Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence

Robert Gibbons

Cornell University and National Bureau of Economic Research

Kevin J. Murphy

Harvard University

This paper studies optimal incentive contracts when workers have career concerns—concerns about the effects of current performance on future compensation. We show that 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 for workers close to retirement because career concerns are weakest for these workers. We find empirical support for this prediction in the relation between chief executive compensation and stock market performance.

I. Introduction

This paper studies career concerns—concerns about the effects of current performance on future compensation—and describes how

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 thank David Card, Henry Farber, Seonghoon Joon, Eugene Kandel, Alan Krueger, Bruce Meyer, Sherwin Rosen, Michael Weisbach, Jerold Zimmerman, and many seminar audiences for helpful comments. Richard Sloan and Kathleen Jones provided able research assistance. Financial support from the following sources is gratefully acknowledged: the National Science Foundation (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).

[*Journal of Political Economy,* 1992, vol. 100, no. 3]

optimal incentive contracts are affected when these 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; for young workers it can 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 (as in expanding organizations) rather than scarce. 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 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 a 10 percent change in shareholder wealth corresponds to 1.7 percent changes in cash compensation for CEOs less than three years from retirement, but only 1.3 percent pay changes for CEOs more than three years from retirement. Thus for a CEO earning $562,000 (the sample average), a 10 percent change in shareholder wealth corresponds to a $9,500 change in cash compensation for a CEO close to retirement, but only a $7,300 change for a CEO far from retirement. We offer no new insights regarding the average magnitude of the relation between CEO pay and performance (see Jensen and Murphy 1990). Instead, we focus on the cross-sectional variation in, rather than the average magnitude of, the pay-performance relation; our results suggest that pay for performance is stronger in years preceding 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

[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.

Fama (1980), Holmström (1982), and MacLeod and Malcomson (1988)—incorporate incentives, learning, and market forces but not risk aversion or contracts.[2] Other dynamic agency models—such as Lambert (1983), Rogerson (1985), Murphy (1986), Holmström and Milgrom (1987), and Fudenberg, Holmström, and Milgrom (1990)—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.[3]

II. 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 (η), the worker's nonnegative effort (a_t), and noise (ϵ_t):

$$y_t = \eta + a_t + \epsilon_t. \tag{1}$$

Before production begins, there is symmetric (but imperfect) information about the worker's ability: the worker and all prospective employers believe that η is normally distributed with mean m_0 and variance σ_0^2. The error terms are normally distributed with mean zero and variance σ_ϵ^2 and are independent of each other and of ability, η.

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

$$U(w_1, \ldots, w_T; a_1, \ldots, a_T) = -\exp\left(-r\left\{\sum_{t=1}^{T} \delta^{t-1}[w_t - g(a_t)]\right\}\right), \tag{2}$$

where w_t is the wage paid in period t and $g(a_t)$ measures disutility of effort. We assume that $g(a_t)$ is convex and satisfies $g'(0) = 0$, $g'(\infty)$

[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 (1992) 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.

$= \infty$, 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) is not the additively separable exponential utility function often encountered in the literature: (2) implies that the worker is indifferent among all deterministic wage streams with constant present value (computed using δ as the discount factor), just as though the worker had access to a perfect capital market.

To keep our theoretical analysis simple, we make two assumptions about contracting possibilities: (1) short-term (i.e., one-period) contracts are linear in output, and (2) long-term (i.e., multiperiod) 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, η) can be interpreted so as to ensure that the optimal contract is linear.

Our assumption that long-term contracts are not feasible can be reinterpreted. We show (in the Appendix, sec. *A*) that this assumption is equivalent to the assumption 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 applied in many settings: see, e.g., Dewatripont [1989] on contracting under adverse selection, Fudenberg and Tirole [1990] on contracting under moral hazard, and Hart and Moore [1988] on incomplete contracting under symmetric information.) In our model, renegotiation between the worker and a single employer mimics one of the effects of competition among the current and prospective employers.

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

The Two-Period Model

The timing of the two-period model is explained as follows. At the beginning of the first period, prospective employers simultaneously offer a worker single-period linear wage contracts of the form $w_1(y_1) = c_1 + b_1 y_1$. (Recall that although information is imperfect it is also symmetric, 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 $w_2(y_2) = c_2 + b_2 y_2$. These second-period contract offers depend implicitly on y_1 because first-period output conveys information about the worker's ability, as described below. The force of our assumption 2, however, is that second-period contracts 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:

$$-E\{\exp\{-r[c_1 + b_1(\eta + a_1 + \epsilon_1) - g(a_1)] - r\delta[c_2 + b_2(\eta + a_2 + \epsilon_2) - g(a_2)]\}\}. \tag{3}$$

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

$$-\exp\{-r\delta^{-1}[c_1 + b_1 y_1 - g(a_1)]\} \cdot E\{\exp\{-r[c_2 + b_2(\eta + a_2 + \epsilon_2) - g(a_2)]\}|y_1\}, \tag{4}$$

so the worker's second-period effort choice problem reduces to

$$\max_{a_2} -E\{\exp\{-r[c_2 + b_2(\eta + a_2 + \epsilon_2) - g(a_2)]\}|y_1\}. \tag{5}$$

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

$$g'(a_2) = b_2. \tag{6}$$

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

$$c_2(b_2) = (1 - b_2)E\{y_2|y_1\}. \tag{7}$$

Given equation (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:

$$E\{y_2|y_1\} = E\{\eta|y_1\} + a_2^*(b_2). \tag{8}$$

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.) When one applies well-known formulas from DeGroot (1970), the conditional distribution of η given the observed first-period output y_1 is then normal with mean

$$m_1(y_1, \hat{a}_1) = \frac{\sigma_\epsilon^2 m_0 + \sigma_0^2(y_1 - \hat{a}_1)}{\sigma_\epsilon^2 + \sigma_0^2} \tag{9}$$

and variance

$$\sigma_1^2 = \frac{\sigma_0^2 \sigma_\epsilon^2}{\sigma_0^2 + \sigma_\epsilon^2}. \tag{10}$$

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 $\Sigma_2^2 \equiv \sigma_1^2 + \sigma_\epsilon^2$, the conditional variance of $\eta + \epsilon_2$ given the observed first-period output y_1.

Substituting the appropriate expressions into (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

$$\begin{aligned} &-E\{\exp(-r\{c_2(b_2) + b_2[\eta + a_2^*(b_2) + \epsilon_2] - g(a_2^*(b_2))\})|y_1\} \\ &= -\exp\{-r[m_1(y_1, \hat{a}_1) + a_2^*(b_2) - g(a_2^*(b_2)) - \tfrac{1}{2} r b_2^2 \Sigma_2^2]\}, \end{aligned} \tag{11}$$

where the right-hand side of (11) is derived using the observation that if x is normally distributed with mean μ and variance σ^2 then

$$E\{\exp(-kx)\} = \exp(-k\mu + \tfrac{1}{2} k^2 \sigma^2). \tag{12}$$

Optimizing (11) and implicitly differentiating (6) yield the first-order condition for b_2^*, the optimal second-period slope:[4]

$$b_2 = \frac{1}{1 + r\Sigma_2^2 g''[a_2^*(b_2)]}. \tag{13}$$

Given the assumption that $g''' \geq 0$, it is straightforward to show both that (11) is strictly quasi-concave, so (13) is sufficient, and that b_2^* decreases with both risk aversion (r) and uncertainty (Σ_2^2).

[4] Note that (13) is identical to the result given in Lazear and Rosen (1981, p. 852, n. 8), who derive it using a second-order Taylor approximation to the more general (static) utility function $U(w - g(a))$. Thus (13) is precisely optimal given (2) and approximately optimal more generally.

A very convenient feature of (13) is that b_2^* is independent of y_1. This fact 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 (7). The reasons that b_2^* is independent of y_1 are that the absence of wealth effects in the utility function (2) implies that optimal incentives depend on the variance but not on the mean of output, and that the variance of beliefs about ability evolves deterministically in the normal learning model, as in (10).

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

$$\begin{aligned} -E\{\exp(-r[c_1 + b_1(\eta + a_1 + \epsilon_1) - g(a_1)] \\ - r\delta\{c_2(b_2^*) + b_2^*[\eta + a_2^*(b_2^*) + \epsilon_2] - g(a_2^*(b_2^*))\})\}. \end{aligned} \tag{14}$$

Substituting (8) and (9) into (7) yields

$$c_2(b_2^*) = (1 - b_2^*)\left[\frac{\sigma_\epsilon^2 m_0 + \sigma_0^2(y_1 - \hat{a}_1)}{\sigma_\epsilon^2 + \sigma_0^2} + a_2^*(b_2^*)\right]. \tag{15}$$

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

$$g'(a_1) = b_1 + \delta\frac{\partial c_2}{\partial a_1} = b_1 + \delta(1 - b_2^*)\frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} \equiv B_1. \tag{16}$$

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, $\delta(1 - b_2^*)\sigma_0^2/(\sigma_0^2 + \sigma_\epsilon^2)$. The career concerns incentive is positive (since [13] implies $0 < b_2^* < 1$), increases with the uncertainty about ability, σ_0^2, and decreases with the uncertainty about production, σ_ϵ^2.

So far we have taken as given the market's second-period conjecture about first-period effort, $\hat{a}_1$. Thus (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 (16) does not depend on $\hat{a}_1$. Therefore, the equilibrium conjecture is

$$\hat{a}_1 = a_1^*(b_1). \tag{17}$$

Competition ensures that firms earn zero expected profits. Because of assumption 2, expected profits must be zero in each period. Hence,

$$c_1(b_1) = (1 - b_1)E(y_1) = (1 - b_1)[m_0 + a_1^*(b_1)]. \tag{18}$$

Substituting $a_1^*(b_1)$ and $c_1(b_1)$ into (14) and applying (12) yield the worker's expected utility (from the perspective of the first period) for an arbitrary b_1:

$$-\exp\{-r[m_0 + a_1^*(b_1) - g(a_1^*(b_1))] - r\delta[m_0 + a_2^*(b_2^*) - g(a_2^*(b_2^*))]\}$$
$$\cdot \exp\{{}^1\!/\!{}_2 r^2[(B_1 + \delta b_2^*)^2 \Sigma_1^2 - 2B_1 \delta b_2^* \sigma_\epsilon^2]\}, \tag{19}$$

where B_1 reflects the sum of explicit and implicit incentives, as defined in (16), and $\Sigma_1^2 \equiv \sigma_0^2 + \sigma_\epsilon^2$ is the variance of $\eta + \epsilon_1$. The optimal first-period slope, b_1^*, thus satisfies the first-order condition

$$b_1 = \frac{1}{1 + r\Sigma_1^2 g''[a_1^*(b_1)]} - \delta(1 - b_2^*)\frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} - \frac{r\delta b_2^* \sigma_0^2 g''[a_1^*(b_1)]}{1 + r\Sigma_1^2 g''[a_1^*(b_1)]}. \tag{20}$$

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 (20) to the single expression in (13), three effects contribute to the result. The first term in (20) reflects a noise reduction effect: learning about the worker's ability causes the conditional variance of output to decline over time ($\Sigma_2^2 < \Sigma_1^2$), so the optimal trade-off between insurance and incentives shifts toward the latter over time—$1/\{1 + r\Sigma_1^2 g''[a_2^*(b_2^*)]\} < b_2^*$. The second term in (20) is the career concerns effect, familiar from (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 (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.[5] The career concerns effect in (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

[5] Suppose, e.g., that there is no noise in production (i.e., $\sigma_\epsilon^2 = 0$). Then first-period output perfectly reveals the worker's ability, so $\sigma_1^2 = 0$. Since there is no uncertainty in the second period (i.e., $\Sigma_2^2 = \sigma_1^2 + \sigma_\epsilon^2 = 0$), 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 (15) is zero. If $\delta = 1$ and $g(a) = a^2/2$, then the total first-period incentive is $B_1 = b_1 = (1 - r\sigma_0^2)/(1 + r\sigma_0^2)$, which clearly can be negative.

the marginal disutility of effort in the first-order condition (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, which may be realistic.

The T-*Period Model*

In the T-period case, the three equations that summarize the workings of the competitive labor market can be written as follows. First, the zero-expected-profit constraint in period t is

$$c_t = (1 - b_t)E\{y_t | y_1, \ldots, y_{t-1}\}. \tag{21}$$

Second, the conditional expectation of output in period t is

$$E\{y_t | y_1, \ldots, y_{t-1}\} = m_{t-1} + \hat{a}_t, \tag{22}$$

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 $(y_1, \ldots, y_{t-1})$ and its conjectures about prior effort levels $(\hat{a}_1, \ldots, \hat{a}_{t-1})$, is

$$m_{t-1}(y_1, \ldots, y_{t-1}; \hat{a}_1, \ldots, \hat{a}_{t-1}) = \frac{\sigma_\epsilon^2 m_0 + \sigma_0^2 \sum_{\tau=1}^{t-1} (y_\tau - \hat{a}_\tau)}{\sigma_\epsilon^2 + (t-1)\sigma_0^2}. \tag{23}$$

Substituting these three equations into the worker's utility function (2) and applying (12) yield the worker's expected utility from a sequence of contracts with arbitrary slopes $(b_1, \ldots, b_T)$ and associated intercepts given by (21), given the market's effort conjectures $(\hat{a}_1, \ldots, \hat{a}_T)$ and the worker's effort choices $(a_1, \ldots, a_T)$. (See the Appendix, sec. B, 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 (16):

$$\begin{aligned} g'(a_t^*) &= b_t + \sum_{\tau=t+1}^{T} \delta^{\tau-t} \frac{\partial c_\tau}{\partial a_t} \\ &= b_t + \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - b_\tau) \frac{\sigma_0^2}{\sigma_\epsilon^2 + (\tau - 1)\sigma_0^2} \equiv B_t. \end{aligned} \tag{24}$$

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

In equilibrium, the market's conjectures are correct: $(\hat{a}_1, \ldots, \hat{a}_T) = (a_1^*, \ldots, a_T^*)$. The worker's expected utility from the sequence of contracts with slopes $(b_1, \ldots, b_T)$ can then be written as

$$-\exp\left(-r\left\{\sum_{t=1}^{T} \delta^{t-1}[m_0 + a_t^* - g(a_t^*)]\right\}\right.$$
$$\left. + \tfrac{1}{2}r^2\left[\left(\sum_{t=1}^{T} \delta^{t-1}B_t\right)^2 \sigma_0^2 + \sum_{t=1}^{T} (\delta^{t-1}B_t)^2\sigma_\epsilon^2\right]\right). \tag{25}$$

For now, take $(b_2, \ldots, b_T)$ as given. Recall that changes in b_1 affect only a_1^*. Since $g'(a_1^*) = B_1$ from (24), it is convenient to optimize (25) with respect to the first-period total incentive, $B_1^*(b_2, \ldots, b_T)$, which solves the first-order condition

$$B_1 = \frac{1}{1 + r\Sigma_1^2 g''(a_1^*)} - \frac{r\sigma_0^2 g''(a_1^*) \sum_{t=2}^{T} \delta^{t-1}B_t}{1 + r\Sigma_1^2 g''(a_1^*)}. \tag{26}$$

Applying (24) then yields the optimal first-period contract slope $b_1^*(b_2, \ldots, b_T)$,

$$b_1 = \frac{1}{1 + r\Sigma_1^2 g''(a_1^*)} - \sum_{t=2}^{T} \delta^{t-1}(1 - b_t)\frac{\sigma_0^2}{\sigma_\epsilon^2 + (t-1)\sigma_0^2}$$
$$- \frac{r\sigma_0^2 g''(a_1^*) \sum_{t=2}^{T} \delta^{t-1}B_t}{1 + r\Sigma_1^2 g''(a_1^*)}. \tag{27}$$

Finally, using (27) recursively yields the optimal slopes $(b_1^*, \ldots, b_T^*)$, beginning with b_T^* and working backward to $b_1^*(b_2^*, \ldots, b_T^*)$.

We conclude our theoretical analysis with a partial characterization of the optimal slopes $\{b_t^*: t = 1, \ldots, T\}$: $b_t^* < b_{t+1}^*$ for all $t < T$. That is, contractual incentives increase monotonically with t and so are strongest for those about to retire. (Again, see the Appendix, sec. *B*, for the derivation.)

III. 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 data sets 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 take actions that increase shareholder wealth (i.e., supply effort, in our simple model). 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.[6]

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 before being appointed CEO, so shareholders (as well as potential alternative employers) may have precise information about the ability of a new CEO.[7] In addition, the expected level of a worker's compensation in our model is determined by a competitive market of prospective employers who continuously revise their bids for the worker's services as new information is revealed regarding the worker's ability. The fact that CEOs rarely leave one firm 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 for CEOs.[8]

[6] 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. We ignore debt by focusing 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.

[7] 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.

[8] 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.

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 briefly discuss below) a change in the firm's environment may renew the shareholder's uncertainty about the CEO's ability, even if his ability in the previous environment had become known.

Second, while 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 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 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.[9]

A Longitudinal Sample of Chief Executive Officers

We test the implications of our career concerns model using data constructed by following all CEOs 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

[9] 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 here. 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 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.

years 1970–88, or a total of 15,148 CEO-years of data. In order to distinguish CEOs close to retirement from those with longer horizons, we initially restrict our analysis to CEOs who left office during the 1970–88 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 on the basis of 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 who left office during the 1970–88 sample period. Panels *A* and *B* show that, on leaving office, the average CEO has been in his position for over 10 years and has been in his firm for almost 30 years. Panel *C* shows that most executives leave their position near normal retirement age: 60 percent were between 60 and 66 when they left, and 31 percent were 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, semiretirement capacity. Vancil (1987) estimates that 75 percent 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 percent die while in office; others resign because of ill health. In our own sample, we find that only 36 of the 1,631 (2.2 percent) 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 postdeparture 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 percent of these CEOs joined another corporation after leaving their firm; another 1.4 percent joined law firms or universities. We were unable to find postdeparture *Who's Who* information on 40 percent 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 who left office during the sample period. Panel *D* shows that about half the CEO-year observations

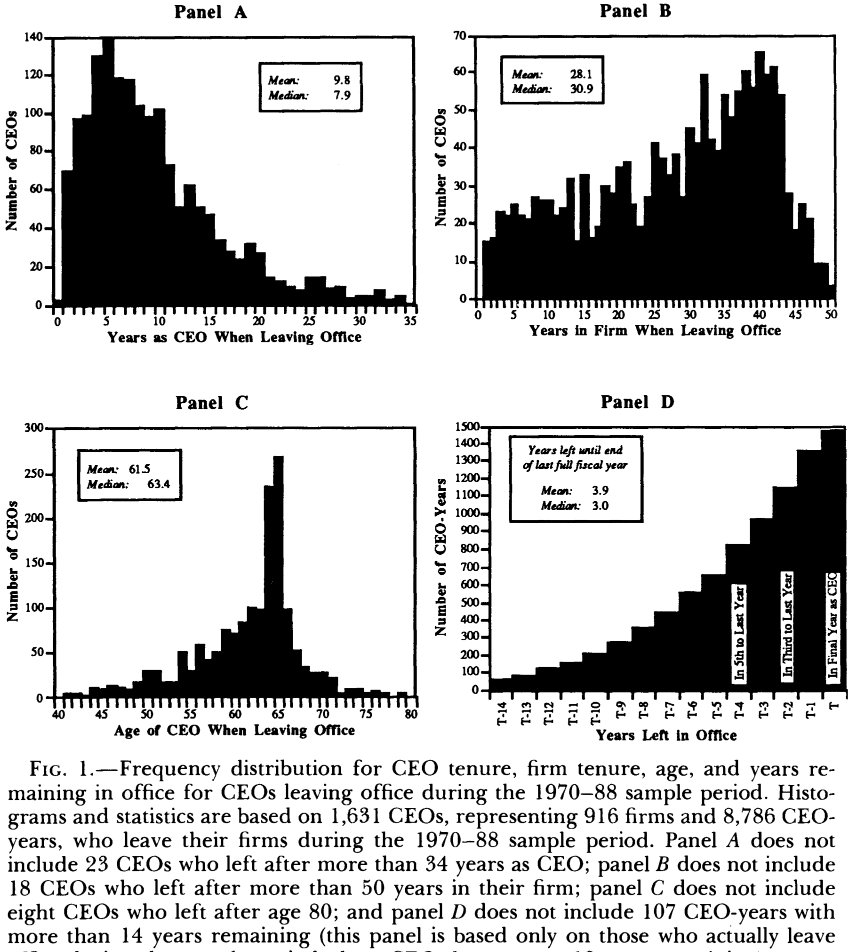

FIG. 1.—Frequency distribution for CEO tenure, firm tenure, age, and years remaining in office for CEOs leaving office during the 1970–88 sample period. Histograms and statistics are based on 1,631 CEOs, representing 916 firms and 8,786 CEO-years, who leave their firms during the 1970–88 sample period. Panel *A* does not include 23 CEOs who left after more than 34 years as CEO; panel *B* does not include 18 CEOs who left after more than 50 years in their firm; panel *C* does not include eight CEOs who left after age 80; and panel *D* does not include 107 CEO-years with more than 14 years remaining (this panel is based only on those who actually leave office during the sample period; these CEOs have *at most* 18 years remaining).

come 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: a CEO-year observation for a CEO who left office during the sample period can be *at most* 18 years before the CEO left office (the number of years covered by our longitudinal data set), but panel *A* reveals that some CEOs hold office more than 18 years.

Figure 1 is based on all 1,631 CEOs who left office during the 1970–88 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-year observation in the subsample represents a CEO who has been in his position for almost nine years and will continue in office for almost four additional years. He receives annual compensation (excluding both grants of and gains from exercising stock options) of \$616,300, of which 90 percent 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 percent per year over the sample period.

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 percent over the sample period (in firms in the completed-spells subsample). We define the change in shareholder wealth, ΔV_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): $\Delta V_t \equiv V_{t-1} r_t$. The sample average change in shareholder wealth is a \$53 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, consistent with 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 overrepresented in the early years of the sample. This overrepresentation is a direct result of our methodology: the completed-spells subsample includes only CEOs who left office

TABLE 1

Tenure, Compensation, and Corporate Summary Statistics for "Completed-Spells Sample" of CEOs Leaving Office during the 1970–88 Sample Period, Grouped by Years Remaining as CEO

Variable	All CEOs Leaving Office during Sample Period (1)	CEOs in Their Last Three Years in Office (2)	CEOs Not in Their Last Three Years in Office (3)	Test Statistic for Difference* (4)
Number of CEO-years	6,737	3,117	3,620	
	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†	3.8	1.0	6.2	t = 93.7*
	B. Compensation Statistics			
Mean salary + bonus	$562,500	$592,800	$536,400	t = 8.0*
Mean total compensation‡	$616,300	$649,000	$588,200	t = 6.8*
Δ(salary + bonus): Mean	$25,500	$28,400	$22,900	t = 1.8
Δ(salary + bonus): Standard deviation	$123,300	$140,400	$106,500	F = 1.7*
Δ%(salary + bonus): Mean	6.6%	7.0%	6.2%	t = 1.3
Δ%(salary + bonus): Standard deviation	23.4%	26.7%	20.0%	F = 1.8*

	C. Firm Characteristics (Millions of Dollars)			
Mean total sales	\$4,239	\$4,417	\$4,087	t = 1.5
Mean market value of stock	\$2,352	\$2,511	\$2,216	t = 2.1*
Δ(shareholder wealth): Mean	−\$53	−\$19	−\$83	t = 1.4
Standard deviation	\$1,892	\$2,077	\$1,717	F = 1.5*
Shareholder return:[§] Mean	1.8%	1.8%	1.8%	t = .1
Standard deviation	31.5%	31.2%	31.8%	F = 1.0*
Mean fiscal year	1977.7	1979.1	1976.5	t = 24.4

NOTE.—The sample is constructed from longitudinal data reported in *Forbes* on 1,292 CEOs serving in 785 firms who leave their firms during the 1970–88 sample period. Monetary variables are in 1988-constant dollars.

* The difference between cols. 2 and 3 is significant at the 5 percent level. F-statistics test the equality of subsample variances with 3,116 and 3,619 degrees of freedom.

† 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 one-half a year remaining. This measure understates actual years remaining since it ignores time served after the end of the CEO's final full fiscal year.

‡ 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.

§ Cumulative rate of return on common stock, defined as log(1 + annual return).

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 $w_t(y_t) = c_t + b_t y_t$ simply becomes $w_t = c_t + b_t \Delta V_t$. 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: $V_{t-1} = E\{y_t | y_1, \ldots, y_{t-1}\}$. Therefore, the change in shareholder wealth in period t is $\Delta V_t = y_t - E\{y_t | y_1, \ldots, y_{t-1}\}$, and the linear compensation contract $w_t = c_t + b_t y_t$ becomes

$$w_t = c_t + b_t(\Delta V_t + E\{y_t | y_1, \ldots, y_{t-1}\}). \tag{28}$$

Both approaches share the feature that the coefficient on ΔV_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 (21) and (22) into (28) yields

$$w_t = m_{t-1} + \hat{a}_t + b_t \Delta V_t, \tag{29}$$

but (29) 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 (23). Because we interpret y_t as end-of-period firm value, however, the first difference of (29) is a more tractable empirical specification: Substituting (22) into the definition of change in shareholder wealth, $\Delta V_t = y_t - E\{y_t | y_1, \ldots, y_{t-1}\}$, 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 (23) then yields

$$m_{t-1} = m_{t-2} + \frac{\sigma_0^2}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} \Delta V_{t-1}. \tag{30}$$

Thus the first difference of (29) 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:

$$\Delta w_t = \Delta \hat{a}_t + b_t \Delta V_t + \left[\frac{\sigma_0^2}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} - b_{t-1}\right]\Delta V_{t-1}, \tag{31}$$

which suggests the empirical specification

$$\Delta w_t = \alpha_t + \beta_t \Delta V_t + \gamma_t \Delta V_{t-1}, \tag{32}$$

where $\alpha_t \equiv \Delta \hat{a}_t$, $\beta_t \equiv b_t$, and $\gamma_t \equiv \{\sigma_0^2/[\sigma_\epsilon^2 + (t-1)\sigma_0^2]\} - b_{t-1}$.

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 (32), we therefore distinguish between the number of years remaining in a CEO's career and the number of years already spent in office (tenure).

HYPOTHESIS 1. With tenure as CEO held constant, the slope of the compensation contract, β_t, increases as the CEO nears retirement.

Career concerns provide weaker incentives as the CEO nears retirement, so the incentives provided by current compensation become stronger.

HYPOTHESIS 2. With years remaining as CEO held constant, the slope of the compensation contract, β_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. Since the CEO's true managerial ability becomes estimated more precisely as tenure increases, the variance of y_t decreases and the pay-performance relation increases as the CEO's tenure increases.

HYPOTHESIS 3. With tenure as CEO held constant, the coefficient on the change in shareholder wealth in period $t - 1$, γ_t, decreases as the CEO nears retirement.

The coefficient on the change in shareholder wealth in period $t - 1$ has two components, $\sigma_0^2/[\sigma_\epsilon^2 + (t-1)\sigma_0^2]$ and $-b_{t-1}$. The first term reflects the effect of ΔV_{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 weaker incentives.

HYPOTHESIS 4. With years remaining as CEO held constant, the coefficient on the change in shareholder wealth in period $t - 1$, γ_t, decreases with tenure as CEO.

Both $\sigma_0^2/[\sigma_\epsilon^2 + (t - 1)\sigma_0^2]$ 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 1 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:

$$\Delta w_{it} = \alpha + \sum_{\tau=0}^{18} \beta_\tau (\text{CEO has } \tau \text{ years left})_{it} \times \Delta V_{it}, \qquad (33)$$

where (CEO has τ years left)$_{it}$ is a dummy variable equal to one if the ith CEO has τ years remaining in his career in fiscal year t. Under hypothesis 1, the slope of the pay-performance relation increases as the CEO nears retirement: $\beta_0 > \beta_1 > \ldots > \beta_{18}$. There are at least four potential problems with estimating (33) directly, however. First, our empirical definition of shareholder wealth change, ΔV_{it}, is net of payments to the CEO; our theoretical measure is gross of 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 Δw_{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 stockholdings 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 (33), as noted in connection with table 1, is that CEOs near retirement are overrepresented in the later years of the completed-spells subsample, so estimates of the β_τ's based on this subsample will reflect secular trends in incentive compensation in addition to the effects of career concerns. As shown in the figures in section C of the Appendix, the relation between pay and performance has increased over the past decade, so the estimated β_τ's for CEOs near retirement are likely to be biased upward. Fourth, and perhaps most important, (33) 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: (33) 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 (33) that allows the pay-performance elasticity to vary across years:

$$\begin{aligned}\Delta \ln(w_{it}) = \alpha &+ \sum_{n=1972}^{1988} \alpha_n(n\text{th year dummy})_{it} \\ &+ \left[\sum_{\tau=0}^{18} \beta_\tau(\text{CEO has } \tau \text{ years left})_{it} \right. \\ &+ \left. \sum_{n=1972}^{1988} \beta_n(n\text{th year dummy})_{it}\right] \times \Delta \ln(V_{it}),\end{aligned} \tag{34}$$

where $\Delta \ln(V_{it}) \approx \ln(1 + r_{it})$ is the continuously accrued rate of return on common stock and $(n\text{th year dummy})_{it}$ 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).[10]

Figure 2 depicts the estimated pay-performance elasticities from (34) for CEOs grouped on the basis of their years remaining as CEO. (To simplify the figure, observations with 15 or more years remaining have been pooled.) The figure is drawn under the assumption of a year effect of .06, which is the average of the 17 estimated year coefficients. Those CEOs in their final year, second-to-last year, and

[10] We also used a more comprehensive measure of compensation that includes fringe benefits and contingent (but nonstock) remuneration. The qualitative results from these regressions are similar to those presented in the tables, but the significance levels are generally much lower, reflecting the noisiness of the data on these additional measures of compensation.

third-to-last year have estimated pay-performance elasticities of .178, .203, and .183, respectively. Each of these estimated elasticities is significantly higher than the estimated elasticities for CEOs in their fifth-to-last year (.119) and sixth-to-last year (.116); 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 percent 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 versus CEOs with more than three years remaining, we divided the completed-spells subsample into two groups and estimated the following regression:

$$\begin{aligned} \Delta \ln(w_{it}) = {} & \alpha_0 + \alpha_1(\text{few years left})_{it} \\ & + \sum_{n=1972}^{1988} \alpha_n(n\text{th year dummy})_{it} \\ & + \left[\beta_0 + \beta_1(\text{few years left})_{it} \right. \\ & \left. + \sum_{n=1972}^{1988} \beta_n(n\text{th year dummy})_{it}\right] \times \Delta \ln(V_{it}), \end{aligned} \tag{35}$$

where $(\text{few years left})_{it}$ is a dummy variable equal to one if the ith CEO is in the last three years of his career in fiscal year t.

Column 1 of table 2 reports estimates of regression (35) for the completed-spells subsample. Those 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 (with a median of .13), indicating that CEOs with many years remaining receive between 0.2 percent and 2.0 percent raises for every 10 percent return realized by shareholders. The few years left × shareholder return coefficient of .0436 implies that CEOs with few years remaining receive an additional 0.44 percent raise for every 10 percent return realized by shareholders. Thus, on the basis of the median year interaction, each 10 percent change in shareholder wealth corresponds to 1.3 percent changes in cash compensation for CEOs more than three years from retirement, compared to 1.7 percent pay changes for CEOs in their final three years. The return × few years left coefficient is highly significant (t = 3.0), which we interpret as empirical support for hypothesis 1.

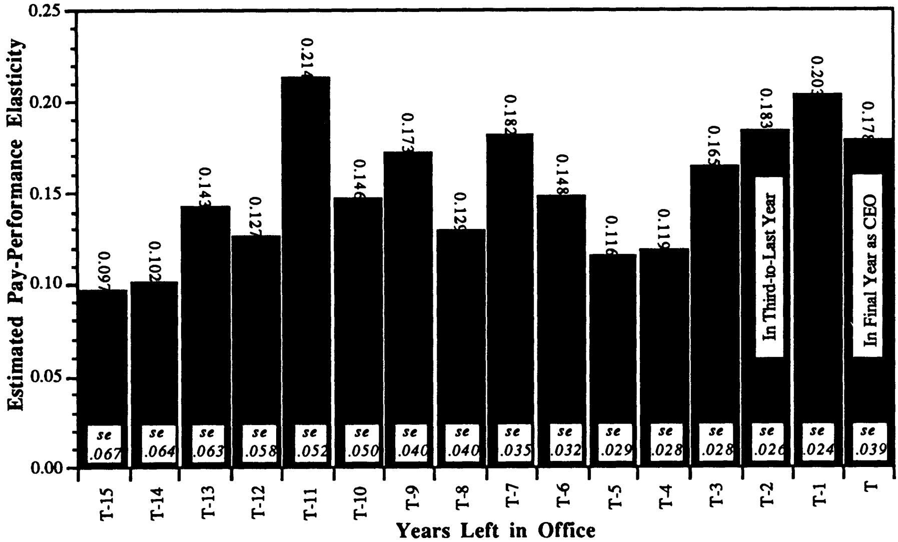

FIG. 2.—Estimated pay-performance elasticities, by years remaining as CEO. The figure depicts estimated coefficients from eq. (34) and is drawn under the assumption of a year effect of .06, which is the average of the 17 estimated year coefficients. Standard errors for each estimated coefficient are in italics. The group $T - 15$ includes a small number of CEOs with more than 15 years remaining.

Our definition of few years left—CEOs in their last three years—is arbitrary. We reestimated 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 few years left × shareholder return 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 that this insignificance reflects both small sample size and non-performance-related payments made to CEOs in their last year. We also reestimated the regression in column 1 of table 2 after eliminating CEOs in their final year. The estimated few years left × shareholder return interaction coefficient in our reestimated regression is positive but only marginally significant.

The results in table 2 support hypothesis 1, but also support an alternative hypothesis: CEOs whose total stay in office is only a few years may have higher pay-performance elasticities than 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 on retirement. None of the qualitative results in table 2 is changed, suggesting that our results are not driven by cross-

TABLE 2

Coefficients of Ordinary Least Squares Regressions of Δ ln(CEO Salary + Bonus) on Shareholder Return, Including Interactions to Allow the Pay-Performance Relation to Vary with Year, Years as CEO, and Years Remaining as CEO

Independent Variable	Predicted Sign	Dependent Variable: Δ ln(CEO Salary + Bonus): Subsample of CEOs Who Retire 1975–88 (1)	(2)	Full Sample (3)
Few years left* (dummy variable)		−.0090	−.0056	−.0085
		(−2.0)	(−1.2)	(−2.0)
Low tenure† (dummy variable)		...	.0396	.0435
			(8.4)	(11.2)
Can't tell if few years left‡ (dummy variable)		...	...	−.0095
				(−1.4)
Few years left × shareholder return	+	.0436	.0429	.0383
		(3.0)	(3.0)	(2.9)
Low tenure × shareholder return	−	...	.0032	.0199
			(.2)	(1.6)
Can't tell if few years left × shareholder return		...	...	.0335
				(1.8)
R^2		.0833	.0929	.0927
Sample size		6,730	6,730	11,360

Note.—The sample is constructed from longitudinal data reported in *Forbes* on 2,236 CEOs serving in 1,204 firms during 1970–88. The subsample in col. 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. *t*-statistics are in parentheses.

* Few years left is a dummy variable that equals one if the CEO is in his last three years as CEO.

† Low tenure is a dummy variable that equals one if the CEO is in his first four years as CEO.

‡ 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.

sectional (negative) correlation between total time in office and the pay-performance elasticity.

Performance Pay and CEO Tenure

Hypothesis 2 predicts that, with years remaining as CEO held 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: compared to column 1, 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 × shareholder 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 η but rather varies over time as $\eta_{t+1} = \eta_t + \nu_t$, where the innovations ν_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 his tenure. Thus hypothesis 2 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 left office during the 1970–88 sample period represents only about half of the 1,900 CEOs (11,359 CEO-years) in the full sample (after first differences are constructed and observations with missing data are eliminated). The CEOs in the full sample but not in the completed-spells subsample include those still serving at the end of the 1988 fiscal year and CEOs of firms that were deleted from the *Forbes* survey.[11] 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 reestimation of column 2 for the *full sample* of CEOs, including those who did not leave their firms during the sample period. The regression contains two additional explanatory variables: "can't tell if few years left," 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, and an interaction of this dummy variable with shareholder return. The results are similar to those re-

[11] A total of 623 firms were deleted from the *Forbes* surveys during the 1970–88 sample period. Of these, 277 were still "going concerns" as of 1989, 290 were acquired by or merged with another firm (127 of these were acquired or merged within 2 years of the *Forbes* delisting), and 56 liquidated, went bankrupt, or went private.

ported 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 3 and 4 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 reestimated the regressions in table 2 after including lagged shareholder return and interactions of lagged return with few years left, low tenure, and the 17 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 3 and 4 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 assumed that the CEO's final year is known with certainty, but some executives may retire earlier or later than they were originally 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 actual years remaining (as measured after retirement has been observed).

As shown in panel *C* of figure 1, many CEOs leave office when they reach age 64 or 65. One estimate of the expected years remaining is therefore the number of years remaining until the CEO reaches 65. We reestimated the regression in column 1 of table 2 after replacing "few years left" with "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 few expected years left × shareholder return was positive but statistically insignificant. Thus hypothesis 1 is not supported when "age 62 or older" is used as a proxy for few expected years remaining.

The board of directors' assessment of a CEO's horizon is based in part on information not available to us, including the CEO's health, potential replacements, and implicit or explicit retirement policies. One interpretation of the insignificance of the few expected years left × shareholder return interaction term is that, because of the information not available to us, actual years remaining (i.e., few years left in table 2) is a better proxy for expected years remaining than the CEO's age is. Furthermore, whatever measure we use to assign CEO-year observations to the few years left category, if our theoretical model is correct, then errors in this assignment process will bias our empirical results against the prediction of the model, as follows. In panel *C* of figure 1, 40 percent of the CEOs left office before the age of 60 or after the age of 66. Thus an older CEO may still have a long remaining career (and therefore would have small explicit incentives, according to our model), so including such a CEO in the few years left category biases downward the coefficient on the few years left × shareholder return interaction term. Similarly, excluding a younger CEO from the few years left category when he is in fact near retirement (and therefore has 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 (33) 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. We now motivate our use of the logarithmic specification to control for size-related heterogeneity, and also describe our attempts to control for other potential (but unknown) forms of heterogeneity.

Column 1 of table 3 reports estimates of the nonlogarithmic version of regression (35) for the 1,292 CEOs in the completed-spells subsample. The coefficient on the interaction Δshareholder wealth × 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, $\Delta V_t \times$ sales and $\Delta V_t \times \text{sales}^2$, to allow the pay-performance relation to vary quadratically with firm size as measured by net sales (in millions of 1988-constant dollars). On the basis of the average of the unreported Δshareholder wealth × year effects, the coefficients in column 2 suggest that CEOs in median-size firms (sales of \$2.0 billion) receive average pay changes of 0.93¢ for every \$1,000 change in shareholder wealth. Those CEOs at the tenth percentile (sales of \$400 million) and ninetieth percentile (sales of \$8.2 billion) receive average pay changes of 0.99¢ and 0.70¢, respectively, for every \$1,000 change in shareholder wealth. Including the size interaction terms in column 2

TABLE 3

Coefficients of Ordinary Least Squares Regressions of Δ(CEO Salary + Bonus) on Δ(Shareholder Wealth) and Δ ln(CEO Salary + Bonus) on Shareholder Return, Including Interactions to Allow the Pay-Performance Relation to Vary with Years Remaining as CEO and with Firm Size as Measured by Sales

Independent Variable	Dependent Variable: Δ(CEO Salary + Bonus) (1)	(2)	Dependent Variable: Δ ln(CEO Salary + Bonus) (3)	(4)
Few years left* (dummy variable)	−3.06	−2.63	−.0090	−.0091
	(−1.0)	(−.9)	(−2.0)	(−2.0)
Δshareholder wealth × sales	...	-4.0×10^{-7}	...	...
		(−4.2)		
Δshareholder wealth × (sales)^2	...	3.1×10^{-12}	...	...
		(3.7)		
Δshareholder wealth × few years left	.0008	.0025	...	...
	(.4)	(1.2)		
Shareholder return × sales	...	...	...	2.7×10^{-6}
				(1.5)
Shareholder return × (sales)^2	...	...	...	-3.2×10^{-12}
				(−.1)
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

Note.—The sample is constructed from longitudinal data reported in *Forbes* on 1,292 CEOs serving in 785 firms who leave their firms during the 1970–88 sample period. Compensation is measured in thousands of 1988-constant dollars; Δ(shareholder wealth) and sales are measured in millions of 1988-constant dollars. All regressions include intercepts and year dummies. Cols. 1 and 2 include Δshareholder wealth and Δshareholder wealth interacted with the year dummies, allowing the pay-performance sensitivity to vary by year; cols. 3 and 4 include shareholder return and shareholder return interacted with the year dummies, allowing the pay-performance elasticity to vary by year. *t*-statistics are in parentheses.

* Few years left is a dummy variable that equals one if the CEO is in his last three years as CEO.

increases the magnitude and significance of the $\Delta V_t \times$ few years left coefficient, although it remains insignificant.

The increase in the interaction Δshareholder wealth × 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. Because the pay-performance *elasticity* is relatively invariant to firm size, one way to control for size-related heterogeneity is to convert the regression variables to percentage changes or logarithmic changes (see, e.g., Coughlan and Schmidt 1985; Gibbons and Murphy 1990).

Column 3 of table 3 repeats the elasticity specification (35) 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 percent 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 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 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:

$$\begin{aligned}\Delta \ln(w_{it}) = {} & \alpha_i + \alpha_1(\text{few years left})_{it} \\ & + \sum_{n=1972}^{1988} \alpha_n(n\text{th year dummy})_{it} \\ & + \Bigg[\beta_i + \beta_1(\text{few years left})_{it} \\ & + \sum_{n=1972}^{1988} \beta_n(n\text{th year dummy})_{it}\Bigg] \times \Delta \ln(V_{it}),\end{aligned} \tag{36}$$

where α_i and β_i are firm-specific intercepts and pay-performance elasticities, respectively.

We estimated regression (36) for the subsample of 625 CEOs from the completed-spells subsample that had at least four years of data.[12] The estimated coefficient on (few years left)$_{it}$ 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 1 is somewhat diminished after

[12] We estimated (36) in two stages. In the first stage, we regressed the dependent and each of the independent variables in (36) on shareholder return, $\Delta \ln(V_{it})$, by CEO. In the second stage, we regressed the residual from the first-stage regression involving the dependent variable in (36) on the residuals from the first-stage regressions involving the independent variables in (36). After standard errors are adjusted for the correct degrees of freedom, the results from the second-stage regression are equivalent to estimating (36) with CEO-specific intercepts and slopes (on shareholder return).

we control for firm-specific heterogeneity by allowing the pay-performance elasticity to vary across firms.

IV. 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, even in the presence of contracts, and that optimal contracts account for these implicit incentives—seem likely to be robust.

The idea that optimal incentive contracts optimize *total* incentives can be applied to promotions.[13] 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 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 on the T-period career we take as exogenous.

Appendix

This Appendix discusses renegotiation-proof long-term contracts and the solution to the T-period model, and presents estimated pay-performance relations by year.

A. *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

[13] Rosen's (1986) model can be interpreted in terms of promotions but does not allow new jobs to have new technologies or workers to be matched to jobs.

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 $w_2(y_1, y_2) = c_2 + b_{21} y_1 + b_{22} y_2$. Renegotiation-proofness then implies that $b_{22} = b_2^*$ from (13) and hence that the worker's second-period effort choice is $a_2^*(b_2^*)$ satisfying (6). Furthermore, given the utility function (2), we can set $b_{21} = 0$: 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,

$$\begin{aligned} -E\{\exp(-r\{c_1 + b_1[\eta + a_1^*(b_1) + \epsilon_1] - g(a_1)\} \\ -r\delta\{c_2 + b_2^*[\eta + a_2^*(b_2^*) + \epsilon_2] - g(a_2^*(b_2^*))\})\}, \end{aligned} \tag{A1}$$

subject to a zero-expected-profit constraint for the firm,

$$c_1 + \delta c_2 = (1 - b_1)[m_0 + a_1^*(b_1)] + \delta(1 - b_2^*)[m_0 + a_2^*(b_2^*)]. \tag{A2}$$

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

$$b_1 = \frac{1}{1 + r\Sigma_1^2 g''[a_1^*(b_1)]} - \frac{r\delta b_2^* \sigma_0^2 g''[a_1^*(b_1)]}{1 + r\Sigma_1^2 g''[a_1^*(b_1)]}, \tag{A3}$$

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 provides exactly the same incentives as the optimal renegotiation-proof long-term contract derived here.

B. *The* T-*Period Model*

We first derive the first-order condition (24), the expected utility (25), and the optimal total incentive given implicitly in (26). We then prove that contractual incentives increase monotonically: for any $t < T$, $b_t^* < b_{t+1}^*$.

Substituting (21), (22), and (23) into the worker's utility function (2) yields the worker's utility from a sequence of contracts with slopes $(b_1, \ldots, b_T)$ and associated intercepts given by (21), given the market's conjecture $(\hat{a}_1, \ldots, \hat{a}_T)$, the worker's effort choices $(a_1, \ldots, a_T)$, and the output realizations $(y_1, \ldots, y_T)$:

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

$$= -\exp\left(-r\left\{\sum_{t=1}^{T}\delta^{t-1}\left[(1-b_t)\hat{a}_t + \frac{(1-b_t)\sigma_\epsilon^2 m_0}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} + b_t y_t - g(a_t) + \frac{(1-b_t)\sigma_0^2}{\sigma_\epsilon^2 + (t-1)\sigma_0^2}\sum_{\tau=1}^{t-1}(y_\tau - \hat{a}_\tau)\right]\right\}\right)$$

$$= -\exp\left(-r\left\{\sum_{t=1}^{T}\delta^{t-1}\left[(1-b_t)\hat{a}_t + \frac{(1-b_t)\sigma_\epsilon^2 m_0}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} + b_t y_t - g(a_t) + \sum_{\tau=t+1}^{T}\delta^{\tau-t}\frac{(1-b_\tau)\sigma_0^2}{\sigma_\epsilon^2 + (\tau-1)\sigma_0^2}(y_t - \hat{a}_t)\right]\right\}\right) \tag{A4}$$

$$= -\exp\left(-r\left\{\sum_{t=1}^{T}\delta^{t-1}\left[\hat{a}_t + \frac{(1-b_t)\sigma_\epsilon^2 m_0}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} - g(a_t) + B_t(y_t - \hat{a}_t)\right]\right\}\right)$$

$$= -\exp\left(-r\left\{\sum_{t=1}^{T}\delta^{t-1}\left[\hat{a}_t + \frac{(1-b_t)\sigma_\epsilon^2 m_0}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} - g(a_t) + B_t(a_t - \hat{a}_t)\right]\right\}\right) \cdot \exp\left\{-r\left[\sum_{t=1}^{T}\delta^{t-1}B_t(\eta + \epsilon_t)\right]\right\},$$

where B_t is defined in (24). To apply (12) to the last term in (A4), note that $\Sigma_{t=1}^{T}\delta^{t-1}B_t(\eta + \epsilon_t)$ is normally distributed with mean M and variance V, where

$$M = \left(\sum_{t=1}^{T}\delta^{t-1}B_t\right)m_0$$

and

$$V = \left(\sum_{t=1}^{T}\delta^{t-1}B_t\right)^2\sigma_0^2 + \sum_{t=1}^{T}(\delta^{t-1}B_t)^2\sigma_\epsilon^2.$$

Maximizing the worker's expected utility with respect to $(a_1, \ldots, a_T)$ then yields the first-order conditions (24). After we impose the equilibrium condition $(\hat{a}_1, \ldots, \hat{a}_T) = (a_1^*, \ldots, a_T^*)$, the worker's expected utility from a sequence of contracts with slopes $(b_1, \ldots, b_T)$ becomes

$$EU = -\exp\left(-r\left\{\sum_{t=1}^{T}\delta^{t-1}\left[a_t^* + \frac{(1-b_t)\sigma_\epsilon^2 m_0}{\sigma_\epsilon^2 + (t-1)\sigma_0^2} - g(a_t^*) + B_t m_0\right]\right\}\right) \cdot \exp\left\{\tfrac{1}{2}r^2\left[\left(\sum_{t=1}^{T}\delta^{t-1}B_t\right)^2\sigma_0^2 + \sum_{t=1}^{T}(\delta^{t-1}B_t)^2\sigma_\epsilon^2\right]\right\},$$

which simplifies to (25). To show that (26)—the first-order condition for the optimal total incentive—is sufficient, note that the expected utility (25) is quasi-concave in B_1 because it is concave for $B_1 < 1$ and (from lemma 1 below) decreasing for $B_1 \geq 1$.

We now show that $b_t^* < b_{t+1}^*$ for any $t < T$. We begin by showing (by induction) that $B_t^* < B_{t+1}^*$ for any $t < T$. Restating (26) yields the first-order condition for B_t^*:

$$1 - B_t^* - rg''(a_t^*)\left(\Sigma_t^2 B_t^* + \sigma_{t-1}^2 \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^*\right) = 0, \tag{A5}$$

where $\Sigma_{t+1}^2 = \sigma_t^2 + \sigma_\epsilon^2$ and $\sigma_t^2 = \sigma_0^2\sigma_\epsilon^2/(t\sigma_0^2 + \sigma_\epsilon^2)$ is the conditional variance of η given the previous t output observations. Since the objective function is quasi-concave, the left side of (A5) is negative (positive) when evaluated at an arbitrary $B_t >$ ($<$) B_t^*. Thus $B_t^* < B_{t+1}^*$ if and only if

$$1 - B_{t+1}^* - rg''(a_{t+1}^*)\left(\Sigma_t^2 B_{t+1}^* + \sigma_{t-1}^2 \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^*\right) < 0, \tag{A6}$$

where $g''(a_{t+1}^*)$ appears in (A6) because $g'(a_t) = B_t$ from (24). Lemma 2 below uses the first-order condition for B_{t+1}^* analogous to (A5) to restate (A6) in a more convenient form. Lemma 3 then uses the induction hypothesis $B_{t+1}^* < B_{t+2}^*$ to prove that this restatement of (A6) holds. To show that $B_t^* < B_{t+1}^*$ for any $t < T$, it then remains to show only that the initial induction step $B_{T-1}^* < B_T^*$ holds, which follows because (A5) applied to period T yields

$$(1 - B_T^*) - rg''(a_T^*)\Sigma_T^2 B_T^* = 0, \tag{A7}$$

so $\Sigma_T^2 < \Sigma_{T-1}^2$ implies

$$(1 - B_T^*) - rg''(a_T^*)(B_T^*\Sigma_{T-1}^2 + \sigma_{T-2}^2\delta B_T^*) < 0,$$

analogous to (A6). Finally, lemma 4 uses another induction argument to show that $b_t^* < b_{t+1}^*$ for any $t < T$.

LEMMA 1. For each $t \in \{1, \ldots, T-1\}$, $\Sigma_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* \geq 0$.

Proof. By induction. For $t = T - 1$, $B_T^* > 0$ from (A7). For $t < T - 1$, suppose $\Sigma_{\tau=t+2}^{T} \delta^{\tau-t-1} B_\tau^* \geq 0$. We must show

$$\sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* = \delta B_{t+1}^* + \delta \sum_{\tau=t+2}^{T} \delta^{\tau-t-1} B_\tau^* \geq 0. \tag{A8}$$

Suppose that (A8) fails. Then (A5) implies

$$B_{t+1}^* = \frac{1 - r\sigma_t^2 g''(a_{t+1}^*) \sum_{\tau=t+2}^{T} \delta^{\tau-t-1} B_t^*}{1 + r\Sigma_{t+1}^2 g''(a_{t+1}^*)} < -\sum_{\tau=t+2}^{T} \delta^{\tau-t-1} B_\tau^*, \tag{A9}$$

which implies

$$\sum_{\tau=t+2}^{T} \delta^{\tau-t-1} B_\tau^* < \frac{-1}{1 + r\sigma_\epsilon^2 g''(a_{t+1}^*)} < 0,$$

which contradicts the induction hypothesis. Q.E.D.

LEMMA 2. $B_t^* < B_{t+1}^*$ if and only if

$$rg''(a_{t+1}^*)\sigma_\epsilon^2(1 - \delta)B_{t+1}^* - (1 - B_{t+1}^*)\left(\frac{\sigma_\epsilon^2}{\sigma_{t-1}^2 + \sigma_\epsilon^2} - \delta\right) > 0. \tag{A10}$$

Proof. We argued above that $B_t^* < B_{t+1}^*$ if and only if (A6) holds. Using (A5) applied to period $t + 1$, we can eliminate the summation in (A6) and rewrite that inequality as

$$1 - B_{t+1}^* - rg''(a_{t+1}^*)(\Sigma_t^2 + \delta\sigma_{t-1}^2)B_{t+1}^*$$
$$- \frac{\delta\sigma_{t-1}^2}{\sigma_t^2}[1 - B_{t+1}^* - rg''(a_{t+1}^*)\Sigma_{t+1}^2 B_{t+1}^*] < 0.$$

Applying the definitions of σ_t^2 and Σ_t^2 then yields (A10). Q.E.D.

LEMMA 3. If $B_{t+1}^* < B_{t+2}^*$, then

$$rg''(a_{t+1}^*)\sigma_\epsilon^2(1 - \delta)B_{t+1}^* - (1 - B_{t+1}^*)\left(\frac{\sigma_\epsilon^2}{\sigma_t^2 + \sigma_\epsilon^2} - \delta\right) > 0, \tag{A11}$$

which implies (A10).

Proof. Using (A5) for periods $t + 1$ and $t + 2$, we get

$$\frac{1 - B_{t+1}^* - rg''(a_{t+1}^*)\Sigma_{t+1}^2 B_{t+1}^*}{rg''(a_{t+1}^*)\sigma_t^2} = \delta B_{t+2}^* + \delta\frac{1 - B_{t+2}^* - rg''(a_{t+2}^*)\Sigma_{t+2}^2 B_{t+2}^*}{rg''(a_{t+2}^*)\sigma_{t+1}^2},$$

or

$$\begin{aligned} &g''(a_{t+2}^*)\sigma_{t+1}^2[1 - B_{t+1}^* - rg''(a_{t+1}^*)\Sigma_{t+1}^2 B_{t+1}^*] \\ &+ \delta g''(a_{t+1}^*)\sigma_t^2[B_{t+2}^* + rg''(a_{t+2}^*)\sigma_\epsilon^2 B_{t+2}^* - 1] = 0. \end{aligned} \tag{A12}$$

Define

$$\begin{aligned} h(B) = {} & g''(a)\sigma_{t+1}^2[1 - B_{t+1}^* - rg''(a_{t+1}^*)\Sigma_{t+1}^2 B_{t+1}^*] \\ &+ \delta g''(a_{t+1}^*)\sigma_t^2[B + rg''(a)\sigma_\epsilon^2 B - 1], \end{aligned}$$

where $g'(a) = B$, so that (A12) may be restated as $h(B_{t+2}^*) = 0$. Then lemma 1 and $g'''(a) \geq 0$ imply that $h'(B) > 0$. Therefore, the induction hypothesis $B_{t+1}^* < B_{t+2}^*$ implies that $h(B_{t+1}^*) < 0$, which simplifies to (A11). To establish (A10) from (A11), note that $\sigma_t^2 < \sigma_{t-1}^2$ and that lemma 1 and (A5) imply $B_{t+1}^* < 1$. Q.E.D.

LEMMA 4. If $B_t^* < B_{t+1}^*$ for all $t < T$, then $b_t^* < b_{t+1}^*$ for all $t < T$.

Proof. Since $b_T^* = B_T^*$ and $b_{T-1}^* < B_{T-1}^*$, we get $b_{T-1}^* < b_T^*$. We now complete the proof by showing that if $b_{t+1}^* < b_{t+2}^* < \ldots < b_T^*$ and $B_t^* < B_{t+1}^*$, then $b_t^* < b_{t+1}^*$. From (24) applied to periods t and $t + 1$ we get

$$B_t^* = b_t^* + \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - b_\tau^*)\frac{\sigma_0^2}{\sigma_\epsilon^2 + (\tau - 1)\sigma_0^2}$$

and

$$\begin{aligned} B_{t+1}^* &= b_{t+1}^* + \sum_{\tau=t+2}^{T} \delta^{\tau-t-1}(1 - b_\tau^*)\frac{\sigma_0^2}{\sigma_\epsilon^2 + (\tau - 1)\sigma_0^2} \\ &= b_{t+1}^* + \sum_{\tau=t+1}^{T-1} \delta^{\tau-t}(1 - b_{\tau+1}^*)\frac{\sigma_0^2}{\sigma_\epsilon^2 + \tau\sigma_0^2}. \end{aligned}$$

Thus $B_t^* < B_{t+1}^*$ implies $b_t^* < b_{t+1}^*$ provided that

$$\sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - b_\tau^*)\frac{\sigma_0^2}{\sigma_\epsilon^2 + (\tau - 1)\sigma_0^2} - \sum_{\tau=t+1}^{T-1} \delta^{\tau-t}(1 - b_{\tau+1}^*)\frac{\sigma_0^2}{\sigma_\epsilon^2 + \tau\sigma_0^2} > 0,$$

which follows from $b_T^* < 1$ and $b_\tau^* < b_{\tau+1}^*$ for $\tau = t + 1, \ldots, T - 1$. Q.E.D.

C. *Estimated Pay-Performance Relations by Year*

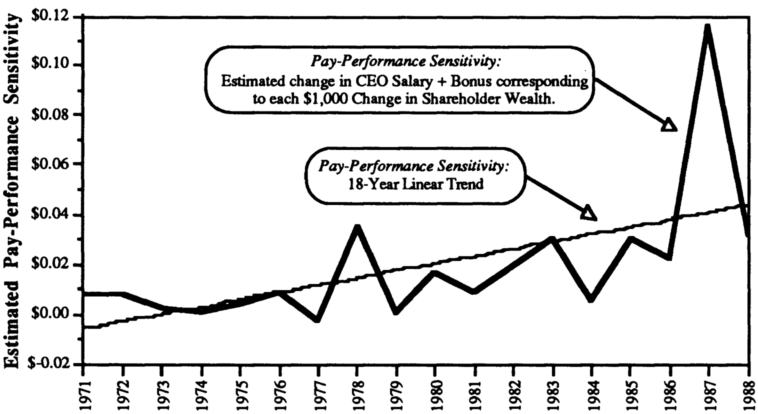

FIG. A1.—Estimated pay-performance sensitivities, by year. Figure is based on annual regressions of Δ(salary + bonus [$ thousands]) on Δ(shareholder wealth [$ millions]) using *Forbes* data on 2,236 CEOs serving in 1,204 firms. Each year's regression is based on approximately 650 observations.

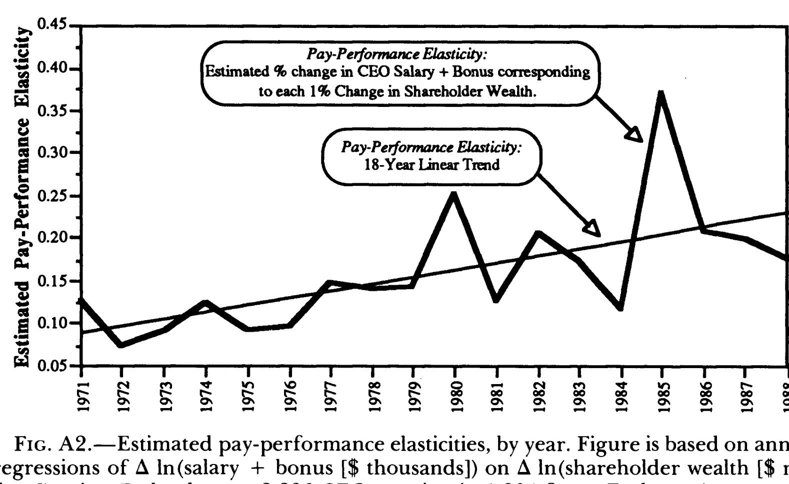

FIG. A2.—Estimated pay-performance elasticities, by year. Figure is based on annual regressions of Δ ln(salary + bonus [$ thousands]) on Δ ln(shareholder wealth [$ millions]) using *Forbes* data on 2,236 CEOs serving in 1,204 firms. Each year's regression is based on approximately 650 observations.

References

Aron, Debra J. "Worker Reputation and Productivity Incentives." *J. Labor Econ.* 5, no. 4, pt. 2 (October 1987): S87–S106.

Baron, David P., and Besanko, David. "Commitment and Fairness in a Dynamic Regulatory Relationship." *Rev. Econ. Studies* 54 (July 1987): 413–36.

Coughlan, Anne T., and Schmidt, Ronald M. "Executive Compensation, Management Turnover, and Firm Performance: An Empirical Investigation." *J. Accounting and Econ.* 7 (April 1985): 43–66.

DeGroot, Morris H. *Optimal Statistical Decisions.* New York: McGraw-Hill, 1970.

Dewatripont, Mathias. "Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts." *Q.J.E.* 104 (August 1989): 589–619.

Fama, Eugene F. "Agency Problems and the Theory of the Firm." *J.P.E.* 88 (April 1980): 288–307.

Freixas, Xavier; Guesnerie, Roger; and Tirole, Jean. "Planning under Incomplete Information and the Ratchet Effect." *Rev. Econ. Studies* 52 (April 1985): 173–91.

Fudenberg, Drew; Holmström, Bengt; and Milgrom, Paul R. "Short-Term Contracts and Long-Term Agency Relationships." *J. Econ. Theory* 51 (June 1990): 1–31.

Fudenberg, Drew, and Tirole, Jean. "Moral Hazard and Renegotiation in Agency Contracts." *Econometrica* 58 (November 1990): 1279–1319.

Gibbons, Robert. "Piece-Rate Incentive Schemes." *J. Labor Econ.* 5, no. 4, pt. 1 (October 1987): 413–29.

Gibbons, Robert, and Murphy, Kevin J. "Relative Performance Evaluation for Chief Executive Officers." *Indus. and Labor Relations Rev.* 43 (suppl.; February 1990): 30S–51S.

Harris, Milton, and Holmström, Bengt. "A Theory of Wage Dynamics." *Rev. Econ. Studies* 49 (July 1982): 315–33.

Hart, Oliver D., and Moore, John. "Incomplete Contracts and Renegotiations." *Econometrica* 56 (July 1988): 755–85.

Holmström, Bengt. "Managerial Incentive Schemes—a Dynamic Perspective." In *Essays in Economics and Management in Honour of Lars Wahlbeck.* Helsinki: Swenska Handelshögkolan, 1982.

Holmström, Bengt, and Milgrom, Paul R. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55 (March 1987): 303–28.

Holmström, Bengt, and Ricart i Costa, Joan E. "Managerial Incentives and Capital Management." *Q.J.E.* 101 (November 1986): 835–60.

Jensen, Michael C., and Murphy, Kevin J. "Performance Pay and Top-Management Incentives." *J.P.E.* 98 (April 1990): 225–64.

Kanemoto, Yoshitsugu, and MacLeod, W. Bentley. "The Ratchet Effect, the Market for Senior Workers, and Intertemporal Commitment." *J. Labor Econ.* 10 (January 1992): 85–98.

Laffont, Jean-Jacques, and Tirole, Jean. "The Dynamics of Incentive Contracts." *Econometrica* 56 (September 1988): 1153–75.

Lambert, Richard A. "Long-Term Contracts and Moral Hazard." *Bell J. Econ.* 14 (Autumn 1983): 441–52.

Lazear, Edward P. "Salaries and Piece Rates." *J. Bus.* 59 (July 1986): 405–31.

Lazear, Edward P., and Rosen, Sherwin. "Rank-Order Tournaments as Optimum Labor Contracts." *J.P.E.* 89 (October 1981): 841–64.

MacDonald, Glenn M. "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information." *A.E.R.* 72 (December 1982): 1038–55.

MacLeod, W. Bentley, and Malcomson, James M. "Reputation and Hierarchy in Dynamic Models of Employment." *J.P.E.* 96 (August 1988): 832–54.

Murphy, Kevin J. "Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts." *Rand J. Econ.* 17 (Spring 1986): 59–76.

Ricart i Costa, Joan E. "Managerial Task Assignment and Promotions." *Econometrica* 56 (March 1988): 449–66.

Rogerson, William P. "Repeated Moral Hazard." *Econometrica* 53 (January 1985): 69–76.

Rosen, Sherwin. "Prizes and Incentives in Elimination Tournaments." *A.E.R.* 76 (September 1986): 701–15.

Smith, Clifford W., Jr., and Warner, Jerold B. "On Financial Contracting: An Analysis of Bond Covenants." *J. Financial Econ.* 7 (June 1979): 117–61.

Vancil, Richard F. *Passing the Baton: Managing the Process of CEO Succession.* Boston: Harvard Bus. School Press, 1987.

Waldman, Michael. "Job Assignments, Signalling, and Efficiency." *Rand J. Econ.* 15 (Summer 1984): 255–67.