsample of CEOs leaving their firms during the sample period: Panel D shows that about half the CEO-year observations are from CEOs in one of their last three full fiscal years in office. Note that the distribution in Panel D is not representative of the full population, since CEOs who leave their firms during the sample period can have at most 18 years left in office—the number of years covered by our longitudinal dataset.
Figure 1 is based on all 1,631 CEOs leaving office during the 1970-1988 sample period. In our empirical work, however, we analyze first-differences and eliminate observations with missing pay and performance data; this leaves 1,292 CEOs representing 785 corporations, or 6,737 CEO-year observations. In what follows we refer to the latter subsample as the *completed-spells* subsample. Table 1 presents summary statistics for the completed-spells subsample. Column (1) shows that the average CEO in the subsample has been in his position for almost nine years, and will continue in his position for an additional three to four years. He receives annual compensation (excluding both grants of and gains from exercising stock options) of $616,300, of which 90% comes in the form of a base salary and annual bonus. All monetary variables in Table 1 and subsequent analyses are adjusted for inflation (using the consumer price index for the closing month of the fiscal year) to represent 1988-constant dollars. Adjusted for inflation, the average CEO's salary and bonus has grown by 6.6% per year.

We matched the *Forbes* compensation data to fiscal-year sales and stock-price performance data obtained from the Compustat data files. Average firm sales exceed $4 billion in 1988-constant dollars (although the median is only $2.2 billion). Shareholders have realized average inflation-adjusted returns of 1.8% over the sample period (in firms in the completed-spells subsample). We define the change in shareholder wealth, $\Delta V_t$, as the inflation-adjusted market value of common stock (in millions) at the beginning of the fiscal year multiplied by the inflation-adjusted rate of return on common stock (including splits and dividends): $\Delta V_t = V_{t-1} r_t$. The sample average change in shareholder wealth is a $56 million loss with a standard deviation of $1.9 billion.

Columns (2) and (3) of Table 1 report the same summary statistics for two subsamples of the completed-spells subsample: column (2) consists of CEO-year observations in which the CEO is in his last three full fiscal years; column (3) consists of the complementary subsample. Panel A shows that CEOs near retirement are older and have served longer in both their firm and their position than CEOs with many years remaining. Panel B shows that CEOs near
retirement are better paid than CEOs with many years remaining. The means and standard deviations of pay changes and percentage pay changes are higher for CEOs near retirement; we interpret these results as indirectly supporting the hypothesis that pay becomes more sensitive to performance (and hence more variable) as CEOs approach retirement. Panel C shows that average firm size (as measured by either sales or market value) and average firm performance (as measured by either the change in shareholder wealth or the shareholders' rate of return) are similar in the two subsamples.

The last row of Table 1 shows that CEOs with many years until retirement are over-represented in the early years of the sample. This over-representation is a direct result of our methodology: the completed-spells subsample includes only CEOs actually leaving office between 1970 and 1988, so there cannot be any CEO-year observations with more than three years remaining in the last three sample years, 1986-88.

An Empirical Specification for CEO Contracts in the Presence of Career Concerns

Our model describes the optimal linear relation between compensation and output in period t, \(w_t\) and \(y_t\). We assume that shareholders seek to maximize their wealth, and we consider two alternative empirical representations of output, \(y_t\), in terms of shareholder wealth, \(V_t\). First, one could interpret \(y_t\) as the change in shareholder wealth in period t; in this case, the linear compensation contract \(w_t(y_t) = c_t + b Ty_t\) simply becomes \(w_t = c_t + b_t V_t\). Alternatively, one could interpret output \(y_t\) as end-of-period firm value, \(y_t \equiv V_t\). In an efficient capital market, beginning-of-period firm value equals the conditional expectation of end-of-period firm value: \(V_{t+1} = E\{y_t|y_1, ..., y_{t-1}\}\). Therefore, the change in shareholder wealth in period t is \(\Delta V_t = y_t - E\{y_t|y_1, ..., y_{t-1}\}\), and the linear compensation contract \(w_t = c_t + b_1 y_t\) becomes

\[
(3.1) \quad w_t = c_t + b_t \{\Delta V_t + E\{y_t|y_1, ..., y_{t-1}\}\}.
\]
Both approaches share the feature that the coefficient on $\Delta V_t$ is $b_t$; consequently, the two approaches yield qualitatively similar results. In what follows, we take the second approach, largely because it is more tractable (as will become clear below).

Substituting (2.21) and (2.22) into (3.1) yields

\begin{equation}
(3.2) \quad w_t = m_{t-1} + \hat{a}_t + b_t \Delta V_t,
\end{equation}

but (3.2) is not convenient to estimate because $m_{t-1}$ is an individual-specific time-varying variable that depends on all prior realizations of output, as described in (2.23). Because we interpret $y_t$ as end-of-period firm value, however, the first-difference of (3.2) is a more tractable empirical specification: Substituting (2.22) into the definition of change in shareholder wealth, $\Delta V_t = y_t - E\{y_t | y_{t-1}, \ldots, y_{t-1}\}$, yields $\Delta V_t = y_t - \hat{a}_t - m_{t-1}$. Substituting the lagged version of this expression, $\Delta V_{t-1} = y_{t-1} - \hat{a}_{t-1} - m_{t-2}$, into (2.23) then yields

\begin{equation}
(3.3) \quad m_{t-1} = m_{t-2} + \frac{\sigma_0^2}{\sigma_e^2 + (t-1)\sigma_0^2} \Delta V_{t-1}.
\end{equation}

Thus, the first-difference of (3.2) implies that year-to-year changes in CEO compensation are a function of the change in CEO effort and this year's and last year's change in shareholder wealth:

\begin{equation}
(3.4) \quad \Delta w_t = \Delta \hat{a}_t + b_t \Delta V_t + \left( \frac{\sigma_0^2}{\sigma_e^2 + (t-1)\sigma_0^2} - b_{t-1} \right) \Delta V_{t-1},
\end{equation}

which suggests the empirical specification

\begin{equation}
(3.5) \quad \Delta w_t = \alpha_t + \beta_t \Delta V_t + \gamma_t \Delta V_{t-1},
\end{equation}

where $\alpha_t \equiv \Delta \hat{a}_t$, $\beta_t \equiv b_t$, and $\gamma_t \equiv \left( \frac{\sigma_0^2}{\sigma_e^2 + (t-1)\sigma_0^2} - b_{t-1} \right)$. 

In our model, we take the worker's total career to be fixed. A worker with fewer periods remaining therefore has worked for more periods, so the market has a more precise belief about his ability. As Panel A of Figure 1 illustrates, however, CEOs have heterogeneous career lengths (i.e., completed durations in office). In formulating testable hypotheses based on the empirical specification (3.5), we therefore distinguish between the number of years remaining in a CEO's career and the number of years already spent in office (tenure). For ease of exposition, we state these hypotheses for the case in which the optimal slope $b_t^*$ increases monotonically with $t$; analogous hypotheses can be derived from the result that $b_t^* < b_T^*$.

**H1.** Holding tenure as CEO constant, the slope of the compensation contract, $\beta_t$, increases as the CEO nears retirement.

Career concerns provide fewer incentives as the CEO nears retirement, so the incentives provided by current compensation increase.

**H2.** Holding years remaining as CEO constant, the slope of the compensation contract, $\beta_t$, increases with tenure as CEO.

The optimal relation between pay and current performance in part reflects the trade-off between the goal of providing the CEO with incentives to increase shareholder wealth and the goal of providing efficient risk-sharing for the risk-averse CEO. The optimal pay-performance relation increases when the variance of output decreases, since at the same pay-performance relation the CEO bears less risk as variance decreases. The CEO's true managerial ability becomes estimated more precisely as tenure increases; consequently, the variance of $y_t$ decreases and the pay-performance relation increases as the CEO's tenure increases.

**H3.** Holding tenure as CEO constant, the coefficient on shareholder-wealth change in period t-1, $\gamma_t$, decreases as the CEO nears retirement.

The coefficient on the change in shareholder wealth in period $t-1$ has two components, $\gamma_t^2/(\sigma_e^2 + (t-1)\sigma_e^2)$ and $-b_{t-1}$. The first term reflects the effect of $\Delta V_{t-1}$ on the updated estimate of ability in period $t$; this effect depends only on tenure as CEO and is independent of years to go. The second term decreases as the CEO nears retirement (i.e., $b_{t-1}$ increases) since career concerns provide smaller incentives.

**H4.** Holding years remaining as CEO constant, the coefficient on the change in shareholder wealth in period t-1, $\gamma_t$, decreases with tenure as CEO.
Both $\sigma^2/(\sigma^2+(t-1)\sigma^2)$ and $-b_{t-1}$ decrease as tenure increases because the estimate of managerial ability becomes more precise.

**Performance Pay and Years Left as CEO**

Hypothesis H1 predicts that the relation between CEO pay changes and current performance will be higher for executives close to retirement. A simple way to evaluate this hypothesis is to estimate a regression that allows the pay-performance relation to vary with years remaining as CEO:

\[
\Delta w_{it} = \alpha + \beta \Delta V_{it} + \sum_{\tau=0}^{18} \beta_{\tau}(\text{CEO has } \tau \text{ years left})_{it} \Delta V_{it},
\]

where $(\text{CEO has } \tau \text{ years left})_{it}$ is a dummy variable equal to one if the $i^{th}$ CEO has $\tau$ years remaining in his career in fiscal year $t$. Under Hypothesis H1, the slope of the pay-performance relation increases as the CEOs nears retirement, $\beta_0 > \beta_1 > \ldots > \beta_{18}$. There are at least four potential problems with estimating (3.6) directly, however. First, our empirical definition of shareholder-wealth change, $\Delta V_{it}$, is net of payments to the CEO, while our theoretical measure is before these payments. This difference has a negligible effect on our analysis, since CEO compensation is trivial compared to changes in the value of the firm. Second, while $\Delta w_{it}$ should include all forms of compensation, the *Forbes* surveys include data on salaries, bonuses, and some minor additional forms of compensation but do not include data on grants of stock and stock options. To the extent that a CEO's holdings of stock in the firm increase with tenure, stockholdings will act to increase incentives as career-concern incentives decrease, so our estimates of the change in the pay-performance relation as the CEO nears retirement may underestimate the shareholders' total response to the issue of decreasing career concerns. To the extent that CEOs decrease their stock holdings as they near retirement, however, our estimates may simply reflect the effort to provide incentives through compensation to replace those formerly induced through stock ownership.
A third problem with direct estimation of (3.6), as noted in connection with Table 1, is that CEOs near retirement are over-represented in the later years of the completed-spells subsample, so the estimates of the $\beta$'s from (3.6) for the CEO subsamples will reflect secular trends in incentive compensation in addition to the effects of career concerns. As shown in Appendix Figure A1, the relation between pay and performance has increased over the past decade, so the estimate of $\beta$'s for CEOs near retirement are likely to be biased upward.

Fourth, and perhaps most important, (3.6) assumes that the relation between pay and performance differs across CEOs (and within a CEO's career) only by the number of years remaining as CEO—(3.6) does not allow for other sources of heterogeneity in the pay-performance relation. One potentially important source of heterogeneity is firm size: the optimal pay-performance relation may decline with firm size both because the variance of changes in shareholder wealth increases with firm size and because the CEO's direct effect on firm value may decrease with firm size.

We show below (in connection with Table 3) that the pay-performance relation (i.e., $\Delta w_t / \Delta V_t$) varies significantly with firm size but that the pay-performance elasticity (i.e., $\Delta \ln(w_t) / \Delta \ln(V_t)$) is almost invariant to firm size. We therefore attempt to control for size-related heterogeneity by converting the regression variables to logarithmic changes and then estimating an elasticity form of (3.6) that allows the pay-performance elasticity to vary across years:

\[
(3.7) \quad \Delta \ln(w_{it}) = \alpha + \sum_{n=1972}^{1988} \alpha_n (n^{th} \text{ year dummy})_{it} \\
+ \left[ \beta_0 + \sum_{\tau=0}^{18} \beta_{\tau} (\text{CEO has } \tau \text{ years left})_{it} + \sum_{n=1972}^{1988} \beta_n (n^{th} \text{ year dummy})_{it} \right] * \Delta \ln(V_{it}),
\]

where $\Delta \ln(V_{it}) = \ln(1+r_{it})$ is the continuously accrued rate of return on common stock and $(n^{th} \text{ year dummy})_{it}$ is a dummy variable equal to one if $n=t$. We define $\Delta \ln(w_{it})$ as the annual logarithmic change in the CEO's salary and bonus (in thousands).\[12\]

\[12\] We also used a more comprehensive measure of compensation that includes fringe benefits, and contingent (but non-stock) remuneration. The qualitative results from these regressions are similar to those presented in the
Figure 2 depicts the estimated pay-performance elasticities from (3.7) for CEOs grouped based on their years remaining as CEO. (To simplify the figure, observations with 15 or more years remaining have been pooled.) The figure is drawn assuming a year effect of .06, which is the average of the 17 estimated year coefficients. CEOs in their final year, second-to-last year, or third-to-last year have estimated pay-performance elasticities of .179, .199, and .184, respectively. Each of these estimated elasticities is significantly higher than the estimated elasticity for CEOs in their fifth to last year (.120) or sixth-to-last year (.115); no other pairs of coefficients in Figure 2 (including pairs involving the highest estimated elasticity for CEOs in their eleventh-to-last year) are significantly different at the 5% level.

Figure 2 suggests that, in general, pay-performance elasticities are higher for CEOs nearing retirement, but also suggests that separate coefficients based on years remaining as CEO are estimated with large standard errors. Since the only significant differences in Figure 2 correspond to CEOs in their last three years vs. CEOs with more than three years remaining, we divided the completed-spell subsample into two groups and estimated the following regression:

\[
\Delta \ln(w_{it}) = \alpha_0 + \alpha_1(Few \ Years \ Left)_{it} + \sum_{n=1972}^{1988} \alpha_n(n^{th} \ year \ dummy)_{it} \\
+ [\beta_0 + \beta_1(Few \ Years \ Left)_{it} + \sum_{n=1972}^{1988} \beta_n(n^{th} \ year \ dummy)_{it}] \ast \Delta \ln(V_{it}),
\]

where \((Few \ Years \ Left)_{it}\) is a dummy variable equal to one if the \(i^{th}\) CEO is in the last three years of his career in fiscal year \(t\).

Column (1) of Table 2 reports estimates of regression (3.8) for the completed-spells subsample. CEOs serving in their last three years as CEO are assigned to the "Few Years Left" category. The (unreported) estimated coefficients on the (Shareholder Return)*(Year) interactions range from .02 to .20, indicating that CEOs with many years remaining receive between .2% and 2.0% raises for every 10% return realized by shareholders. The
Return*(Few Years Left) coefficient is also positive, suggesting that CEOs with few years remaining receive an additional .44% raise for every 10% return realized by shareholders. The Return*(Few Years Left) coefficient is highly significant (t=3.0), which we interpret as empirical support for the Hypothesis H1.

Our definition of Few Years Left—CEOs in their last three years—is admittedly arbitrary, and we re-estimated the regression in column (1) of Table 2 for three alternative definitions of Few Years Left: executives in their last one, two, and four years as CEO. The estimated Return*(Few Years Left) interaction coefficient is positive in all cases and is significant except when Few Years Left is defined as CEOs in their final year; we believe this insignificance reflects both small sample size and non-performance-related payments made to CEOs in their last year.

The results in Table 2 support hypothesis H1, but also support an alternative hypothesis: CEOs whose total stay in office is only a few years may have higher pay-performance elasticities than do CEOs who stay in office longer, independent of years remaining until retirement. We tested this alternative hypothesis by repeating the regressions in Table 2 for the subsample of CEOs who had spent at least five years in office upon retirement. None of the qualitative results in Table 2 are changed, suggesting that our results are not driven by cross-sectional (negative) correlation between total time in office and the pay-performance elasticity.

Performance Pay and CEO Tenure

Hypothesis H2 predicts that, holding years remaining as CEO constant, the pay-performance elasticity increases with tenure because managerial ability becomes estimated with less uncertainty. We test this hypothesis in the regression reported in column (2) of Table 2, which includes controls for CEO tenure: the new regressors are "Low Tenure," a dummy variable that equals one if the CEO is in his first four years as CEO, and Low Tenure interacted with shareholder return. Our model predicts a negative coefficient on the (Low
Tenure)\(^*(\text{Return})\) interaction. The positive (but insignificant) coefficient in column (2) is inconsistent with this hypothesis. Under a plausible alternative formulation of the theory, however, the pay-performance relation is not predicted to increase with tenure: suppose (as in Holmström, 1982) that managerial ability is not a fixed parameter \(\eta\) but rather varies over time as \(\eta_{t+1} = \eta_t + \nu_t\), where the innovations \(\nu_t\) are independent and Normally distributed. In the steady state, the market's uncertainty about the CEO's ability is time-invariant and so independent of the CEO's tenure. Thus, hypothesis H2 would not be a prediction of this alternative model.

*Estimates Based on the Full Sample*

The completed-spells subsample of 1,292 CEOs (6,737 CEO-years) who actually leave their firm during the 1970-1988 sample period represents only about half of the 1,900 CEOs (11,359 CEO-years) in the full sample (after constructing first differences and eliminating observations with missing data). CEOs in the full sample but not in the completed-spells subsample include CEOs still serving at the end of the 1988 fiscal year and CEOs of firms that were deleted from the *Forbes* survey.\(^{13}\) Data from these additional observations are useful for two reasons. First, although we cannot tell whether the CEO's final observed year corresponds to one of his last three years in office, for many CEO-year observations we can tell whether he has more than three years remaining as CEO. Second, data from these additional observations can be used to estimate more precisely the year effects and shareholder return*year interactions in columns (1) and (2) of Table 2.

Column (3) of Table 2 reports results from the re-estimation of column (2) for the *full sample* of CEOs, including CEOs who do not leave their firms during the sample period. The regression contains two additional explanatory variables: "Can't Tell If Few Years Left," a

---

13 A total of 623 firms were deleted from the *Forbes* surveys during the 1970-1988 sample period. Of these, 277 are still "going concerns" as of 1989, 290 were acquired by or merged with another firm (127 of these were acquired or merged within two years of the *Forbes* delisting), and 56 liquidated, went bankrupt, or went private.
dummy variable that equals one if the CEO-year observation is in the last three years of the data on a CEO who does not belong to the completed-spells subsample, and an interaction of this dummy variable with Shareholder Return. The results are similar to those reported in column (2) and are consistent with the primary implication of the career-concerns theory: the pay-performance elasticity increases as the CEO approaches retirement.

Years Remaining and Pay for Lagged Performance

Hypotheses H3 and H4 predict that the relation between current CEO pay changes and the previous year's performance should decline with both years remaining as CEO and CEO tenure. We re-estimated the regressions in Table 2 after including lagged shareholder return and interactions of lagged return with Few Years Left, Low Tenure, and the seventeen year dummy variables. The magnitude and significance of the coefficients reported in Table 2 were not affected by including lagged performance, reflecting the low correlation between current and past stock returns. The coefficients on the new interaction terms—(Lagged Return)*(Few Years Left) and (Lagged Return)*(Low Tenure)—were insignificant in all regressions. Thus, hypotheses H3 and H4 are not supported by our data.

Performance Pay and Expected Years Left as CEO

In formalizing the theory of optimal contracts in the presence of career concerns, we assume that the CEO's final year is known with certainty, but it seems likely that many executives retire earlier or later than they were expected to retire. The effect of career concerns on optimal explicit incentives may therefore depend more on the expected years remaining as CEO than on the CEO's actual years remaining (as measured after retirement has been observed).

As shown in Panel C of Figure 1, many executives leave their position when they reach age 64 or 65. One estimate of the expected years remaining is the number of years remaining until the CEO reaches 65. We re-estimated the regression in column (1) of Table 2 after re-
defining "Few Years Left" as "Few Expected Years Left," a dummy variable that equals one if the CEO is age 62 or older (i.e., within three years of reaching age 65, or older than age 65). The coefficient on the interaction term (Return)*(Few Expected Years Left) was positive but statistically insignificant. Thus, hypothesis H1 is not supported when "years remaining until age 65" is used as a proxy for expected years remaining.

We also conducted a three-stage investigation in which we first used the completed-spells subsample to estimate the CEO’s years remaining using information available to the board at the beginning of each period, including age, firm and CEO tenure, and prior sales, stock-price performance, and compensation. We then used these estimated parameters from the first stage to compute estimated years remaining for all CEO-year observations in the full sample, and set the dummy variable (Few Expected Years Left) equal to one for an observation if the estimated horizon was less than three years. Finally, we used this new dummy variable to re-estimate versions of regression 3.8) with and without year and tenure effects. The coefficient on the interaction term (Return)*(Few Expected Years Left) was positive but insignificant (using the uncorrected standard errors) in all regressions. Thus hypothesis H1 is not supported when expected years remaining is substituted for actual years remaining in this way.

The board of directors’ assessment of a CEO’s horizon is obviously based on information not available to us, including the CEO’s health, potential replacements, and implicit or explicit mandatory retirement policies. One interpretation of the insignificance from both experiments reported above—using either age-based or regression-based proxies for expected years remaining—is that actual years remaining (i.e., "Few Years Left" in Table 2) is a better proxy for expected years remaining than is either the age-based or the regression-based proxy. Both the age- and regression-based proxies tend to classify older executives into the Few Expected Years Left group, while younger CEOs tend to be excluded from this group. But many older CEOs have long remaining careers (and therefore have small explicit incentives, according to our model); including these CEOs in our "Few Expected Years Left" category
biases downward the coefficient on the (Return)*(Few Expected Years Left) interaction term if the market expects a lengthy subsequent career. Similarly, excluding young CEOs from the "Few Expected Years Left" category when they are in fact near retirement (and therefore have large explicit incentives, according to our model) also biases the interaction coefficient downward.

Other Sources of Heterogeneity in the Pay-Performance Relation

As noted previously, regression (3.6) allows the pay-performance relation to vary with years remaining as CEO, but ignores other potential sources of pay-performance heterogeneity, such as the effects of firm size. In this section, we justify our use of the logarithmic specification to control for size-related heterogeneity, and we also describe our attempts to control for other potential (but unknown) forms of heterogeneity.

Column (1) of Table 3 reports estimates of the non-logarithmic version of regression (3.8) for the 1,292 CEOs in the completed-spells subsample. The coefficient on the interaction \( \Delta(\text{Shareholder Wealth})*(\text{Few Years Left}) \) is positive but insignificant, suggesting that the pay-performance relation is independent of years remaining as CEO. The regression in column (2) includes two interaction terms, \( \Delta V_t*\text{Sales} \) and \( \Delta V_t*\text{Sales}^2 \), to allow the pay-performance relation to vary quadratically with firm size as measured by net sales (in millions of 1988-constant dollars). Based on average values for the \( (\Delta \text{Shareholder Wealth})*(\text{Year}) \) effects, the coefficients suggest that CEOs in median size firms (sales of $2.0 billion) receive average pay changes of $.93\varepsilon$ for every $1,000 change in share watershed wealth. CEOs at the 10th percentile (sales of $400 million) and 90th percentile (sales of $8.2 billion) receive average pay changes of $.99\varepsilon$ and $.70\varepsilon$, respectively, for every $1,000 change in shareholder wealth. Including the size interaction terms in column (2) increases the magnitude and significance of the \( \Delta V_t*(\text{Few Years Left}) \) coefficient, although it remains insignificant.

The increase in the interaction \( \Delta(\text{Shareholder Wealth})*(\text{Few Years Left}) \) from column (1) to column (2) of Table 3 suggests that it is important to control for size-related
heterogeneity in the pay-performance relation when testing for career concerns in CEO compensation contracts. One common way of controlling for size-related heterogeneity, based on the observation that the pay-performance elasticity is relatively invariant to firm size, is to convert the regression variables to percentage changes or logarithmic changes; see, for example, Coughlan and Schmidt (1985) and Gibbons and Murphy (1990).

Column (3) of Table 3 simply repeats the elasticity specification (3.8) reported in column (1) of Table 2. Column (4) of Table 3 includes interaction variables to allow the pay-performance elasticity to vary quadratically with size as measured by net sales. Although the size-interaction variables are individually insignificant, they are jointly significant (at the 2% level), indicating that pay-performance elasticities are slightly higher for larger firms. The estimated coefficient on the (Shareholder Return)*(Few Years Left) interaction in column (4), however, is extremely close in magnitude and significance to its counterpart in column (3). These results suggest that the logarithmic specification removes the size bias from the estimated career-concerns coefficient.

Although the logarithmic specification controls for size-related heterogeneity in the pay-performance relation, other sources of potentially important heterogeneity remain. Some of the heterogeneity reflects secular trends, which can be controlled for by including year interaction variables, but other possible sources of heterogeneity include reflect firm or industry factors not yet incorporated into our theory or specification. We attempted to control for these factors by allowing the pay-performance elasticity to vary across firms:

\[
\Delta \ln (w_{it}) = \alpha_i + \alpha_1 (\text{Few Years Left})_{it} + \sum_{n=1972}^{1988} \alpha_n (n^{th} \text{ year dummy})_{it} \\
+ [\beta_i + \beta_1 (\text{Few Years Left})_{it} + \sum_{n=1972}^{1988} \beta_n (n^{th} \text{ year dummy})_{it}] \ast \Delta \ln (V_{it}),
\]

where $\alpha_i$ and $\beta_i$ are firm-specific intercepts and pay-performance elasticities, respectively.
We estimated (3.9) for the subsample of 625 CEOs from the completed-spells subsample that had at least four years of data.\textsuperscript{14} The estimated coefficient on (Few Years Left)\textsubscript{11} is positive (\(\beta_1 = .0297\)) but insignificant (t=1.4). The point estimate, however, is similar in magnitude to the coefficient in column (1) of Table 2 (\(\beta_1 = .0436\)), which constrained the pay-performance elasticities to be equal across firms. Thus, our support of hypothesis H1 is somewhat diminished after controlling for firm-specific heterogeneity by allowing the pay-performance elasticity to vary across firms.

4. Conclusion

The driving force behind our theoretical analysis is that an individual’s actions are influenced by career concerns. We extend previous research by showing that career concerns can have important effects on incentives even in the presence of contracts. We also show that optimal compensation contracts neutralize career-concern incentives by optimizing the total incentives from the contract and from career concerns—explicit contractual incentives are high when implicit career-concern incentives are low, and vice versa. In developing our model, we have kept the analysis tractable by making strong assumptions. Weaker assumptions would undoubtedly allow additional effects to emerge, but our primary qualitative results—that career concerns affect incentives and that optimal contracts account for these implicit incentives—seem likely to be robust.

Perhaps the most natural application of the idea that optimal incentive contracts optimize total incentives is to promotions.\textsuperscript{15} It would be useful to integrate our model of optimal contracts in the presence of career concerns with a dynamic model of learning and job

\textsuperscript{14} We estimated (3.9) in two stages. In the first stage, we regressed the dependent and each of the independent variables in (3.9) on shareholder return, \(\Delta \ln(V_{11})\), by CEO. In the second stage, we regressed the residual from the first-stage regression involving the dependent variable in (3.9) on the residuals from the first-stage regressions involving the independent variables in (3.9). After adjusting standard errors for the correct degrees of freedom, the results from the second-stage regression are equivalent to estimating (3.9) with CEO-specific intercepts and slopes (on shareholder return).

\textsuperscript{15} Rosen’s (1986) model can be interpreted in terms of promotions but does not allow new jobs to have new technologies or for workers to be matched to jobs.
assignment. The latter could involve symmetric learning, as in MacDonald (1982) and Murphy (1986), or asymmetric learning (where the current employer learns a worker's ability faster than a prospective employer does), as in Waldman (1984) and Ricart i Costa (1988). Such a model would endogenize transitions between jobs, which would improve upon the T-period career we take as exogenous. We expect the development of such models, together with the development of new data sources on internal organization, to be an important part of future research on incentives and organizations.
Appendix

This Appendix discusses four issues: (1) the difference between the multiplicative exponential utility function (2.2) and the additive exponential utility function; (2) the assumption that short-term contracts are linear; (3) renegotiation-proof long-term contracts; and (4) the solution to the T-period model.

1) The Utility Function

We now adopt the more conventional assumptions that the worker’s exponential utility function is additively separable and that the worker has no access to credit. In the two-period case, these assumptions imply that

\[ (A1.1) \quad U(w_1, w_2; a_1, a_2) = -\exp (-r[w_1 - g(a_1)]) - \delta \exp (-r[w_2 - g(a_2)]) . \]

In order to focus on the role of career concerns, we continue to impose assumptions (A1) and (A2): short-term contracts are linear and long-term contracts are not feasible. The second-period problem is then the same whether we use (2.2) or (A1.1) for the utility function; in particular, (2.13) and (2.15) continue to hold. Given the market’s conjecture \( \hat{a}_1 \) and equilibrium play in the second period, however, the worker’s first-period problem is now to choose \( a_1 \) to maximize

\[ (A1.2) \quad -\mathbb{E}\left\{ \exp \left(-r[c_1 + b_1(a_1 + \epsilon_1) - g(a_1)]\right) \right\} \]

\[ -\delta \mathbb{E}\left\{ \exp \left(-r[c_2(b_2) + b_2^*(\hat{a}_1 + a_2^* + \epsilon_2) - g(a_2^*(b_2^*))]\right) \right\} \]

\[ = -\exp \left(-r[m_0 + (1-b_1)\hat{a}_1 + b_1a_1 - g(a_1) - \frac{1}{2} rb_1^2 \Sigma_1^2] \right) \]

\[ - \delta \exp \left(-r[m_0 + a_2^*(b_2^*) - g(a_2^*(b_2^*)) + (1-b_2^*)\rho(a_1-\hat{a}_1) - \frac{1}{2} r \Sigma_1^2(b_2^*(1-\rho^2) + \rho^2)] \right) . \]
where \( \rho \equiv \sigma^2_0 / (\sigma^2_0 + \sigma^2_\epsilon) \). The worker's optimal first-period effort, \( a_1^* (b_1) \), therefore satisfies a first-order condition analogous to (2.16):

\[
(A1.3) \quad g'(a_1) = b_1 + \delta (1 - b_2^*) \frac{\sigma^2_0}{\sigma^2_0 + \sigma^2_\epsilon} \cdot \frac{\text{EXP2}}{\text{EXP1}} = B_1 , \quad \text{where}
\]

\[
\text{EXP1} = \exp \left( -r \left[ (1 - b_1) a_1 + b_1 a_1 - g(a_1) - \frac{1}{2} r b_2^2 \sigma^2_1 \right] \right) \quad \text{and}
\]

\[
\text{EXP2} = \exp \left( -r \left[ a_2^* (b_2^*) - g(a_2^* (b_2^*)) + (1 - b_2^*) \rho (a_1^* - a_1) - \frac{1}{2} r \sigma^2_1 [b_2^2 (1 - \rho^2) + \rho^2] \right] \right) .
\]

Thus, the worker's equilibrium expected utility given an arbitrary first-period slope \( b_1 \) is

\[
(A1.4) \quad - \exp \left( -r \left[ m_0 + a_1 - g(a_1) - \frac{1}{2} r b_2^2 \sigma^2_1 \right] \right)
\]

\[
- \delta \exp \left( -r \left[ m_0 + a_2^* (b_2^*) - g(a_2^* (b_2^*)) - \frac{1}{2} r \sigma^2_1 [b_2^2 (1 - \rho^2) + \rho^2] \right] \right) ,
\]

where \( a_1 \) is given by (A1.3). The optimal \( b_1 \) therefore satisfies the first-order condition

\[
(A1.5) \quad [1 - g'(a_1)] a_1' = r b_1 \sigma^2_1 .
\]

Implicitly differentiating (A1.3) and solving for \( a_1' \) implies that

\[
(A1.6) \quad b_1 = \frac{1 - (B_1 - b_1) \left( 1 + r \Sigma^2_1 g''(a_1^*(b_1)) \right)}{1 + r \Sigma^2_1 g''(a_1^*(b_1))} .
\]

It remains only to show that \( b_1 \) as given by (A1.6) is less than \( b_2 \) as given by (2.13). Intuitively, this inequality holds for the same reasons given in the text: noise reduction, career concerns, and human-capital insurance. Formally, suppose \( b_1 > b_2 \). Since \( g'(a_1) = B_1 > b_1 > \)
\( b_2 = g'(a_2) \) and \( g'' > 0 \), it then follows that \( a_1 > a_2 \). Since \( g'' \geq 0 \), we have \( g''(a_1) \geq g''(a_2) \), so \( \sum g''(a_1) \geq \sum g''(a_2) \), but this implies \( b_1 < b_2 \), a contradiction.

2) Linear Short-Term Contracts

To keep our theoretical analysis simple, we assumed that short-term (i.e., one-period) contracts are linear in output. Holmström and Milgrom (1987) demonstrate that a model much like the single-period version of our model (but lacking the uncertainty about the agent's ability, \( \eta \)) can be reinterpreted so as to ensure that the optimal contract is linear. The key step in the Holmström-Milgrom model is to reinterpret output in the single-period model as the aggregate of outputs over time in an underlying dynamic model, as follows:

Suppose a worker is to be employed for a year, but can control the production process on a day-to-day basis. Each day's production is an independent Bernoulli trial (i.e., output is either high or low, and increased effort on a given day increases the probability of output being high that day). The worker observes output each day, and so can choose future effort levels in response to past output realizations. After rewriting the worker's utility function in (2.2) to accommodate this underlying dynamic model, Holmström and Milgrom show that the optimal compensation contract for the year is equivalent to repeating the static contract that is optimal for each one-day incentive problem. Therefore, the optimal contract for the year depends on only the aggregate output produced during the year and is linear in this aggregate.\(^{16}\) In the limit (as the output that a Bernoulli trial represents shrinks from that of a day to that of an instant, in a suitably normalized fashion), the worker continuously controls the drift of a one-dimensional Brownian motion. Again, the optimal compensation contract is linear in aggregate output, and now aggregate output is Normally distributed, just as is \( \varepsilon_1 \) in (2.1).

In the Holmström-Milgrom model, a contract that is linear in year-end aggregate output provides the agent with a constant incentive for effort on each day of the year (while a step-
function contract of the kind analyzed by Mirrlees (1974) would provide very different incentives). We believe that there is a good deal to be said for providing the agent with a constant incentive for effort, even when uncertainty about the agent’s ability creates a new need for insurance. Given the Holmström-Milgrom linearity result, our assumption (A1) amounts to assuming that this insurance is provided by reducing the slope of the linear contract that would be optimal in the absence of such uncertainty.

3) Renegotiation-proof Long-Term Contracts

We now relax our assumption that long-term contracts are not feasible. Instead, we assume that (linear) long-term contracts are feasible but must be Pareto-efficient (i.e., renegotiation-proof) at each date. To keep the argument simple, we consider only the two-period case. Suppose that at the beginning of the first period, prospective employers simultaneously offer a worker two-period contracts of the form \((w_1(y_1), w_2(y_1, y_2))\). We assume that the contract the worker accepts binds both the firm and the worker, but that the parties will renegotiate the contract if it is Pareto-inefficient at the beginning of the second period (i.e., after first-period output is observed).

We continue to assume that one-period contracts must be linear: \(w_1(y_1) = c_1 + b_1 y_1\) and \(w_2(y_1, y_2) = c_2 + b_2 1 y_1 + b_2 2 y_2\). Renegotiation-proofness then implies that \(b_2 2 = b_2^*\) from (2.13) and hence that the worker’s second-period effort choice is \(a_2^*(b_2^*)\) satisfying (2.6). Furthermore, given the utility function (2.2), we can set \(b_2 1 = 0\): any desired dependence of first- and second-period wages on first-period output can be replicated through the appropriate choice of \(b_1\). Thus, the worker’s optimal first-period effort choice \(a_1^*(b_1)\) is given by the first-order condition \(g'(a_1) = b_1\). The first-period competition among prospective employers therefore amounts to choosing \(c_1, b_1,\) and \(c_2\) to maximize the worker’s expected utility,

\[
(A3.1) \quad - E \left\{ \exp \left( - r [c_1 + b_1 (\eta + a_1^*(b_1) + \epsilon_1) - g(a_1)] \right. \\
\left. - r [c_2 + b_2^*(\eta + a_2^*(b_2^*) + \epsilon_2) - g(a_2^*(b_2^*))] \right) \right\},
\]
subject to a zero-expected-profit constraint for the firm,

\[(A3.2) \quad c_1 + \delta c_2 = (1-b_1)[m_0 + a_1^*(b_1)] + \delta(1-b_2^*)[m_0 + a_2^*(b_2)] \, .\]

Arguments parallel to those in the text then show that the optimal first-period slope satisfies

\[(A3.3) \quad b_1 = \frac{1}{1 + r \sum \sigma^2 g''[a_1^*(b_1)]} \frac{r \delta b_2^* \sigma^2 g''[a_1^*(b_1)]}{1 + r \sum \sigma^2 g''[a_1^*(b_1)]}, \]

which is precisely the optimized value of $B_1$ given in the text—the optimal total incentive in the career-concerns model. Thus, the sequence of one-period contracts we identify as optimal in the text provide exactly the same incentives as would the optimal renegotiation-proof long-term contract derived here.

4) The T-Period Model

We first derive the first-order conditions (2.24) and the optimal total incentives and contract slopes (2.26) and (2.27). We then prove that for any $t < T$, $b_t^* < b_T^*$.

Substituting (2.21), (2.22), and (2.23) into the worker's utility function (2.2) yields the worker's utility from a sequence of contracts with slopes $(b_1, \ldots, b_T)$ and associated intercepts given by (2.21), given the market's conjectures $(\hat{a}_1, \ldots, \hat{a}_T)$, the worker's effort choices $(a_1, \ldots, a_T)$, and the output realizations $(y_1, \ldots, y_T)$:

\[(A4.1) \quad U = -\exp\left(-r\left\{\sum_{t=1}^{T} \delta^{t-1} [c_t + b_t y_t - g(a_t)]\right\}\right) \]

\[= -\exp\left(-r\left\{\sum_{t=1}^{T} \delta^{t-1} \left[ (1-b_t)\hat{a}_t + \frac{(1-b_t)\sigma_e^2 m_0}{\sigma_e^2 + (t-1)\sigma_o^2} + b_t y_t - g(a_t) \right]\right\}\right) \]
\[
\frac{(1-b_t)\sigma^2_0}{\sigma^2_c + (t-1)\sigma^2_0} \sum_{t=1}^{T-1} \left( y_t - \hat{\alpha}_t \right) )
\]

\[= -\exp \left( -r \sum_{t=1}^{T} \delta^{t-1} \left[ \frac{(1-b_t)\sigma^2_0}{\sigma^2_c + (t-1)\sigma^2_0} \left( y_t - \hat{\alpha}_t \right) \right] \right) \]

\[= -\exp \left( -r \sum_{t=1}^{T} \delta^{t-1} \left[ \frac{(1-b_t)\sigma^2_0}{\sigma^2_c + (t-1)\sigma^2_0} m_0 - g(a_t) + B_t(y_t - \hat{\alpha}_t) \right] \right) \]

\[= -\exp \left( -r \sum_{t=1}^{T} \delta^{t-1} \left[ \frac{(1-b_t)\sigma^2_0}{\sigma^2_c + (t-1)\sigma^2_0} m_0 - g(a_t) + B_t(a_t - \hat{\alpha}_t) \right] \right) \]

\[\cdot \exp \left( -r \sum_{t=1}^{T} \delta^{t-1} B_t (\eta + \epsilon_t) \right) \}

where \( B_t \) is defined in (2.24).

To apply (2.12) to the last term in (A4.1), note that \( \sum_{t=1}^{T} \delta^{t-1} B_t (\eta + \epsilon_t) \) is Normally distributed with mean \( M \) and variance \( V \), where

\[M = \sum_{t=1}^{T} \delta^{t-1} B_t \] m_0 \quad \text{and} \quad \[V = \left( \sum_{t=1}^{T} \delta^{t-1} B_t \right)^2 \sigma^2_0 + \sum_{t=1}^{T} \left( \delta^{t-1} B_t \right)^2 \sigma^2_c .\]

Maximizing the worker's expected utility with respect to \((a_1, ..., a_T)\) then yields the first-order conditions (2.24). After imposing the equilibrium condition \((\hat{a}_1, ..., \hat{a}_T) = (a_1^*, ..., a_T^*)\), the worker's expected utility from a sequence of contracts with slopes \((b_1, ..., b_T)\) becomes
Robert Gibbons and Kevin J. Murphy

(A4.2) \[ EU = -\exp\left(-r\left\{ \sum_{t=1}^{T} \delta^{t-1} \left[ a_t^* + \frac{(1-b_t)\sigma^2 \mu_0}{\sigma^2 + (1-1)^2 \sigma_0^2} - g(a_t^*) + B_t \mu_0 \right]\right)\) \]

\[ \cdot \exp\left(\frac{1}{2} r^2 \left\{ \left(\sum_{t=1}^{T} \delta^{t-1} B_t \right)^2 \sigma^2_t + \sum_{t=1}^{T} \left(\delta^{t-1} B_t \right)^2 \sigma_0^2 \right\} \right) , \]

which simplifies to (2.25) and then yields the optimal total incentives given in (2.26) and the optimal contract slopes given in (2.27).

We now show that \( b_t^* < b_T^* \) for any \( t < T \). The proof involves the following Lemma.

**Lemma.** For each \( t \in \{1, \ldots, T-1\} \), \( \sum_{t=t+1}^{T} \delta^{t-1} B_t^* \geq 0 \).

**Proof.** By induction. For \( t = T-1 \), (2.26) implies \( B_T^* > 0 \) because

(A4.3) \[ (1 - B_T^*) - r g''(a_t) \Sigma^2 B_T^* = 0 , \]

where (for any \( t \) \( \Sigma^2 = \sigma^2 + \sigma^2_\epsilon \) and \( \sigma^2 = \sigma^2_\eta \sigma^2_\epsilon / (\sigma^2_\epsilon + \sigma^2_\epsilon) \) is the conditional variance of \( \eta \) given the previous \( t \) output observations. For \( t < T-1 \), suppose \( \sum_{t=t+2}^{T} \delta^{t-1} B_t^* \geq 0 \). We must show

(A4.4) \[ \sum_{t=t+1}^{T} \delta^{t-1} B_t^* = \delta B_{t+1}^* + \delta \sum_{t=t+2}^{T} \delta^{t-1} B_t^* \geq 0 , \text{ or } B_{t+1}^* = \sum_{t=t+2}^{T} \delta^{t-1} B_t^* . \]

But the first-order condition for \( B_{t+1}^* \), which is adapted from (2.26), is

(A4.5) \[ B_{t+1}^* = \frac{1 - r \sigma^2 g''(a_{t+1}) \sum_{t=t+2}^{T} \delta^{t-1} B_t^*}{1 + r \Sigma^2 g''(a_{t+1})} , \]

which implies the second inequality in (A4.4). Q.E.D.

We first show that \( b_t^* < b_T^* \). Restate the first-order condition for \( B_t^* \), (A4.5), as...
(A4.6) \[(1 - B_t^*) - \log''(a_t) \left[ B_t^* \Sigma_t^2 + \sigma_t^2 \sum_{\tau=1+1}^T \delta^{\tau-1} B_t^* \right] = 0 \].

Since the objective function (analogous to (2.26) but restated for period t) is concave in \(B_t\), the lefthand side of (A4.6) is negative (positive) when evaluated at \(B_t > (<) B_t^*\). Thus, \(B_t^* < B_T^*\) if and only if

(A4.7) \[(1 - B_T^*) - \log''(a_T) \left[ B_T^* \Sigma_T^2 + \sigma_T^2 \sum_{\tau=1+1}^T \delta^{\tau-1} B_T^* \right] < 0 \],

where \(g''(a_t)\) in (A4.6) becomes \(g''(a_T)\) in (A4.7) because \(g'(a_t^*(B_t)) = B_t\) from (2.24). Substituting (A4.3) into (A4.7) and applying \(\Sigma_T^2 < \Sigma_T^2\) and the Lemma shows that \(B_t^* < B_T^*\).

To show that \(b_t^* < b_T^*\), it now suffices to show that

(A4.8) \[\sum_{\tau=t+1}^T \delta^{\tau-1} (1-b_t^*) \frac{\sigma_0^2}{\sigma_e^2 + (\tau-1)\sigma_0^2} \geq 0\]

in (2.24); that is

(A4.9) \[\delta (1-b_{t+1}^*) \frac{\sigma_0^2}{\sigma_e^2 + \tau\sigma_0^2} + \delta \sum_{\tau=t+2}^T \delta^{\tau-1} (1-b_t^*) \frac{\sigma_0^2}{\sigma_e^2 + (\tau-1)\sigma_0^2} \geq 0\],

which follows if \(b_t^* < 1\) for all \(t\). Since the lefthand side of (A4.6) evaluated at \(B_t = 1\) is negative, \(B_t^* < 1\) for all \(t\). Given this, (A4.9) follows by induction: For \(t = T-1\) we have that \(b_T^* < 1\). For \(t < T-1\), suppose the second term in (A4.9) is positive. Then the fact that \(B_{t+1}^* < 1\) implies that \(b_{t+1}^* < 1\), which completes the proof.
References


Figure 1

Frequency Distributions for CEO Tenure, Firm Tenure, Age, and Years Remaining in Office for CEOs Leaving Office During the 1970-1988 Sample Period

Panel A

<table>
<thead>
<tr>
<th>Mean: 9.8</th>
<th>Median: 7.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years as CEO When Leaving Office</td>
<td></td>
</tr>
</tbody>
</table>

Note: Histogram does not include 23 CEOs who left after more than 34 years as CEO.

Panel B

<table>
<thead>
<tr>
<th>Mean: 28.1</th>
<th>Median: 30.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years in Firm When Leaving Office</td>
<td></td>
</tr>
</tbody>
</table>

Note: Histogram does not include 18 CEOs who left after more than 50 years in their firm.

Panel C

<table>
<thead>
<tr>
<th>Mean: 61.5</th>
<th>Median: 63.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of CEO When Leaving Office</td>
<td></td>
</tr>
</tbody>
</table>

Note: Histogram does not include 8 CEOs who left after age 80.

Panel D

<table>
<thead>
<tr>
<th>Mean: 3.9</th>
<th>Median: 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years Left in Office</td>
<td></td>
</tr>
</tbody>
</table>

Note: Histogram does not include 107 CEO-years for CEOs with more than 14 years remaining. (Histogram is based only on CEOs who actually leave office during the sample period; these CEOs have at most 18 years remaining).

* Histograms and statistics are based on 1,631 CEOs, representing 916 firms and 8,786 CEO-Years, leaving their firms during the 1970-1988 sample period. The empirical results in the paper are based on the subsample that remains after constructing first-differences and eliminating observations with missing performance data. This subsample includes 1,292 CEOs serving in 785 firms, or a total of 6,737 CEO-years.
Table 1
Tenure, Compensation, and Corporate Summary Statistics for “Completed-Spell Sample” of CEOs Leaving Office During the 1970-1988 Sample Period, Grouped by Years Remaining as CEO

<table>
<thead>
<tr>
<th>Variable</th>
<th>All CEOs leaving office during Sample Period</th>
<th>CEOs in their Last Three Years in Office</th>
<th>CEOs not in their Last Three Years in Office</th>
<th>Test Statistic for Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CEO-Years</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6,737</td>
<td>3,117</td>
<td>3,620</td>
<td></td>
</tr>
</tbody>
</table>

**Panel A.**
**CEO Characteristics**
- Mean Age                                      | 59.1                                         | 61.2                                      | 57.3                                        | \( t=29.4^* \)               |
- Mean Tenure in Firm                            | 26.8                                         | 28.4                                      | 25.4                                        | \( t=10.8^* \)               |
- Mean Tenure as CEO                             | 8.9                                          | 9.6                                       | 8.2                                         | \( t=8.6^* \)                |
- Mean Years remaining as CEO until end of final fiscal year\(^c\) | 3.8                                          | 1.0                                       | 6.2                                         | \( t=93.7^* \)               |

**Panel B.**
**Compensation Statistics**
- Mean Salary + Bonus                            | $562,500                                     | $592,800                                  | $536,400                                    | \( t=8.0^* \)                |
- Mean Total Compensation\(^d\)                  | $616,300                                     | $649,000                                  | $588,200                                    | \( t=6.8^* \)                |
- \( \Delta \) (Salary + Bonus)\(^c\) Mean     | $25,500                                      | $28,400                                   | $22,900                                     | \( t=1.8 \)                  |
- \( \Delta \) (Salary + Bonus)\(^c\) Std Dev  | $123,300                                     | $140,400                                  | $106,500                                    | \( F=1.7^* \)                |
- \( \Delta \% \) (Salary + Bonus)\(^c\) Mean  | 6.6\%                                        | 7.0\%                                     | 6.2\%                                       | \( t=1.3 \)                  |
- \( \Delta \% \) (Salary + Bonus)\(^c\) Std Dev| 23.3\%                                       | 26.7\%                                    | 20.0\%                                      | \( F=1.8^* \)                |

**Panel C.**
**Firm Characteristics ($millions)**
- Mean Total Sales                               | $4,239                                       | $4,417                                    | $4,087                                      | \( t=-1.5 \)                 |
- Mean Market Value of Stock                     | $2,352                                       | $2,510                                    | $2,215                                      | \( t=2.1^* \)                |
- \( \Delta \) (Shareholder Wealth)\(^c\) Mean | $53                                          | $19                                       | $83                                         | \( t=1.4 \)                  |
- \( \Delta \) (Shareholder Wealth)\(^c\) Std Dev| $1,892                                       | $2,077                                    | $1,717                                      | \( F=1.5^* \)                |
- Shareholder Return\(^e\) Mean                  | 1.8\%                                        | 1.8\%                                     | 1.8\%                                       | \( t=0.1^* \)                |
- Shareholder Return\(^e\) Std Dev               | 31.5\%                                       | 31.2\%                                    | 31.8\%                                      | \( F=1.0^* \)                |

Mean Fiscal Year                                | 1979.5                                       | 1979.1                                    | 1976.5                                      | \( t=24.4 \)                 |

\(^a\) The sample is constructed from longitudinal data reported in *Forbes* on 1,292 CEOs serving in 785 firms who leave their firms during the 1970-1988 sample period. Monetary variables are in 1988-constant dollars.

\(^b\) * indicates that the difference between columns (2) and (3) is significant at the 5% level. \( F \)-statistics test the equality of subsample variances with 3,116 and 3,619 degrees of freedom.

\(^c\) (Tenure as CEO) and (Years Remaining as CEO) are evaluated in the middle of the fiscal year. Thus, a CEO retiring at fiscal year-end would have .5 years remaining. This measure understates actual years remaining since it ignores time served after the end of the CEO’s final full fiscal year.

\(^d\) Total Compensation includes salary, bonus, value of restricted stock, savings and thrift plans, and other benefits but does not include the value of stock options granted or the gains from exercising stock options.

\(^e\) Cumulative rate of return on common stock, defined as \( \log(1 + \text{annual return}) \).
Figure 2
Estimated Pay-Performance Elasticities, by Years Remaining as CEO

The figure depicts estimated coefficients from equation (3.7) in the text, given below, and is drawn assuming a year effect of .06, which is the average of the 17 estimated year coefficients. The group T-15 includes a small number of CEOs with more than 15 years remaining.

\[
\Delta \ln(w_{it}) = \alpha + \sum_{n=1972}^{1988} \alpha_{n}(n^{th} \text{ year dummy})_{it} \\
+ \left[ \beta_0 + \sum_{t=0}^{18} \beta_t(\text{CEO has } t \text{ years left})_{it} + \sum_{n=1972}^{1988} \beta_n(n^{th} \text{ year dummy})_{it} \right] \Delta \ln(V_{it})
\]

Each of the estimated elasticities for CEOs in their final year, second-to-last year, or third-to-last year (.179, .199, and .184, respectively) is significantly higher than the estimated elasticity for CEOs in their fifth to last year (.120) or sixth-to-last year (.115); no other pairs of coefficients in the figure (including pairs involving the highest estimated elasticity, for CEOs in their eleventh-to-last year) are significantly different at the 5% level.
### TABLE 2

Coefficients of Ordinary Least Squares Regressions of $\Delta \ln(\text{CEO Salary} + \text{Bonus})$ on Shareholder Return, including interactions to allow the Pay-Performance Relation to vary with Year, Years as CEO, and Years Remaining as CEO* (t-statistics in parentheses)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Predicted Sign</th>
<th>Subsample of CEOs who Retire Between 1975-1988</th>
<th>Full Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Few Years Left(^b)</td>
<td>-.0090</td>
<td>-.0056</td>
<td>-.0085</td>
</tr>
<tr>
<td>(Dummy Variable)</td>
<td>(-2.0)</td>
<td>(-1.2)</td>
<td>(-2.0)</td>
</tr>
<tr>
<td>Low Tenure(^c)</td>
<td></td>
<td>0.0396</td>
<td>0.0434</td>
</tr>
<tr>
<td>(Dummy Variable)</td>
<td></td>
<td>(8.4)</td>
<td>(11.2)</td>
</tr>
<tr>
<td>Can't tell if Few Years Left(^d)</td>
<td></td>
<td>-0.0095</td>
<td></td>
</tr>
<tr>
<td>(Dummy Variable)</td>
<td></td>
<td></td>
<td>(-1.4)</td>
</tr>
<tr>
<td>(Few Years Left) (^+) *(Shareholder Return)</td>
<td>0.0436</td>
<td>0.0429</td>
<td>0.0383</td>
</tr>
<tr>
<td></td>
<td>(3.0)</td>
<td>(3.0)</td>
<td>(2.9)</td>
</tr>
<tr>
<td>(Low Tenure) (^-) *(Shareholder Return)</td>
<td></td>
<td>0.0032</td>
<td>0.0199</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>(Can't tell if Few Years Left) (^*) *(Shareholder Return)</td>
<td></td>
<td>-0.0095</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-1.4)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0833</td>
<td>0.0929</td>
<td>0.0927</td>
</tr>
<tr>
<td>Sample Size</td>
<td>6,730</td>
<td>6,730</td>
<td>11,359</td>
</tr>
</tbody>
</table>

* The sample is constructed from longitudinal data reported in *Forbes* on 2,236 CEOs serving in 1,204 firms from 1970-1988. The subsample in column (1) includes 1,292 CEOs serving in 785 firms leaving their firms during the sample period. Compensation is measured in thousands of 1988-constant dollars. All regressions also include intercepts, year dummies, Shareholder Return and Shareholder Return interacted with the year dummies, allowing the pay-performance elasticity to vary by year.

\(^b\) (Few Years Left) is a dummy variable that equals one if the CEO is in his last three years as CEO.

\(^c\) (Low Tenure) is a dummy variable that equals one if the CEO is in his first four years as CEO.

\(^d\) (Can't tell if Few Years Left) is a dummy variable that equals one if the CEO-year observation is in the last three years of the data on a CEO who does not belong to the completed-spells subsample.
| Independent Variable | \text{Dependent Variable:} & \Delta(\text{CEO Salary + Bonus}) & \text{Dependent Variable:} & \Delta\ln(\text{CEO Salary + Bonus}) |
|----------------------|-----------------|-----------------|-----------------|-----------------|
|                      | (1)             | (2)             | (3)             | (4)             |
| Few Years Left\(^b\) | -3.06           | -2.63           | -.0090          | -.0091          |
| (Dummy Variable)     | (-1.0)          | (-0.9)          | (-2.0)          | (-2.0)          |
| \((\Delta\text{Shareholder Wealth})*(Sales)\) | —               | -4.1x10\(^{-7}\) | —               | —               |
|                      |                 | (-4.2)          |                 |                 |
| \((\Delta\text{Shareholder Wealth})*(Sales)^2\) | —               | 3.1x10\(^{-12}\) | —               | —               |
|                      |                 | (3.7)           |                 |                 |
| \((\Delta\text{Shareholder Wealth})*(Few Years Left)\) | .0008           | .0025           | —               | —               |
|                      | (0.4)           | (1.2)           |                 |                 |
| \((\text{Shareholder Return})*(Sales)\) | —               | —               | 2.7x10\(^{-6}\) | (1.5)           |
|                      |                 |                 | (3.0)           |                 |
| \((\text{Shareholder Return})*(Sales)^2\) | —               | —               | -3.2x10\(^{-12}\) | (-0.1)          |
|                      |                 |                 | (3.0)           |                 |
| \((\text{Shareholder Return})*(Few Years Left)\) | —               | —               | .0436           | .0432           |
|                      |                 |                 | (3.0)           | (3.0)           |
| \(R^2\)             | .0426           | .0452           | .0833           | .0845           |
| Sample Size          | 6,727           | 6,727           | 6,730           | 6,730           |

\(^a\) The sample is constructed from longitudinal data reported in Forbes on 1,292 CEOs serving in 785 firms who leave their firms during the 1970-1988 sample period. Compensation is measured in thousands of 1988-constant dollars; \(\Delta(\text{Shareholder Wealth})\) and \(\Delta\ln(\text{CEO Salary + Bonus})\) are measured in millions of 1988-constant dollars. All regressions include intercepts and year dummies. Columns (1) and (2) include \(\Delta\text{Shareholder Wealth}\) and \(\Delta\text{Shareholder Wealth}\) interacted with the year dummies, allowing the pay-performance sensitivity to vary by year; columns (3) and (4) include Shareholder Return and Shareholder Return interacted with the year dummies, allowing the pay-performance elasticity to vary by year.

\(^b\) (Few Years Left) is a dummy variable that equals one if the CEO is in his last three years as CEO.
Appendix Figure A1

**Figure A1**
Estimated pay-performance sensitivities, by year. Figure is based on annual regressions of $\Delta$(Salary + Bonus-$\text{thousands}$) on $\Delta$(Shareholder Wealth-$\text{millions}$) using *Forbes* data on 2,236 CEOs serving in 1,204 firms. Each year's regression is based on approximately 650 observations.

**Figure A2**
Estimated pay-performance elasticities, by year. Figure is based on annual regressions of $\Delta \ln$(Salary + Bonus-$\text{thousands}$) on $\Delta \ln$(Shareholder Wealth-$\text{millions}$) using *Forbes* data on 2,236 CEOs serving in 1,204 firms. Each year's regression is based on approximately 650 observations.