

MANAGERIAL COMPENSATION AND THE COST OF
MORAL HAZARD*

BY MARY M. MARGIOTTA AND ROBERT A. MILLER^{†,1}

*A. T. Kearney, U.S.A.
Carnegie Mellon University, U.S.A.*

This article investigates managerial compensation and its incentive effects. Our econometric framework is derived from a multiperiod principal-agent model with moral hazard. Longitudinal data on returns to firms and managerial compensation are used to estimate the model. We find that firms would incur large losses from ignoring moral hazard, whereas managers only require moderate additional compensation for accepting a contract that ties their wealth to the value of the firm. Thus the costs of aligning hidden managerial actions to shareholder goals through the compensation schedule are much less than the benefits from the resulting managerial performance.

1. INTRODUCTION

Previous research has established several empirical regularities that relate managerial compensation to financial and accounting measures of firm performance. First, executive compensation (whether measured as salary and bonus alone or including the value of nonmonetary compensation such as stocks and options granted) increases with stock market returns on equity and accounting measures of net income (see, e.g., Murphy, 1985). Second, Lambert and Larcker (1987) report that greater weight in the compensation package is given to measures of firm performance that are more closely related to managerial inputs. For example, given his or her own firm's profits, a manager is rewarded more if rival firms perform poorly than if the whole industry does well (see Antle and Smith, 1985, 1986).

These findings have been interpreted as evidence for the presence of moral hazard that arises when shareholders delegate authority to an executive whose interests differ from those of the shareholders and whose actions are not observed directly. Although they may be rationalized by agency theory [as developed by Holmstrom (1979, 1982) and others], the empirical regularities just mentioned also could emerge in economies where no informational asymmetries exist. For example,

* Manuscript received September 1996; revised August 1998.

[†] E-mail: ramiller@andrew.cmu.edu.

¹ We are indebted to Rick Antle and Abbie Smith for making their data available to us, to Rafael Mendible for research assistance, and to two referees for their comments. This article has benefited from presentations at the May 1990 conference on empirical applications of structural models organized by the Econometric Society and SSRI at the University of Wisconsin, as well as seminars at several business schools and economics departments. Our work was partially funded by the National Science Foundation Grant SES-9108659.

in competitive equilibrium, workers are paid the value of their marginal product, which is partly determined by the firm's product price; to the extent that price movement reflects changing demand (as opposed to supply) conditions, compensation and firm performance are positively correlated, the first finding. Moreover, the services of superior managers presumably command a premium, which could explain the second.

There is, nevertheless, another reason to believe moral hazard is important in models of executive decision making and compensation; as a fraction of their annual income, a typical manager holds a sizable quantity of his or her firm's stock, exposing himself or herself to idiosyncratic fluctuations in its return. For example, the data used in this study show that the variability in managerial income attributable to holding the firm's stocks and options on such stocks far exceeds the variability in cash and bonus.² If agents are risk-averse, such portfolios seem suboptimal. In principle, there are reasons why risk-averse managers voluntarily hold stock in their own firms. First, if the value of the manager's specific capital (such as his or her bargaining power with the board of directors) is negatively correlated with these idiosyncratic fluctuations, then he or she may choose to offset this risk by purchasing shares in the firm. If this factor were important, however, it would help smooth total payments to managers of large corporations, not exacerbate volatility, which is what our data show. Second, if externalities generated by the firm (such as pollution) harm the manager and are greater when fluctuations are positive, then risk aversion could lead the manager to insure himself or herself by holding inside stock. On the other hand, the value of his or her other personal assets (including human capital and maybe housing stock) is probably positively correlated with the firm's value, and in this case, he or she would prefer to take a negative position with respect to the firm for the purposes of diversification. Third, managers may indeed pay lower transaction costs from holding portfolios of their own firms, but this seems small compared with the high amount of undiversified risk they expose themselves to in the process. To summarize, the rationale for managers holding financial claims to their own firms seems somewhat contrived in economies lacking private information.

In a widely cited article, Jensen and Murphy (1990) have argued that while the annual change in executive compensation is positively related to changes in shareholder wealth (as agency theory predicts), the magnitude of the response coefficient is inconsequential. The authors recognize that if agents are risk-averse, the only reason for imposing additional risk on them is to induce an action that they would not otherwise undertake. However, they apparently overlook the possibility that only a small stake in the firm may be necessary to dissuade the manager from pursuing actions that are disastrous for shareholders at large; consequently, the performance-to-compensation ratio the authors advocate as an empirical benchmark for measuring moral hazard is questionable on theoretical grounds. An alternative approach, taken here, is to evaluate the importance of moral hazard from estimated structural parameters of a model in which managerial preferences, the effects of his or her actions on the firm's returns, and shareholders' information sets are explicitly laid out. This article, then, is an empirical application of the principal-agent paradigm to

² See Table 1 in Section 8, which describes the data set, for more detail on this point.

investigate managerial compensation schemes. It estimates a simple dynamic model of optimal contracting in the presence of moral hazard that arises because of differences over optimal work effort and the inability of shareholders to monitor effort. Our approach is similar to Ferrall and Shearer (1999), who estimate a model based on a linear contracting and compare it with alternative schemes, including the interim efficient contract that forms the benchmark for this article.

Sections 3 through 7 comprise the theory and in doing so extend the literature on optimal contracting for moral hazard to economies supporting time-varying interest rates and stock market volatility. The model is described in Section 2, drawing extensively on the principal-agent frameworks of Malcolmson and Spinnewyn (1988), Fudenberg et al. (1990), and Rey and Salanie (1990). Two critical assumptions of this extension are that a complete set of markets exist but that the principal directly observes contingent claims traded by the agent on his or her own account. The latter assumption, which roughly corresponds to public disclosure laws currently in place, prevents the agent from undoing the risk the principal imposes on the agent to make the contract incentive compatible. At the same time, the notion that firms are value-maximizers can be formally defended. Because shareholders may trade in consumption claims that are contingent on any output realization, potential conflicts among themselves about goals their manager should pursue [discussed in Grossman and Hart (1979), for example] are eliminated. Moreover, given the empirical results of Altug and Miller (1990) from panel data on male labor supply in all occupations, in which there is little evidence against the hypothesis that markets are competitive and complete, the approach taken here, ignoring other distortions in the economy, and thus disregarding second-best considerations, seems a reasonable empirical benchmark to gauge the importance of moral hazard in this part of the labor sector, the market for managerial services.³

Then in Section 3 we analyze the manager's intertemporal consumption problem given his or her labor decisions and an implied distribution of future compensation when an economy with aggregate shocks is supported by complete competitive markets complemented by public disclosure conditions that allow shareholders to prevent the manager from diversifying the uncertainty of his or her future income. Here we highlight the importance of a public disclosure condition, necessary to supplement the standard conditions that are required to enforce solvency conditions that are necessary when there is trading in contingent claims. To induce the manager to work hard within the firm, shareholders must respect two constraints analyzed in Section 4, respectively called *participation* (which makes outside offers look sufficiently unappealing to the manager) and *incentive compatibility* (which aligns his or her personal objectives within the firm with those of shareholders).

Section 5 characterizes the optimal contract. Following Grossman and Hart (1983), we derive the schedule that minimizes the expected costs to shareholders from inducing the manager to act in their best interests. Because we assume that the manager's preferences in current consumption are exponential, wealth plays a

³ For critiques of assuming complete markets to identify structural parameters in empirical models of intertemporal labor supply and consumption decisions, see Miller and Sieg (1997) and the survey by Miller (1997).

limited role in the optimal contract. Given the constraints imposed by the data used in our empirical analysis, which contains no information on the manager's wealth apart from the firm-specific financial instruments he or she holds, this simplification makes the empirical analysis more tractable. Although the optimal contract does not have a closed form but must be solved numerically, its main features are evident from casual inspection. In particular, the dependence of managerial compensation on the firm's abnormal returns arises from the likelihood ratio of the probability density function for shirking relative to the corresponding density for diligence. If this ratio is decreasing in returns, then the optimal contract requires compensation to be monotone increasing in returns. The short-term optimal contract derived in Sections 3 through 5 survives a relaxation of two assumptions that may look questionable from an empiricist's perspective. First, we establish that the optimal long-term contract is essentially a replication of the short-term contract derived in the preceding section. (Thus the fact that managers tend to hold their jobs for more than a year may not bode badly for our theory.) Second, we also note that, with some minor modifications, the theory readily extends to situations where several managers all work for the one firm, a useful extension for our application because the data set contains information on the three top managers rather than just the CEO. These extensions are undertaken in Section 6.

Finally, Section 7 concludes our theoretical discussion by deriving three measures of the importance of moral hazard in this model. The first of these measures is the difference between the expected compensation paid under the optimal contract inducing high effort and the fixed amount that would be paid if the shareholders observed effort and chose to make the manager work hard. This cost is the risk premium that must be paid to the manager and represents the costs to shareholders of hidden actions. The second measure calculates the amount of additional compensation required to compensate the manager for taking higher effort. The sum of the first two measures represents the amount of total additional compensation paid to induce high effort when actions are the private information of managers. The final measure calculates the gross increase in shareholder wealth from implementing the optimal contract. It is calculated as the expected gross rise in firm values when managers choose high effort before subtracting the costs associated with inducing high effort, namely, the first two measures.

The second half of our paper, Sections 8 through 11, applies the theory to a panel of firms and their managers. Much of the data set, described in Section 8, was originally compiled by Masson (1971) and substantially extended by Antle and Smith (1985, 1986). The variables of concern include characteristics of the firm (such as its industry) and their managers (such as their ages and position within the firm), the return on the firm's assets, the return on the market portfolio, and other aggregate variables such as the interest rate, managerial compensation, and that portion of his or her wealth portfolio comprising financial claims on his or her firm. As a preliminary exercise to check its comparability with previous work, and as further motivation for our own structural analysis, we conducted several diagnostic checks with data, and these are also reported in Section 8. To summarize, while cash and bonus payouts to managers are insensitive to the firm's abnormal returns, more inclusive measures of compensation are significantly positively correlated, a finding

that is consistent with other data sets on managerial compensation.⁴ Although our finding that executive income increases \$16.57 when the abnormal income to the firm rises by \$1000 is an order of magnitude higher than the numbers reported in Jensen and Murphy (1990), their basic point, that the total loss to managers is much less than the total loss to shareholders from the firm failing, seems unimpeachable. In addition, we find that separations are not highly correlated with poor firm performance within the group of survivor firms we analyze, weak evidence that the act of severance is seldom used as a disciplinary tool to elicit superior effort.

The parameterization of abnormal returns, identification, and estimation are discussed in Section 9. The distributional assumption made for abnormal returns, truncated normal, was partly selected for its three-parameter flexibility and partly because of the empirical distribution itself, which bears more than a passing resemblance to our estimated parameterization. While the manager never disobeys any instructions he or she receives, so that points on the shirking distribution are never directly observed, the model is nevertheless identified, because the losses to the company and the personal nonpecuniary gains to him or her from disobeying the shareholders are reflected in the compensation plan itself. A sequential procedure was used to estimate a parameter vector determining the manager's utility function, the probability distribution characterizing excess returns to the company, and the dependence of both on what actions the manager takes.

Section 10 discusses the empirical results, and the last section offers some concluding remarks about possible directions for future research. Overall, the model does a reasonable job of fitting the data. Neither the ranking nor the signs of parameters presumed to be true by the theoretical model are violated by the parameter estimates we obtain, several of our estimates are quite precise, and the overidentifying restrictions we impose in estimation are not rejected by the data. We find firms would incur very large losses from not designing a compensation contract that induces the manager to act in the interests of shareholders. Compared with these potential losses, risk-averse managers receive only moderate increases in their expected compensation for accepting fluctuations in their wealth driven by the volatility of firm returns about the market portfolio. Similarly, the nonpecuniary benefits to managers from acting against the firm's interests are quite modest. Thus the cost of enforcing the incentive compatibility condition associated with the optimal contract is very small compared with the substantial benefits from aligning the hidden actions of managers with the goals of their shareholder clientele through the compensation schedule.

2. THE MODEL

This section lays out the theoretical framework on which our estimation is based. Consider some indivisible physical plants or firms owned by some well-diversified shareholders. The output of each depends, stochastically, on the effort its manager makes and also on general economic conditions. The economic conditions are observed by shareholders, but they cannot monitor managerial effort perfectly. The

⁴ See, for example, the recent paper by Hall and Liebman (1998).

firms are infinitely lived, but a sequence of executives successively manages them, if only because people have finite working lives. Managers are hired by the firms' owners from a population that looks identical both before and after the manager has acquired (firm-specific or general) working experience. There are no individual characteristics that make for good or bad job matches in this framework [in contrast to, say, Miller (1984)]. Also, well-functioning markets exist to help managers smooth their respective consumption streams, so these firms need not serve as financial intermediaries.

At the beginning of a typical period, denoted t , first the results of last period's production are revealed, allowing the return on each firm's stock to be calculated. Each manager, generically denoted by n , is then compensated according to the executive compensation plan previously agreed to in period $t - 1$. Next, the n th manager chooses his or her consumption, denoted c_{nt} , his or her asset portfolio e_{nt} , and whether to retire or not l_{0nt} . The shareholders to whom he or she is accountable observe these choices (c_{nt} , e_{nt} , l_{0nt}). In particular, we assume that extant public disclosure laws prevent managers from anonymously trading the assets of their own firms. Then the shareholders announce a production and compensation plan for period t . (Alternatively, competing candidates to manage the firm could propose their respective production and compensation plans so that the shareholders can vote among themselves which contract to accept.) Finally, the (designated) manager chooses his or her effort level (l_{1nt} , l_{2nt}).

An equilibrium for this model is defined as a set of market-clearing prices for contingent claims to consumption, plus a set of contingent compensation schemes announced by shareholders, that are individually optimal for the respective parties to sequentially implement, given the constraints they face. In equilibrium, there are two constraints production and compensation plans must satisfy. First, the manager's expected utility from working for the firm must exceed his or her expected utility from any alternative use of his or her time. The second constraint is motivated by the fact that in the model, the manager's actions, which affect the probability distribution determining the firm's output, are not observed by its residual claimants, the shareholders. Consequently, the executive compensation plan must tie the manager's interests to those of his or her firm's shareholders in order to induce him or her to work on their behalf. Apart from managerial effort, all information is public, and there are no barriers to trade. Markets to contingent claims are competitive and complete. Consequently, the only reason why first best consumption and leisure allocations are not attained revolves around the problem of aligning managerial with shareholder interests. We assume that shareholders can monitor portfolio decisions of managers, so managers cannot insure themselves against outcomes that pay them poorly by making offsetting trades in contingent claims—in other words, that laws against insider trading are easy to effectively enforce.

Formally, uncertainty about public information is treated as a probability space (Ω, F, P) . The element $\omega \in \Omega$ signifies a particular realization of all publicly observed (random) variables in this economy from time 0 onward. Let F_0, F_1, F_2, \dots, F denote the increasing sequence of σ -algebras that characterizes how public information accumulates over time, and denote by E_t the expectations operator associated with F_t ; thus the random variables with t subscripts defined

below are F_t measurable. Following Debreu's (1959) treatment of uncertainty, the commodity space for consumption goods is formed from consumption units at date 0 plus claims to consumption units at subsequent dates that are contingent on how history unfolds. Markets exist for all claims to consumption that are contingent on any subset $A_t \in F_t$. Let Λ_t , a measure on F_t for each $t \in \{0, 1, 2, \dots\}$, denote date 0 prices of contingent claims to consumption to be delivered in period t , and denote by λ_t its Radon-Nikodym derivative. Thus the Λ_t sequence represents the set of equilibrium prices for the contingent consumption claims, denominated in consumption date 0 units, so $E[\lambda_t]$, for example, is the number of consumption units forgone in date 0 to obtain a sure consumption unit in date t , meaning that $\{E[\lambda_t]\}^{-1} - 1$ is the t period interest rate.

The choices of a manager are over the consumption and leisure allocations he or she receives each period of his or her life. Let $c_{nt} \in (-\infty, \infty)$ denote consumption by agent n in period t . There are three levels of labor activity, and we express the n th manager's choice in period t by a vector $l_{nt} = (l_{0nt}, l_{1nt}, l_{2nt})$, where $l_{jnt} \in \{0, 1\}$ is an indicator function for choice $j \in \{0, 1, 2\}$ and

$$(1) \quad \sum_{j=0}^{j=2} l_{jnt} = 1$$

If $l_{0nt} = 1$, we say the manager has retired, and this activity is publicly observed; for expositional convenience, we also assume that retirement is an irreversible decision, that is, $l_{0nt} = 1$ if $l_{0n,t-1} = 1$ and $t \leq \bar{n}$. The other two effort levels, respectively called *shirking* (designated by setting $l_{1nt} = 1$) and *working diligently* (when $l_{2nt} = 1$), are private information to the manager.

Preferences are parameterized as a time additively separable utility function in which consumption enters exponentially and leisure multiplicatively. Thus lifetime utility (from birth at \underline{n} to the exogenously determined date of death at $\bar{n} + 1$) can be expressed as

$$(2) \quad - \sum_{t=\underline{n}}^{t=\bar{n}} \beta^t (\alpha_0 l_{0nt} + \alpha_1 l_{1nt} + \alpha_2 l_{2nt}) \exp(-\rho c_{nt})$$

where $\beta \in (0, 1)$ is the common subjective discount factor, α_j is a utility parameter associated with choosing l_{jnt} , and ρ is the coefficient of absolute risk aversion. We assume that $\alpha_2 > \alpha_1$, or that diligence is more distasteful than shirking. This assumption is the vehicle by which the manager's preferences are not aligned with shareholder interests. We are not suggesting that managers are inherently lazy, merely that their personal goals do not automatically motivate them to maximize the value of the firm even if their compensation is independent of the firm's performance. One might conjecture that retirement is more enjoyable than working hard ($\alpha_2 > \alpha_0$), although this is not necessarily true for high-ranking management; perhaps the fact that they receive positive compensation is only partial compensation for the risk their undiversified portfolio exposes them to. Finally, the ordering of α_0 and α_1 is unclear because it depends, among other things, on perks associated with shirking.

Let w_{nt} denote compensation received by the n th manager at the beginning of period t , let $a_{n,t-1}$ denote the value of shareholder equity at the end of period $t-1$, and define $z_{nt} \equiv w_{nt}/a_{n,t-1}$ as the direct (accounting) effect of managerial compensation on the firm's returns realized at the beginning of next period, t . We denote by π_{nt} the return to the n th manager's firm at time t , define π_t as the return on the market portfolio, and let

$$(3) \quad x_{nt} \equiv \pi_{nt} - \pi_t + z_{nt}$$

Without further assumptions, Equation (3) stands as a tautology defining x_{nt} as gross abnormal returns to the n th manager's firm in period t , net of returns from holding the market portfolio, but not accounting for the costs of managerial compensation. This study assumes that x_{nt} is a random variable that depends on the manager's effort activity choice in the previous period but, conditional on $(I_{1n,t-1}, I_{2n,t-1})$, is independently and identically distributed across (n, t) . Given $I_{jn,t-1} = 1$ for $j \in \{1, 2\}$, we denote the probability density function of x_{nt} by $f_j(x_{nt})$. These assumptions about x_{nt} rule out several features of interest that may warrant further research. First, our analysis ignores the potential for private information about the abilities or managerial preferences, as well as actions taken earlier than the previous period, to affect current returns. Second, a more sophisticated model might analyze environments where publicly observed variables such as past returns and/or hidden actions such as managerial initiative affect higher-order moments of x_{nt} . For example, can managers affect the covariance of x_{nt} with π_t and in this way affect the value of the firm? Third, industry effects could be incorporated. If x_{nt} contained an industry effect (that reflected the ability of the industry to limit rivalry, attract subsidies from taxpayers, and so on), the determinants of an optimal compensation package would typically distinguish between those factors which affect the whole industry and those which only affect the firm.

Note that if $f_1(x_{nt}) > 0$ and $f_2(x_{nt}) = 0$ for all x_{nt} in any set with strictly positive measure, such as an open interval that occurs with strictly positive probability if the agent shirks, then either a two-part contract that generated the first best outcome would be optimal (as in Mirrlees, 1976) or, in the event of the manager being revealed to shirk just before his or her retirement, an optimal contract would require his or her remaining lifetime utility to attain its lower bound.⁵ Alternatively, if $f_2(x_{nt}) > 0$ and $f_1(x_{nt}) = 0$ for all x_{nt} in any set with strictly positive measure, then one can show that the highest payments to managers would occur in this region. Since managerial compensation is, roughly speaking, increasing in returns to the firm's assets and managers are not typically threatened with abject poverty for poor performance (or even conditions that remotely resemble those declaring personal bankruptcy), both specifications are somewhat counterfactual. Therefore, we shall assume that $f_1(x_{nt})$ and $f_2(x_{nt})$ have the same support, denoted by $[\psi, \infty)$.

⁵ This lower bound should not necessarily be interpreted as the worst imaginable utility but could be determined by laws about personal bankruptcy that prescribe what property must be forfeited, what may be retained, and how future opportunities for earning and consuming are affected.

3. INTERTEMPORAL CONSUMPTION

At the heart of the optimal contract for this model is the indirect utility function of the manager, which maps his or her expected utility as a function of the relevant security prices, the portion of his or her wealth that can be fully diversified, and the probability distribution for any unanticipated changes in his or her wealth induced by undiversifiable components to his or her compensation package. In particular, obtaining the valuation function at the time the manager enters his or her final period of work is a useful step toward deriving the solution to the optimal contract. For this reason, we now derive the valuation function for the following problem in consumption smoothing. The manager enters period t with wealth endowment of e_{nt} and chooses his or her consumption c_{nt} . At the beginning of the next period, he or she receives some additional, random amount of income $w_{n,t+1}$, and beyond that time, he or she consumes from his or her own augmented wealth until period $\bar{n} + 1$, when he or she dies. The easiest way to solve this problem is to split it into two phases. First, we derive the indirect utility function for the agent on entering retirement at the beginning of period $t + 1$. On retirement, each manager faces a single budget constraint for the remainder of his or her life, which implies that his or her intertemporal consumption problem under uncertainty is straightforward to solve. Second, following Bellman's (1957) principle, we solve a two-period problem, where utility for the second period is just the indirect utility function derived in the first part.

Proceeding accordingly, we now derive the valuation function, or indirect utility, from choosing a consumption sequence that maximizes

$$(4) \quad -E_t \left[\alpha_0 \sum_{s=t}^{\bar{n}} \beta^s \exp(-\rho c_{ns}) \right]$$

subject to the budget constraint

$$(5) \quad E_t \left(\sum_{s=t}^{\bar{n}} \lambda_s c_{ns} \right) \leq \lambda_t e_{nt}$$

This is a consumer choice problem with a linear budget constraint and additively separable utility over commodities, where the objects of choice are claims to contingent commodities. The only nonstandard aspect of the problem, compared with Debreu's (1959: Chap. 7) treatment, for example, is that the commodity space is infinite dimensional. All the contingent claims prices λ_s in periods $s \in \{t, \dots, \bar{n}\}$ help determine the optimal consumption stream. Nevertheless, these preferences fall within the HARA class [as defined by Merton (1971), for example], so only a very small number of securities are required to characterize the optimal financial portfolio. In particular, let p_{tu} denote the price of a bond that, contingent on the history through to date t , pays a unit of consumption in periods t through u .

$$(6) \quad p_{tu} \equiv E_t \left(\sum_{s=t}^u \lambda_s \right)$$

Similarly, let q_{tu} denote the price of a security that pays the random quantity $(\ln \lambda_s - s \ln \beta)$ consumption units in periods t through u :

$$(7) \quad q_{tu} \equiv E_t \left[\sum_{s=t}^u \lambda_s (\ln \lambda_s - s \ln \beta) \right]$$

Taking the first-order condition for the consumer's optimization problem, substituting the Frisch demands back into the multiperiod utility function, and solving the marginal utility of wealth using the (remaining) lifetime budget constraint, we obtain, following Rubinstein (1981), the indirect utility for the manager on his or her retirement.

PROPOSITION 1. *The value function obtained from choosing the consumption sequence to maximizes Equation (4) subject to Equation (5) is*

$$(8) \quad -\alpha_0 p_{t\bar{n}} \exp \left(-\frac{q_{t\bar{n}} + \rho \lambda_t e_{nt}}{p_{t\bar{n}}} \right)$$

The second step of the intertemporal consumption problem described at the beginning of this section is to solve a two-period problem where the functional form of utility in the second period comes from Equation (8), implying that the two-period expected utility is

$$(9) \quad -\alpha_j \beta^t \exp(-\rho c_{nt}) - \alpha_0 E_t \left[p_{t+1, \bar{n}} \exp \left(-\frac{q_{t+1, \bar{n}} + \rho \lambda_{t+1} e_{n, t+1} + \rho \lambda_{t+1} w_{n, t+1}}{p_{t+1, \bar{n}}} \right) \right]$$

while the budget constraint is

$$(10) \quad \lambda_t c_{nt} + E_t(\lambda_{t+1} e_{n, t+1}) \leq \lambda_t e_{nt}$$

This formulation is an application of Bellman's (1957) principle: The optimized value of lifetime utility is equal to a two-stage problem in which one first derives the indirect utility from tomorrow onward and then maximizes over the sum of today's direct utility and the indirect utility from tomorrow onward. If Equation (10) were the only constraint, then the additional wealth generated from the executive compensation received at the beginning of period $t+1$ would be valued at $E_t(\lambda_{t+1} w_{n, t+1})$, and the solution to this problem would be almost indistinguishable from maximizing

$$(11) \quad -\alpha_j \beta^t \exp(-\rho c_{nt}) - \alpha_0 E_t \left[p_{t+1, \bar{n}} \exp \left(-\frac{q_{t+1, \bar{n}} + \rho \lambda_{t+1} e_{n, t+1}}{p_{t+1, \bar{n}}} \right) \right]$$

subject to

$$(12) \quad \lambda_t c_{nt} + E_t(\lambda_{t+1} e_{n, t+1}) \leq \lambda_t e_{nt} + E_t(\lambda_{t+1} w_{n, t+1})$$

The solution to both problems yields identical optimal current consumption allocations and net wealth in the second period, with their common valuation function given by the next proposition.

PROPOSITION 2. *The value function obtained from choosing $(c_{nt}, e_{n,t+1})$ either to maximize Equation (9) subject to Equation (10) or to maximize Equation (11) subject to Equation (12) is*

$$-p_{t\bar{n}} \alpha_j^{\lambda_t / p_{t\bar{n}}} \alpha_0^{(p_{t\bar{n}} - \lambda_t) / p_{t\bar{n}}} \exp \left[- \frac{\rho \lambda_t e_{nt} + \rho E_t(\lambda_{t+1} w_{n,t+1}) + q_{t\bar{n}}}{p_{t\bar{n}}} \right]$$

This proposition establishes that the compensation schedule by itself cannot provide incentives to work hard when markets are complete because it shows that paying a contract of $w_{n,t+1}$ in period $t + 1$ is essentially the same as paying a lump sum of $E_t(\lambda_{t+1} w_{n,t+1})$ in period t , leading the manager to shirk. If the manager is allowed to trade his or her firm’s financial securities anonymously, then no contract enforcing incentive compatibility exists because the manager can and would nullify its incentive effects by trading contingent claims to nullify their incentive effects. An argument by contradiction thus establishes that without public disclosure laws, the optimal contract is to pay the manager to shirk, as described in Proposition 3 below. Therefore, in this model, in contrast to the model of Fudenberg et al. (1990), for example, where the securities market is not sufficiently developed to allow the manager to undo the potential incentive effects of contracting, public disclosure laws are necessary to enforce incentive schemes.

In order to induce hard work, shareholders must be able to affect the manager’s asset portfolio. Because the manager can only affect $x_{n,t+1}$ in period t , which by assumption is an independent random variable, a public disclosure law requiring the manager to announce his or her asset position with his or her own firm suffices to empower shareholders to design contracts that have incentive effects. This is accomplished by announcing a contingent plan at date t that defines the random variable $w_{n,t+1}$ as net compensation, or net changes in the manager’s wealth that will be realized from his or her claims on the firm’s specific assets, that is directly induced by $x_{n,t+1}$ at the beginning of period $t + 1$. Formally, $w_{n,t+1}$ is a real-valued function defined on Ω measurable with respect to the probability space (Ω, F_{t+1}, P) . Therefore, the choice of $w_{n,t+1}$ determines how events that unfold at the very beginning of period $t + 1$ map into compensation subsequently paid to the manager and includes returns on the firm assets that were held over from previous periods by the manager, either voluntarily or involuntarily (because of covenants prohibiting sale before prescribed dates).

In this framework, $w_{n,t+1}$ is announced after the manager’s consumption and savings choices are announced, but despite appearances, the sequencing of these events is unimportant. Suppose, for example, the manager had an opportunity to trade in financial securities after the compensation plan was announced, shareholders knew this opportunity existed, and they observed the returns on the manager’s financial assets at the beginning of the next period before payment for the previous

period’s work was made (another way of embodying the public disclosure condition). Then the shareholders could neutralize any attempt by the manager to hedge against compensation plans entailing uncertain outcomes by picking a compensation plan that fully accounted for his or her retrospectively observed trading strategy.

Because the model assumes that the firm’s size is negligible relative to the economy, $x_{n,t+1}$ has no aggregate effects, which implies that the prices of contingent claims to consumption are unaffected by the manager’s effort level. Therefore, under a public disclosure law of the type described earlier, Equation (9) simplifies to

$$(13) \quad -\alpha_j \beta^t \exp(-\rho c_{nt}) - \alpha_0 E_t(v_{n,t+1} | I_{ntj} = 1) E_t \left[p_{t+1, \bar{n}} \exp\left(-\frac{q_{t+1, \bar{n}} + \rho \lambda_{t+1} e_{t+1, \bar{n}}}{p_{t+1, \bar{n}}}\right) \right]$$

where $v_{n,t+1}$ is a random variable implicitly chosen by shareholders via $w_{n,t+1}$ defined as

$$(14) \quad v_{n,t+1} \equiv \exp\left(\frac{-\rho \lambda_{t+1} w_{n,t+1}}{p_{t+1, \bar{n}}}\right)$$

This random variable has an economics interpretation: The quotient of $p_{t+1, \bar{n}}$ and λ_{t+1} is the current price in period $t + 1$ consumption units for a consumption unit to be received each period from $t + 1$ to \bar{n} . Therefore, $\lambda_{t+1} w_{n,t+1} / p_{t+1, \bar{n}}$ is the amount of consumption that can be permanently increased in each period from $t + 1$ until death at $\bar{n} + 1$ that can be purchased with compensation $w_{n,t+1}$. This implies that $v_{n,t+1}$ is the multiplicative increment in lifetime utility from spending the compensation in that way. Given his or her anticipated effort level, the manager chooses $(c_{nt}, e_{n,t+1})$ to maximize Equation (13) subject to the budget constraint Equation (10). Substituting $\alpha_0 E(v_{n,t+1} | I_{ntj} = 1)$ for α_0 into Equation (11) and $\lambda_t e_{nt}$ for $\lambda_t e_{nt} + E_t(\lambda_{t+1} w_{n,t+1})$ into Equation (12), Proposition 2 now implies that the value function for the manager choosing effort level j is

$$(15) \quad -p_{t\bar{n}} \alpha_j^{\lambda_t / p_{t\bar{n}}} \alpha_0^{(p_{t\bar{n}} - \lambda_t) / p_{t\bar{n}}} \left[E_t(v_{n,t+1} | I_{ntj} = 1) \right]^{(p_{t\bar{n}} - \lambda_t) / p_{t\bar{n}}} \times \exp\left(-\frac{\rho \lambda_t e_{nt} + q_{t\bar{n}}}{p_{t\bar{n}}}\right)$$

4. FEASIBLE CONTRACTS

The Introduction mentioned two constraints the contract must satisfy. The first one, participation, requires shareholders (the principal) to offer the manager (agent) a contract that is at least as attractive as immediate retirement (his or her best alternative). The other constraint, incentive compatibility, only applies if shareholders want the manager to work diligently. Feasible contracts are those satisfying both the participation and incentive compatibility constraints.

The participation constraint is found by comparing the expected utility from leaving the workforce immediately with the expected utility from working one more period and then retiring. It requires the difference to be negative. After factoring out $-p_{i\bar{n}} \exp[-(\rho\lambda_t e_{nt} + q_{i\bar{n}})/p_{i\bar{n}}]$ from Equations (15) and (8), the participation constraint

$$\alpha_0 \geq \alpha_j^{\lambda_t/p_{i\bar{n}}} \alpha_0^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}} [E_t(v_{n,t+1} | I_{ntj} = 1)]^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}}$$

is derived. Dividing both sides by $\alpha_j^{\lambda_t/p_{i\bar{n}}} \alpha_0^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}}$ and raising them to the power $p_{i\bar{n}}/(p_{i\bar{n}} - \lambda_t)$, we obtain

$$(16) \quad (\alpha_0/\alpha_j)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} \geq E_t(v_{n,t+1} | I_{ntj} = 1)$$

The incentive compatibility constraint requires the manager to prefer diligent work over shirking, the benefits of greater expected compensation offsetting the certain nonpecuniary loss of job satisfaction. Now from Equation (15), the difference in working hard versus shirking is proportional to

$$(17) \quad \alpha_1^{\lambda_t/p_{i\bar{n}}} [E_t(v_{n,t+1} | I_{nt1} = 1)]^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}} - \alpha_2^{\lambda_t/p_{i\bar{n}}} [E_t(v_{n,t+1} | I_{nt2} = 1)]^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}}$$

Let $g(x_{nt})$ denote the ratio of the probability density functions from shirking versus working diligently, defined as a positive real-valued function on the support of x_{nt} as the mapping

$$g(x_{nt}) \equiv \frac{f_1(x_{nt})}{f_2(x_{nt})}$$

Substituting $g(x_{n,t+1})$ into $E_t(v_{n,t+1} | I_{nt1} = 1)$, Equation (17) implies that the incentive compatibility constraint is

$$E_t [g(x_{n,t+1}) v_{n,t+1} | I_{nt2} = 1]^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}} \geq (\alpha_2/\alpha_1)^{\lambda_t/p_{i\bar{n}}} E_t(v_{n,t+1} | I_{nt2} = 1)^{(p_{i\bar{n}} - \lambda_t)/p_{i\bar{n}}}$$

which, on factoring and rearranging, simplifies to

$$(18) \quad E_t [v_{n,t+1} g(x_{n,t+1}) - v_{n,t+1} (\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} | I_{nt2} = 1] \geq 0$$

5. COST MINIMIZATION

The (common) objective of the shareholders is straightforward to write down when the firm’s assets comprise an infinitesimal fraction of those in the economy. Shareholders fully diversify all firm-specific uncertainty and collectively behave like a risk-neutral agent, choosing, period by period, the manager’s effort level and offering him or her a feasible contract that minimizes the discounted value of

expected compensation $E_t(w_{n,t+1})$ or, equivalently, from Equation (14), maximizes $E_t(\ln v_{n,t+1})$. To determine whether the firm should offer the manager an incentive-compatible compensation package that elicits diligent work versus a (lower cost) scheme that provides him or her with the nonpecuniary benefit of low effort, shareholders compare the costs and benefits of both schemes.

Suppose, momentarily, that shareholders decided that the manager should shirk. Since he or she is risk-averse and shareholders behave as if they are risk-neutral, it is optimal for them to pay him or her a constant wage that raises his or her wealth just enough to offset the nonpecuniary benefits of retirement. This scheme fully insures him or her against idiosyncratic fluctuations in π_{nt} about π_t . Using the participation constraint Equation (16), we obtain the formula for optimal compensation for the shirking case, denoted $w_{1n,t+1}$.

PROPOSITION 3. *If $l_{1nt} = 1$, then the cost-minimizing feasible contract is*

$$(19) \quad w_{1n,t+1} = \rho^{-1} \left(\frac{p_{t+1,\bar{n}}}{\lambda_{t+1}} \right) \left(\frac{\lambda_t}{p_{t\bar{n}} - \lambda_t} \right) \ln(\alpha_1/\alpha_0)$$

Actually, any contract that pays Equation (19) in expectation is cost-minimizing for $l_{1nt} = 1$. The manager purchases contingent claims to offset any deviation of the contract from Equation (19). Since shareholders anticipate the manager to eliminate all idiosyncratic risk, monitoring his or her financial portfolio is redundant; there is separation between the compensation contract and decisions concerning the financial portfolio. This contrasts with the case of working diligently, i.e., setting $l_{2nt} = 1$.

The timing in the model implies that shareholders observe the asset position of the manager at the beginning of each period before actual compensation is paid out or any assets are traded. This means that shareholders can compensate the manager on the basis of changes in his or her net asset position, i.e., changes in his or her own holdings of the firm, accounting for any trading in the firm's securities undertaken by the manager on the firm's securities. Thus, without loss of generality, the cost-minimizing contract for diligent effort is stated in terms of overall exposure to idiosyncratic fluctuations in the firm's returns.

PROPOSITION 4. *There is a unique, strictly positive solution for η to*

$$(20) \quad \int \left[\eta (\alpha_2/\alpha_1)^{\lambda_t/(p_{t\bar{n}} - \lambda_t)} - \eta g(x) + 1 \right]^{-1} f_2(x) dx = 1$$

Denote this unique solution by η_{nt} . If $l_{2nt} = 1$, then the cost-minimizing feasible contract is

$$(21) \quad w_{2n,t+1} = \rho^{-1} \left(\frac{p_{t+1,\bar{n}}}{\lambda_{t+1}} \right) \left(\frac{\lambda_t}{p_{t\bar{n}} - \lambda_t} \right) \ln(\alpha_2/\alpha_0) + \rho^{-1} \left(\frac{p_{t+1,\bar{n}}}{\lambda_{t+1}} \right) \ln \left[1 + \eta_{nt} (\alpha_2/\alpha_1)^{\lambda_t/(p_{t\bar{n}} - \lambda_t)} - \eta_{nt} g(x_{n,t+1}) \right]$$

Equation (21) has a straightforward interpretation. Two expressions comprise the optimal contract. Comparing Equation (19) with

$$w_{0n,t+1} \equiv \rho^{-1} \left(\frac{p_{t+1,\bar{n}}}{\lambda_{t+1}} \right) \left(\frac{\lambda_t}{p_{t\bar{n}} - \lambda_t} \right) \ln(\alpha_2/\alpha_0)$$

one can immediately deduce that $w_{0n,t+1}$, the first expression in Equation (21), is the amount the manager would be paid if he or she was asked to work diligently and his or her effort could be monitored without cost. The second expression therefore is attributable to moral hazard. As the proof to Proposition 4 shows, η_{nt} is ratio of the Lagrange multipliers for the two constraints in the shareholders' minimization problem. The lower the shadow value of relaxing the incentive compatibility constraint relative to the participation constraint, the smaller is η_{nt} . As η_{nt} approaches 0, this expression disappears altogether. Note, too, that the probability distribution for idiosyncratic first output is more concentrated in the lower part of the support if the manager shirks rather than works diligently. Thus $g(x_{n,t+1})$, the likelihood ratio of the respective probability density functions, tends to decline as $x_{n,t+1}$ increases. In the special case where $g(x_{n,t+1})$ is monotone decreasing, $w_{2n,t+1}$ is increasing in $x_{n,t+1}$ over the whole support $[\psi, \infty)$.

Regardless of whether shareholders require the manager to work hard or shirk, the optimal contract does not expose the manager to general fluctuations in the economy. Mathematically, this is evident from Equations (19) and (21), which show that aggregate conditions in the economy only affect the optimal contract through the terms $p_{t+1,\bar{n}}/\lambda_{t+1}$ and $\lambda_t/(p_{t\bar{n}} - \lambda_t) = (p_{t\bar{n}}/\lambda_t - 1)^{-1}$. The former is the real value of a console bond at the beginning of period $t + 1$ that pays a sure unit of consumption from periods $t + 1$ through \bar{n} , whereas the latter also depends only on the real value of a console bond at the beginning of period t that pays a sure unit of consumption from periods t through \bar{n} . The values of both are computed directly from short- and long-term real interest rates. Therefore, the optimal contract only depends on the contingent claims prices through their effects on interest rates. Intuitively, shareholders have no reason to expose their manager to uncertainty generated by probability distributions over which the manager has no control.

6. TWO EXTENSIONS

Before turning to measures of the importance of moral hazard and the empirical analysis itself, we now briefly discuss a property of the basic model and an extension to it, both of which are motivated by two features within our data set. The first is that most of the executives within the sample held their positions for more than one (annual) period. This stylistic fact begs the question as to whether long-term contracts could induce the same performance at a lower expected cost than short-term contracts. Second, the data we analyze contain information on the top three executives for each of the sampled firms, not just the CEO. In our data, the level of risk exposure to the firm's idiosyncratic income that compensation packages impose seems to depend on the position. This suggests that different jobs within upper-level

management have different characteristics. Accordingly, the first point of this section is to show that the cost-minimizing contract derived in the preceding sections cannot be improved on by entering into a longer-term relationship with the firm. Meaningful long-term contracting may be an equilibrium outcome of a different model of private information, but not this one. And the second point is to demonstrate a simple way of extending this principal-agent model to deal with multiple agents when the expected cost to the principal from an agent deviating from a prescribed Nash equilibrium strategy for all agents depends on the position the agent holds within the firm.

Fudenberg et al. (1990) have analyzed a principal-agent model where the manager's utility from consumption each period is additively separable over time, firms maximize the expected value of discounted income flows, the interest rate is constant, and there is no aggregate uncertainty within the economy. They established that if an optimal long-term contract exists, it can be implemented by a sequence of short-term contracts. Moreover, when utility from current consumption is exponential, wealth and consumption need not be monitored to implement the optimal contract. Here we extend their model to economies where there are aggregate shocks but complete markets by incorporating aggregate uncertainty, thus providing a rationale for the stock market and other financial securities and allowing future interest rates to be stochastic. This difference, however, does not affect their main results, as the next proposition shows. Time additivity in both preferences and technology and the short-term nature of private information are key features in demonstrating that the problem has a recursive structure with complete information at the beginning of each period. These features imply that the optimal long-term contract can be written as a sequence of one-period contracts, whereas the assumption of exponential utility is used to show that monitoring the agent's wealth is redundant.

PROPOSITION 5. The optimal long-term contract can be written as a sequence of short-term contracts in which the expected profit to shareholders every period is zero, the actions and payment plans are identical to those which would be offered in a one-period problem, and the available technology is the same and a manager of the same age retires at the end of each period.

As mentioned earlier, the data list compensation on the top three executives for each of the sampled firms, which we exploit by extending the model to account for a shareholder (principal) contracting with multiple managers (agents). This second modification is accomplished in the following way: Suppose that there are K managers working for firm n at time t , designated $k \in \{1, \dots, K\}$, each of whom determines his or her retirement date and effort level. First, we extend the definition of preferences to be job-specific, replacing $(\alpha_0, \alpha_1, \alpha_2)$ with $(\alpha_{0k}, \alpha_{1k}, \alpha_{2k})$ for each $k \in \{1, \dots, K\}$. Analogous to the single-agent problem, the k th manager's action is characterized by the threuple $l_{nt}^{(k)} \equiv (l_{0ntk}, l_{1ntk}, l_{2ntk})$, where $l_{jntk} \in \{0, 1\}$ for each (j, n, t, k) and $\sum_{j=0}^2 l_{jntk} = 1$. We now define $f(x_{nt} | l_{nt}^{(1)}, \dots, l_{nt}^{(K)})$, the probability

density function for x_{nt} , conditional on the effort levels of the agents, by

$$f(x_{nt} | l_{nt}^{(1)}, \dots, l_{nt}^{(K)}) = \begin{cases} f_2(x_{nt}) & \text{if } \sum_{i=1}^K l_{2ntk} = K \\ f_{1k}(x_{nt}) & \text{if } \sum_{h=1}^K l_{2nth} = K - 1 \text{ and } l_{1ntk} = 1 \\ f_1(x_{nt}) & \text{if } \sum_{k=1}^K l_{2ntk} < K - 1 \end{cases}$$

Let $F_1(x_{nt})$, $F_2(x_{nt})$, and $F_{k1}(x_{nt})$ denote the probability distribution functions, respectively, associated with $f_1(x_{nt})$, $f_2(x_{nt})$, and $f_{1k}(x_{nt})$. We assume a stochastic dominance condition, that the inequalities $F_2(x_{nt}) \leq F_{1k}(x_{nt}) \leq F_1(x_{nt})$ hold for each for all $k \in \{1, \dots, K\}$, and again impose the assumption of a common support (ψ, ∞) .

Given Nash responses by managers (which limit their degree of coordination in thwarting the goals of the shareholders), the contract simplifies to the one-agent case with some minor notational alterations. A sketch of the approach suffices to characterize the main features of the optimal contract. There are only three basic cases to consider: Either everybody should shirk, all but one manager should work diligently, or everybody should work diligently. Following the argument used to prove Proposition 3, the cheapest way of retaining a manager at effort level $l_{jntk} = 1$ is to pay him or her the contract specified in Equation (19) after substituting $\ln(\alpha_{1k}/\alpha_{0k})$ for $\ln(\alpha_1/\alpha_0)$. If shareholders wish the k th manager to work hard in period t , the optimal contract depends on whether some other manager is shirking or not. If not, in a Nash equilibrium, the only potential deviation shareholders must guard against is that one manager will shirk when all the others work diligently. Consequently, the optimal contract for this case merely involves replacing, in Proposition 4, the preference parameters $(\alpha_0, \alpha_1, \alpha_2)$ with $(\alpha_{0k}, \alpha_{1k}, \alpha_{2k})$ and the likelihood ratio $g(x_{nt})$ with $g_{2k}(x_{nt}) \equiv f_{1k}(x_{nt})/f_2(x_{nt})$. If shareholders prescribe that (only) one other manager shirks, say h , then their aim is to draw the firm's idiosyncratic shock from the $F_{h1}(x_{nt})$ probability distribution rather than $F_2(x_{nt})$ or $F_1(x_{nt})$. In this case, the optimal contract for k is to replace the preference parameters $(\alpha_0, \alpha_1, \alpha_2)$ with $(\alpha_{0k}, \alpha_{1k}, \alpha_{2k})$, the likelihood ratio $g(x_{nt})$ with $g_{1h}(x_{nt}) \equiv f_1(x_{nt})/f_{1k}(x_{nt})$, and the probability density function $f_2(x_{nt})$ with $f_{1k}(x_{nt})$ in Proposition 4.

7. THE IMPORTANCE OF MORAL HAZARD

Three different measures of assessing the importance of moral hazard are suggested by this framework. The first measure is the loss shareholders incur from not observing the manager's actions directly. A second measure is the value to the manager of the compensating differential from working diligently versus shirking. Third is the income loss a firm would sustain from signing a contract with a manager to shirk.

The first measure, denoted Δ_{1t} , is the maximum amount shareholders would pay to solve the moral hazard problem. In this framework, Δ_{1t} is the expected difference between the payout under a system of perfect monitoring $w_{0n,t+1}$ less the payout

under the optimal compensation scheme $w_{2n,t+1}$.

$$\begin{aligned} \Delta_{1t} &= E_t(w_{2n,t+1} | l_{2nt} = 1) - E_t(w_{0n,t+1} | l_{2nt} = 1) \\ &= \rho^{-1} E_t(1 + p_{n,t+1}) \int \ln \left[1 + \eta_{nt} (\alpha_2 / \alpha_1)^{\lambda_t / (\rho_{i\bar{n}} - \lambda_t)} - \eta_{nt} g(x) \right] f_2(x) dx \end{aligned}$$

The second line gives an explicit expression for this expected difference, where η_{nt} is defined in Equation (20). This measure is a lower bound on the shareholders' willingness to pay for a perfect monitor, because it is based on asking the manager to perform the same tasks. If, however, the manager's actions could be monitored perfectly, it sounds plausible that shareholders would modify the manager's job description to better exploit the monitoring technology for the benefit of the firm.

The next measure we investigate evaluates the manager's nonpecuniary benefits from pursuing his or her own goals within the firm. Denoted Δ_{2t} , it is the expected difference required to keep utility constant at the reservation level between managerial compensation for diligent work under perfect monitoring and compensation for shirking.

$$\begin{aligned} \Delta_{2t} &= E_t(w_{0n,t+1} | l_{2nt} = 1) - E_t(w_{1n,t+1} | l_{1nt} = 1) \\ &= \rho^{-1} E_t[(1 + p_{n,t+1})] \lambda_t (p_{i\bar{n}} - \lambda_t)^{-1} \ln(\alpha_1 / \alpha_2) \end{aligned}$$

Clearly, Δ_{2t} is increasing in the relative attractiveness of shirking to hard work. Again, the measure has a meaningful interpretation only within the context of the equilibrium contract itself, because the nature of moral hazard critically depends on the job description for l_{2nt} . Our framework assumes that the nonpecuniary aspects of work satisfaction are one-dimensional, entering utility through the scalar α . Therefore, l_{1nt} can be interpreted as those activities undertaken at work yielding the lowest α in absolute magnitude, conditional on shareholders requesting the manager undertake l_{2nt} . The second measure compares these two activities and ignores everything else.

The third measure is the expected gross output loss to the firm switching from the distribution of abnormal returns for diligent work to the distribution for shirking, i.e., the difference between the expected output to the plant from the manager pursuing the firm's goals versus his or her own, before netting out expected managerial compensation. In our framework, this is

$$\begin{aligned} \Delta_{3t} &= E_t(a_{nt} x_{n,t+1} | l_{2nt} = 1) - E_t(a_{nt} x_{n,t+1} | l_{1nt} = 1) \\ &= a_{nt} \int x [f_2(x) - f_1(x)] dx \end{aligned}$$

Against this output reduction is a savings in managerial compensation coming from two terms, the cost of having a perfect monitor and the cost of inducing the manager to work diligently when a perfect monitor is in place. Subtracting from Δ_{3t} the sum of Δ_{1t} and Δ_{2t} , we obtain the net income loss a firm would sustain from signing a shirking contract with a manager, in which l_{1nt} is selected and recom-

pensed appropriately. Substituting the expressions defining Δ_{1t} through Δ_{3t} , we obtain

$$\begin{aligned} \Delta_t &\equiv \Delta_{3t} - \Delta_{2t} - \Delta_{1t} \\ &= a_{nt} [E_t(x_{n,t+1} | l_{2nt} = 1) - E_t(x_{n,t+1} | l_{1nt} = 1)] \\ &\quad - E_t(w_{2n,t+1} | l_{2nt} = 1) - E_t(w_{0n,t+1} | l_{1nt} = 1) \\ &= a_{nt} [E_t(\pi_{n,t+1} | l_{2nt} = 1) - E_t(\pi_{n,t+1} | l_{1nt} = 1)] \end{aligned}$$

where the last line follows from the definition of $z_{n,t+1} \equiv w_{n,t+1}/a_{nt}$ and Equation (3). As our discussion in the preceding section established, this net amount represents the value of enforcing public disclosure laws that make contracts requiring the manager to work hard feasible. From the perspective of public policy, this measure is arguably the most relevant. However, it also suffers from the same deficiency as Δ_{1t} because of its focus on just one alternative. If, for example, laws against the invasion of privacy rendered it impossible for managers to credibly inform shareholders about their asset position with the firm so that feasible contracts inducing hard work no longer exist, then presumably the job statements of managers would be rewritten to improve on the shirking contract. In this respect, Δ_t overstates the value of being able to write and enforce a compensation plan that is contingent on the firm's idiosyncratic returns.

8. THE DATA

Our empirical work used data on compensation packages for the top three executives of 34 firms for the period 1948 through 1977, originally collected by Masson (1971) and extended by Antle and Smith (1985, 1986), as well as time-series data on stock market returns, interest, and inflation rates. Appendix B describes how the variables used in this study were constructed. Our description of the data in relation to its applicability to this model is in four parts. First, we provide a cross-sectional summary of the data set, and then we present some of its time-series features.

8.1. *Cross-Sectional Summary Statistics.* There are 306 executives in the survey, of which exactly one-third are CEOs. Of the 306 executives, 210 had left their firm before 1977 when the panel ended, and for the purposes of our study, these observations are defined as separations. The average length of tenure is 5.6 years, with a standard deviation of 5.4 years (high because of several 1-year stints).

Table 1 provides some summary statistics of compensation within the three industries. Many of them were approaching retirement, averaging 57.1 years old (with a standard deviation 7.9 years), so roughly 8 years away from the then statutory retirement age of 65. While the sample average age of CEOs by industry is higher than the average for non-CEOs, the difference is insignificant in each case. The table shows that CEOs are paid more than non-CEOs but that their compensation exhibits more variability, thus possibly suggesting that CEOs have stronger incen-

tives to engage in non-value-maximizing activities. Note, too, that compensation appears to differ across the three industries, with aerospace looking more lucrative than the other two. Relative to its mean, the variation in salary and bonus is less than that of the other components to total compensation. This feature suggests that measures of compensation that exclude managerial income from holding and granting financial securities whose value is affected by the firm's abnormal returns are unlikely to capture the main performance-enhancing characteristics of the compensation package.

TABLE 1
CROSS-SECTIONAL INFORMATION ON EXECUTIVE COMPENSATION IN 1967 us\$
(STANDARD DEVIATIONS IN PARENTHESES)

		Aerospace	Chemicals	Electronics	All
After-tax compensation	All	126,822 (410,590)	86,094 (493,072)	76,958 (428,977)	93,395 (468,892)
	CEO	144,731 (475,720)	120,618 (602,724)	96,688 (610,090)	121,840 (579,963)
	Non-CEO	118,211 (375,687)	69,283 (428,969)	64,522 (257,306)	79,082 (400,985)
Pretax salary and bonus	All	136,408 (61,319)	121,786 (63,111)	82,223 (34,787)	119,594 (67,757)
	CEO	175,965 (58,025)	154,324 (70,338)	106,522 (38,316)	151,388 (67,263)
	Non-CEO	117,386 (53,305)	105,943 (52,440)	66,907 (21,073)	13,596 (51,904)
After-tax value of options granted	All	16,821 (58,726)	11,759 (46,206)	16,947 (65,051)	13,505 (51,874)
	CEO	19,463 (61,331)	14,525 (55,171)	19,721 (77,674)	16,322 (60,354)
	Non-CEO	15,551 (57,474)	10,412 (41,100)	15,200 (55,780)	12,088 (46,987)
Return on stock held	All	8,797 (294,955)	-5,601 (477,451)	(2,872) (384,681)	-2,229 (411,763)
	CEO	6,790 (306,196)	5,835 (531,743)	(5,681) (564,701)	4,264 (499,470)
	Non-CEO	9,763 (289,817)	-11,169 (400,172)	(1,102) 203,096	-5,497 (359,753)
Pretax value of stock bonus	All	0	273 (1,852)	518 (4,821)	249 (2,341)
	CEO	0	239 (1,496)	0	154 (1,205)
	Non-CEO	0	290 (2,003)	845 (6,139)	296 (2,705)
Return on options held	All	17,530 (163,653)	1,444 (114,107)	3,507 (92,296)	5,080 (123,808)
	CEO	18,002 (218,544)	5,523 (156,450)	1,386 (110,359)	7,419 (165,162)
	Non-CEO	17,303 (129,605)	-548 (86,293)	4,845 79,080	3,903 (96,559)
Age	All	55.5 (7.1)	57.7 (8.1)	56.7 (8.4)	57.1 (7.9)
	CEO	57.6 (6.9)	58.0 (7.7)	57.5 (7.0)	57.9 (7.5)
	Non-CEO	54.5 (6.9)	57.6 (8.5)	56.2 (9.1)	56.8 (8.4)

TABLE 2
CROSS-SECTIONAL INFORMATION ON FIRMS (SALES, EQUITY, AND ASSETS ARE IN MILLIONS OF 1967 US\$;
STANDARD DEVIATIONS IN PARENTHESES)

	Aerospace	Chemicals	Electronics	All
Abnormal returns	0.069 (0.368)	0.001 (0.239)	0.052 (0.395)	0.022 (0.291)
Return on assets	0.087 (0.044)	0.149 (0.066)	0.105 (0.086)	0.129 (0.071)
Sales	107.59 (187.54)	71.94 (113.67)	18.73 (30.55)	71.06 (127.68)
Debt-to-equity ratio	1.00 (0.50)	0.33 (0.13)	0.68 (0.55)	0.52 (0.43)
Common equity	21.25 (39.08)	39.30 (63.30)	9.30 (20.79)	31.10 (55.47)
TOTAL ASSETS	40.60 (74.49)	31.71 (48.17)	8.62 (13.32)	29.99 (52.36)

Turning to statistics that describe the firms these executives manage, Table 2 shows that the firms are not highly leveraged, the average debt equity ratio ranging from 0.3 in chemicals to 1.0 in the aerospace industry. Overall, the average of abnormal returns across all three industries is less than one-tenth of its standard deviation. The hypothesis that $E(x_t) = 0$ cannot be rejected by these data in any of the three industries at standard significance levels. This provides some empirical justification for ignoring risk premia. On the other hand, the test itself is masked by survivorship bias induced by the sample selection procedure, so we are wary of making too much of the empirical regularity that even within each industry, the firms seem representative of the market portfolio. Further evidence about abnormal returns from the cross section is displayed by the smoothed frequencies in Figures 1 through 3. Note that the sample standard deviations of abnormal returns differ across industries (again suggesting that the probability distribution of abnormal returns differs by industry) and that the respective sample distributions are skewed toward the right tail.

8.2. *Time-Series Summary Statistics.* The next two tables summarize the time-series properties of our data set. Table 3 exhibits real interest rates, market returns, and abnormal returns by industry for the data period. The three decades were characterized by low interest rates (the one period real rate averaging 0.4 percent), and the market performed reasonably well (with an average value weighted return of 8.7 percent). The electronics industry stocks in our sample performed worse than the market over half the time (57 percent), compared with 43 percent for the aerospace industry and 46 percent for chemical industry.

Time-series averages and standard deviations of compensation components (enumerated in 1967 US\$) are given in Table 4. There is no noticeable time trend to total compensation within this 30-year phase. The standard deviation of executive compensation is typically about twice the sample mean. Yearly averages across the 34 firms vary quite significantly, and total compensation averaged across the 34 firms is not positive every year. For example, average compensation was \$326,920 in 1959 but negative \$161,610 in 1961. Salary and bonus appear to increase over time, with a low

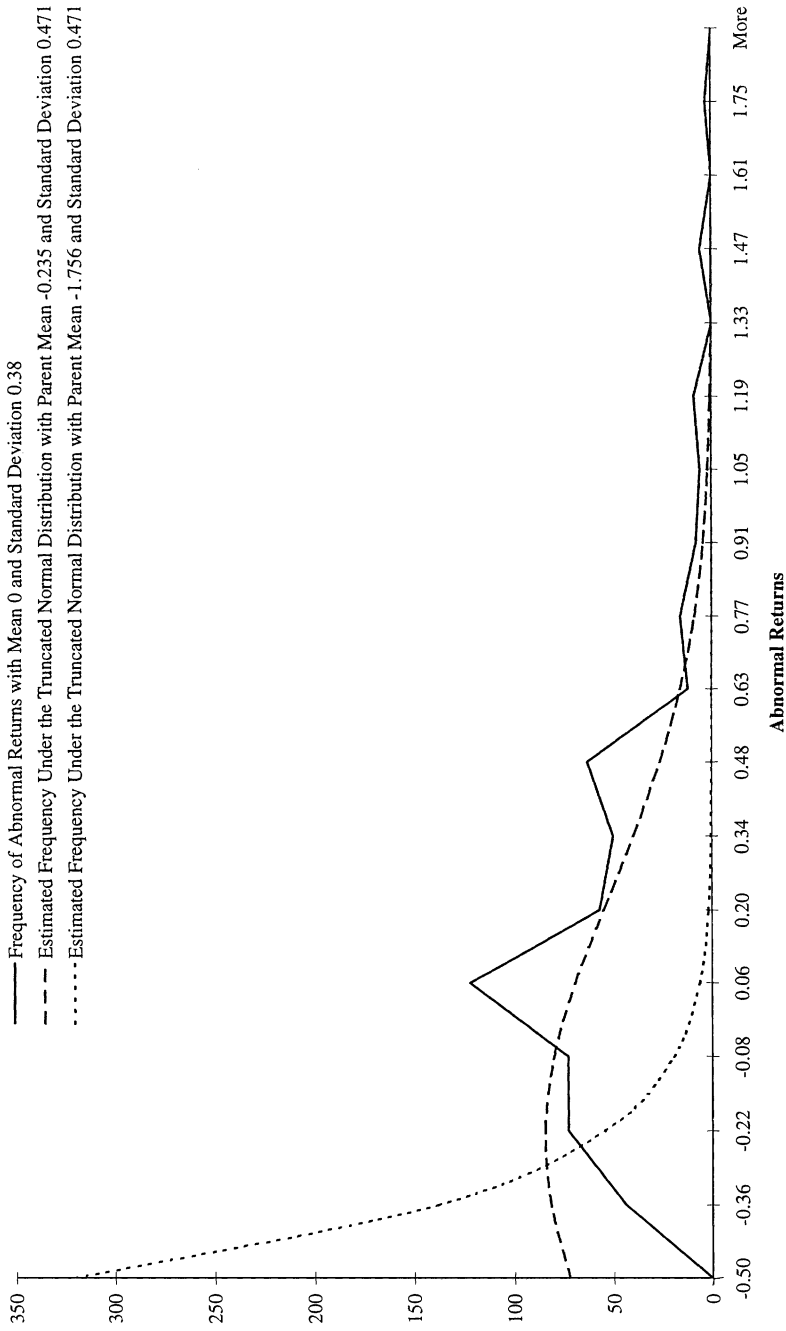


FIGURE 1
AEROSPACE INDUSTRY ABNORMAL RETURNS

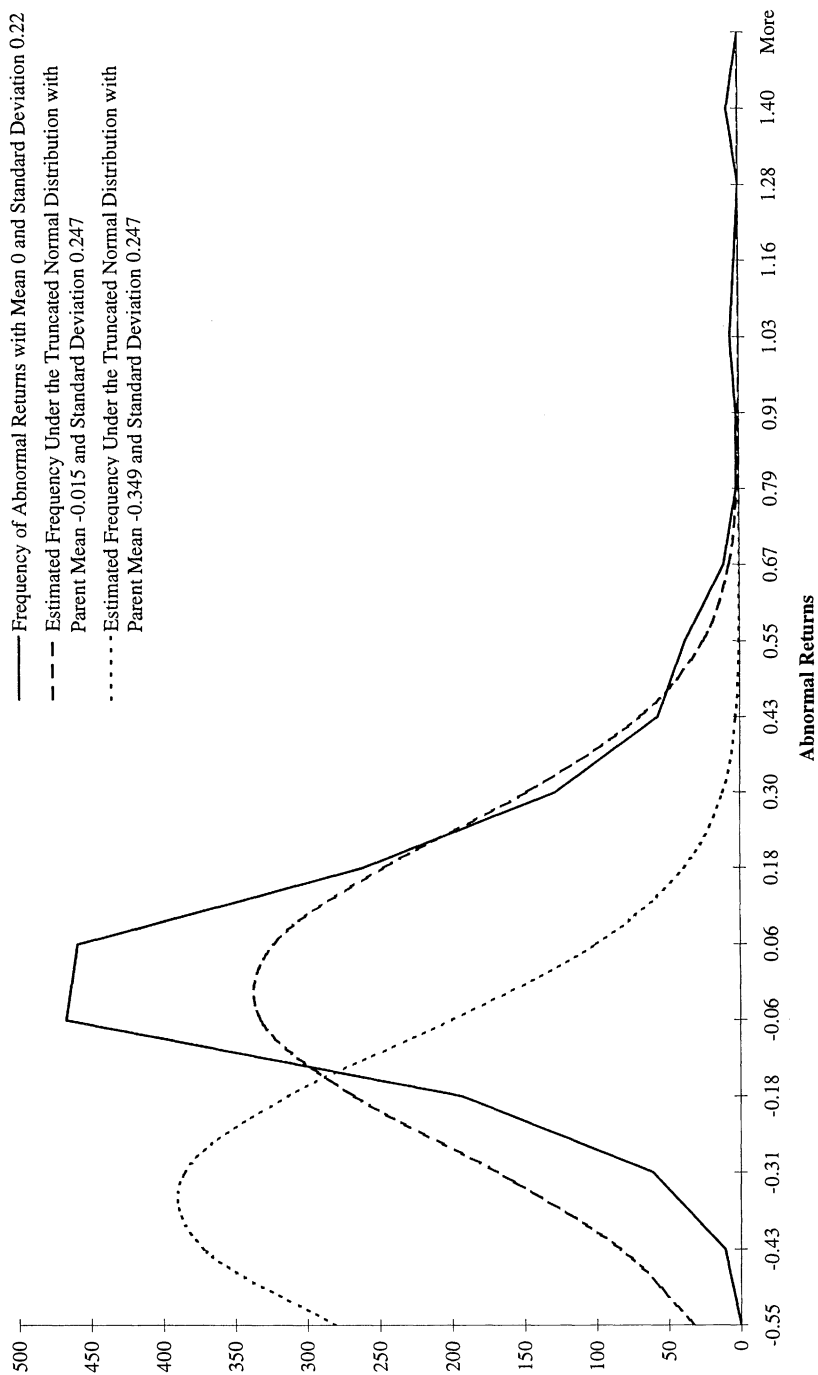


FIGURE 2
CHEMICALS INDUSTRY ABNORMAL RETURNS

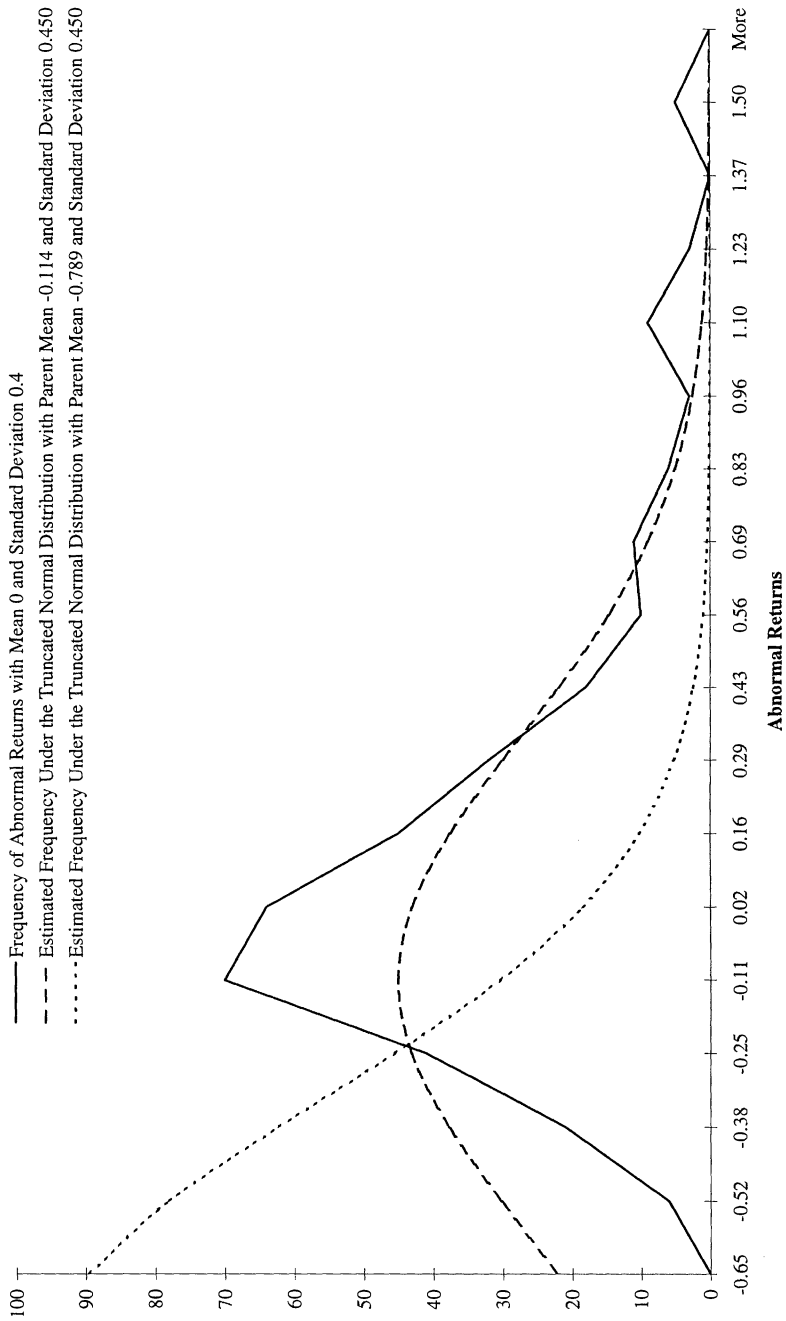


FIGURE 3

ELECTRONICS INDUSTRY ABNORMAL RETURNS

TABLE 3
 TIME-SERIES INFORMATION ON INTEREST RATES AND ABNORMAL RETURNS
 (STANDARD DEVIATIONS IN PARENTHESES)

Year	Inflation Rate	Interest Rate	Value Weighted Return on Stocks	Abnormal Returns			
				All	Aerospace	Chemicals	Electronics
1948	0.030	-0.019	0.000	0.062 (0.137)	0.195 (0.283)	-0.086 (0.155)	0.052 (0.159)
1949	-0.021	0.032	0.227	-0.019 (0.202)	-0.037 (0.179)	0.000 (0.141)	-0.071 (0.314)
1950	0.059	-0.044	0.226	-0.017 (0.174)	0.383 (0.236)	0.069 (0.206)	0.193 (0.286)
1951	0.060	-0.042	0.140	0.040 (0.239)	-0.116 (0.121)	0.087 (0.178)	-0.016 (0.481)
1952	0.008	0.010	0.125	-0.053 (0.174)	0.132 (0.138)	-0.109 (0.138)	-0.019 (0.233)
1953	0.007	0.012	-0.004	0.023 (0.193)	0.053 (0.345)	0.010 (0.151)	0.051 (0.244)
1954	-0.007	0.017	0.514	-0.022 (0.200)	0.025	-0.059 (0.178)	0.135 (0.269)
1955	0.004	0.014	0.248	-0.013 (0.215)	-0.004 (0.384)	0.021 (0.168)	-0.175 (0.189)
1956	0.030	-0.003	0.052	0.014 (0.220)	0.126 (0.321)	0.019 (0.198)	-0.118 (0.196)
1957	0.029	0.004	-0.131	-0.030 (0.172)	-0.101 (0.145)	-0.038 (0.170)	0.081 (0.196)
1958	-0.018	0.001	0.423	-0.025 (0.214)	-0.039 (0.087)	-0.010 (0.211)	-0.109 (0.360)
1959	0.017	0.017	0.112	0.077 (0.269)	-0.004 (0.447)	0.063 (0.204)	0.241 (0.297)
1960	0.014	0.016	-0.005	-0.019 (0.231)	0.073 (0.280)	-0.047 (0.191)	0.057 (0.364)
1961	0.007	0.017	0.266	-0.107 (0.234)	0.040 (0.338)	-0.105 (0.188)	-0.207 (0.191)
1962	0.013	0.015	-0.111	-0.071 (0.146)	0.030 (0.161)	-0.055 (0.093)	-0.230 (0.186)
1963	0.016	0.015	0.195	-0.065 (0.200)	-3.490 (0.133)	0.011 (0.126)	-0.07 (0.250)
1964	0.010	0.026	0.153	-0.035 (0.238)	0.080 (0.484)	-0.020 (0.149)	-0.205 (0.091)
1965	0.019	0.021	0.119	0.122 (0.293)	0.363 (0.329)	0.059 (0.188)	0.103 (0.524)
1966	0.035	0.015	-0.119	0.014 (0.240)	0.028 (0.072)	-0.057 (0.187)	0.270 (0.366)
1967	0.030	0.013	0.231	-0.016 (0.126)	-0.009 (0.125)	-0.014 (0.136)	-0.037 (0.090)
1968	0.047	0.007	0.077	-0.097 (0.171)	-0.227 (0.186)	0.036 (0.140)	-0.180 (0.179)
1969	0.062	0.007	-0.151	-0.144 (0.175)	-0.289 (0.153)	-0.101 (0.177)	-0.151 (0.106)
1970	0.056	0.010	-0.041	-0.104 (0.207)	-2.310 (0.204)	0.01 (0.150)	-3.68 (0.072)
1971	0.033	0.011	0.122	0.050 (0.280)	0.245 (0.365)	0.008 (0.267)	-0.023 (0.010)
1972	0.034	0.007	0.138	-0.029 (0.209)	-0.006 (0.145)	-0.046 (0.231)	0.014 (0.226)
1973	0.087	-0.013	-0.236	-0.024 (0.270)	-2.460 (0.146)	0.088 (0.268)	-0.19 (0.009)

TABLE 3
 TIME-SERIES INFORMATION ON INTEREST RATES AND ABNORMAL RETURNS
 (STANDARD DEVIATIONS IN PARENTHESES) (CONTINUED)

Year	Inflation Rate	Interest Rate	Value Weighted Return on Stocks	Abnormal Returns			
				All	Aerospace	Chemicals	Electronics
1974	0.126	-0.039	-0.350	0.138 (0.378)	0.131 (0.305)	0.227 (0.400)	-0.210 (0.127)
1975	0.069	-0.009	0.288	0.218 (0.265)	0.250 (0.321)	0.188 (0.265)	0.290 (0.251)
1976	0.050	0.001	0.202	0.089 (0.330)	0.310 (0.220)	-0.280 (0.294)	0.530 (0.197)
1977	0.067	-0.012	-0.108	0.081 (0.280)	0.237 (0.270)	0.007 (0.289)	0.162 (0.277)

standard deviation compared with the standard deviation of total compensation. The average change in wealth associated with changes in the value of options and stocks held is substantial and has a very high variance. It is these changes that account for the reported negative compensation. Again, Table 4 suggests that ignoring the options and stock holdings of managers would seriously bias the analysis against the substantiating claim that managerial compensation results from contracts written to overcome the effects of private information.

8.3. *Retirement.* We mentioned that 210 of the 306 executives leave the sample before 1977 (or the last year for which the firm survives in the data). Although our model does not account for separation, it is consistent with an extended framework in which managers decide at the beginning of a period whether to work one or more periods or not. Then the optimal compensation package is formed for those who elect to stay on the job. An alternative view is that involuntary separation may form part of an optimal contract for a more sophisticated production and information technology than the one developed here. Unfortunately, the data contain little information on these separations. Possible explanations include death, retirement, dismissal for some misconduct, or voluntary quits. If death and voluntary retirement explained separations, the probability of a separation would increase with the age of an executive. However, if the separations observed were dismissals, we might expect the probability of a separation to decline with firm performance.

As a diagnostic check of our model specification, we estimated the conditional probability of a separation on several variables of interest using a probit. Recall l_{0nt} to be an indicator variable for retirement, where $l_{0nt} = 1$ if the n th executive retires in period t and 0 otherwise. We assume

$$l_{0nt} = \begin{cases} 1 & \text{if } \kappa_{nt}B + \epsilon_{nt} > 0 \\ 0 & \text{if } \kappa_{nt}B + \epsilon_{nt} \leq 0 \end{cases}$$

where ϵ_{nt} is identically and independently distributed as a standard normal random variate, κ_{nt} is a row vector of covariates of interest, and B is a conformable column vector of coefficients to be estimated.

TABLE 4
 TIME-SERIES INFORMATION ON EXECUTIVE COMPENSATION IN THOUSANDS OF 1967 US\$
 (STANDARD DEVIATIONS IN PARENTHESES)

	After-Tax Compensation Salary	Pretax Salary & Bonus	Value of Options Granted	Return on Stock Held	Value of Stock Bonus	Return on Options Held	Average Age
1948	39.98 (233.31)	93.89 (50.74)	2.09 (148.35)	-22.84 (226.40)	0.00 0.00	0.13 (1.32)	55 (8)
1949	63.66 (77.93)	100.59 (53.48)	1.00 (81.19)	5.87 (69.08)	0.13 (0.75)	0.37 (10.76)	55 (8)
1950	126.57 (177.55)	109.60 (53.78)	2.36 (147.85)	49.47 (171.90)	0.18 (1.03)	4.77 (20.05)	56 (7)
1951	179.15 (401.99)	107.55 (51.24)	15.94 (804.44)	102.50 (391.41)	0.12 (0.67)	6.20 (35.98)	56 (7)
1952	-31.78 (321.04)	103.87 (50.33)	2.73 (109.63)	-83.43 (313.93)	0.00 0.00	-5.69 (49.89)	57 (7)
1953	71.79 (193.25)	108.29 (56.58)	3.39 (158.42)	10.87 (176.78)	0.00 0.00	-0.09 (51.12)	57 (7)
1954	89.41 (738.13)	109.42 (54.66)	11.68 (377.18)	-38.74 (637.34)	0.00 0.00	48.55 (231.46)	57 (8)
1955	66.87 (266.95)	112.68 (58.23)	6.15 (393.52)	-14.65 (221.88)	0.13 (0.93)	10.65 (87.34)	58 (8)
1956	92.66 (421.21)	118.10 (64.02)	15.32 (570.01)	8.69 (380.37)	0.18 (1.24)	1.86 (69.68)	58 (8)
1957	27.97 (301.98)	113.75 (59.47)	7.19 (321.22)	-43.38 (262.97)	0.04 (0.42)	-9.43 (91.88)	58 (8)
1958	164.85 (431.71)	107.27 (57.88)	27.19 (819.33)	41.68 (363.72)	0.05 (0.32)	39.25 (109.47)	58 (8)
1959	325.92 (730.54)	113.04 (62.64)	14.37 (397.21)	207.83 (632.87)	0.07 (0.46)	35.25 (224.95)	57 (9)
1960	14.10 (660.68)	111.52 (66.14)	7.33 (194.76)	-73.33 (601.75)	0.09 (0.62)	16.52 (159.28)	57 (8)
1961	-161.61 (627.18)	112.72 (63.15)	10.07 (287.64)	-214.50 (538.65)	0.15 (0.88)	-19.58 (192.54)	56 (8)
1962	-71.71 (313.32)	109.88 (53.79)	17.97 (389.35)	-102.87 (272.83)	0.24 (1.39)	-47.90 (98.26)	56 (7)
1963	48.10 (183.80)	113.92 (56.39)	3.33 (153.12)	-24.04 (145.71)	0.22 (1.29)	7.02 (58.53)	56 (7)
1964	88.00 (292.06)	120.98 (61.61)	9.49 (417.13)	-30.22 (251.29)	0.23 (1.32)	12.05 (71.24)	56 (6)
1965	370.70 (781.54)	125.61 (60.94)	13.58 (630.55)	188.80 (722.51)	0.29 (1.62)	80.35 (193.35)	58 (7)
1966	58.89 (534.93)	131.23 (60.65)	10.12 (526.93)	-14.32 (462.25)	0.00 0.00	-29.74 (121.41)	60 (9)
1967	353.54 (667.37)	133.92 (65.90)	22.74 (1025.60)	182.16 (564.56)	0.00 0.00	58.57 (216.83)	57 (7)
1968	-5.94 (545.01)	135.08 (63.68)	19.70 (600.01)	-88.28 (509.98)	0.76 (3.84)	-22.88 (91.14)	58 (8)
1969	-1.10 371.93	133.71 (63.28)	20.69 (473.30)	-76.98 (331.52)	1.11 (5.15)	-40.80 (73.88)	58 (7)
1970	-33.92 (312.30)	116.48 (57.83)	18.89 (484.17)	-125.43 (334.73)	0.51 (2.46)	-25.43 (51.77)	58 (9)
1971	132.30 (453.43)	117.14 (56.89)	25.26 (747.68)	22.29 (407.36)	0.40 (1.96)	7.92 (67.91)	56 (8)
1972	162.10 (310.39)	133.21 (63.56)	33.27 1026.24	15.63 (223.28)	0.52 (2.51)	17.84 (67.95)	56 (8)
1973	16.68 (581.38)	137.44 (63.00)	15.42 (419.24)	-67.35 (556.76)	0.27 (1.59)	-33.21 (117.56)	56 (8)
1974	131.98 (355.79)	149.90 (69.71)	13.80 (343.35)	7.73 (290.01)	1.43 (8.39)	-11.26 (84.62)	57 (7)

TABLE 4
 TIME-SERIES INFORMATION ON EXECUTIVE COMPENSATION IN THOUSANDS OF 1967 US\$
 (STANDARD DEVIATIONS IN PARENTHESES) (CONTINUED)

	After-Tax Compensation Salary	Pretax Salary & Bonus	Value of Options Granted	Return on Stock Held	Value of Stock Bonus	Return on Options Held	Average Age
1975	293.06 (483.31)	142.74 (66.63)	16.22 (356.10)	117.65 (431.84)	0.00 0.00	58.13 (143.54)	57 (9)
1976	211.14 (342.73)	153.66 (71.24)	36.44 (909.67)	32.30 (205.93)	1.13 (6.62)	32.78 (113.67)	58 (9)
1977	73.93 (314.31)	168.50 (72.75)	26.96 (517.33)	-12.46 (215.05)	0.00 (0.01)	-69.19 (165.39)	65 (12)

Our findings in Table 5 show that neither abnormal stock returns nor returns on assets are significant variables in explaining separations. This casts doubt on the hypothesis that dismissals are an important method of disciplining managers. Both age and a retirement-age dummy are significant, suggesting that retirement can explain many of the separations. Rank within the firm is positively related, implying that subordinates are more likely to leave than CEOs. Finally, although the index coefficient on firm performance is insignificant, executive compensation is negatively correlated with retirement, a counterintuitive finding, at least within the theoretical framework we postulate, where compensation and retirement are unrelated. Perhaps this contemporaneous correlation is attributable to tax considerations, which may provide incentives to defer compensation until after retirement, but a definitive answer to this question must await further research.

8.4. *Sensitivity of Executive Compensation to Changes in Shareholder Wealth.* Because the population of firms and executives in this data set differs in many ways from more recent data sets that have been used, we ran several regressions to

TABLE 5
 RESULTS FROM PROBIT MODEL: PROBABILITY OF SEPARATIONS
 STANDARD ERRORS IN PARENTHESES

Variable	Coefficient
Age	0.029 (0.005)
Retirement age dummy	0.961 (0.116)
Abnormal return	-0.075 (0.156)
Return on assets	0.533 (0.539)
Position	0.247 (0.084)
Compensation	-2.01E-07 9.14E-08
Constant	-3.5 (0.325)

NOTE: Number of observations = 2536; log-likelihood = -649.31.

compare the empirical relationship between changes in various measures of executive compensation and firm value with those of other studies. For the purposes of comparability, the variables described in the footnote to Table 6 were constructed to roughly conform with those defined in Jensen and Murphy (1990), with two noteworthy exceptions. First, changes in both managerial income and shareholder wealth from holding the market portfolio were excluded from our analysis, meaning that only the effects of changes in shareholder wealth attributable to firm-specific abnormal returns were investigated. Second, historical (nominal minus inflation) real interest rates were used to compute present values.

Ordinary least-squares regressions were run for all the executives in our sample and for CEOs alone, with several definitions of executive income. Table 6 displays the regression results. Almost all the coefficients are significant. Following the methodology proposed by Jensen and Murphy (1990), a measure of the sensitivity of executive compensation to shareholder wealth is to predict the change in compensation from a \$1000 increase in shareholder wealth with the estimated linear model. Focusing, for argument's sake, on the whole sample, the second part of column 2 shows that after accounting for the lagged effect, the change in an executive's cash and bonus is $\$0.07 - \$0.06 = \$0.02$. This number increases to $\$9.47 - \$7.20 = \$2.27$ after accounting for the value of stocks and options granted, improved retirement benefits, and related items (see Column 3). Then, from the second part of column 4, if we treat the change in salary and bonus as permanent, the effect rises further to $\$11.03 - \$1.19 = \$9.84$. The return on stocks held contributes another \$4.00, and the return on options held contributes \$2.73. Aggregating over all the items, we find that an increase in shareholder wealth from favorable abnormal returns by \$1000 raises the most encompassing measure of executive compensation by $\$9.84 + \$4.00 + \$2.73 = \16.57 , an order of magnitude larger than Jensen and Murphy's estimate of \$3.25 but hardly enough to reverse their basic point that, collectively, shareholders have far more at stake than the executives who run their firms.

9. IDENTIFICATION AND ESTIMATION

The data just described on compensation to the three top executives, abnormal firm returns, and security prices were used to identify and estimate the truncated normal parameterization of the model. They are ordered by $n \in \{0, 1, \dots, N\}$, each observation referring to one of the 3 executive positions in one of the 34 firms in one of the 30 years. The intuitive basis for identification in this framework stems from the idea that graphing fluctuations in realized compensation against the firm's abnormal returns trace out the compensation schedule and that the distribution of abnormal returns itself can be estimated nonparametrically from realizations over time. Therefore, the curvature of the compensation schedule is informative about expected firm losses if the manager shirks, the extra utility the manager would derive from shirking, and his or her attitude toward risk. Indeed, the compensation contract derived in the preceding sections maps the prices of observed securities and firm returns into compensation received by the manager. To avoid stochastic singularity, we postulated a measurement error that induces a discrepancy between the observed compensation and actual compensation $w_{n,t+1}$. Accordingly, for each $n \in \{1, \dots, N\}$,

TABLE 6
OLS ESTIMATES OF PAY-PERFORMANCE SENSITIVITY COMPONENTS (DEPENDENT VARIABLES IN THOUSANDS OF 1967 US\$;
STANDARD ERRORS IN PARENTHESES)

Independent Variable	1		2		3		4				5		6		
	Change in Pretax Salary and Bonus		Change in Pretax Salary and Bonus		Change in Total Pay ²		Change in the Sum of Total Pay and PV (Salary and Bonus) ³				Return on Stock Held		Return on Options Held		
	CEO	All	CEO	All	CEO	All	CEO	All	All	CEO	All	CEO	All	CEO	All
Intercept	3.577 (0.895)	3.577 (0.4251)	3.560 (0.8883)	3.639 (0.4297)	-8.956 (35.690)	-5.296 (22.050)	-40.940 (66.160)	9.536 (18.069)	5.690 (16.796)	-3.700 (8.020)	8.700 (5.688)	6.423 (2.439)			
Change in shareholders' wealth from abnormal returns ¹	0.00011 (0.00004)	0.00011 (0.00002)	0.00004 (0.00008)	0.00007 (0.00003)	0.01672 (0.0032)	0.00947 (0.00167)	0.01730 (0.004)	0.01103 (0.0014)	0.00376 (0.00085)	0.00400 (0.0004)	0.00426 (0.00051)	0.00253 (0.00018)			
Lagged change in shareholders' wealth from abnormal returns															
			-0.0001 (0.00005)	-0.000058 (0.00002)	-0.00720 (0.00181)	-0.00720 (0.0011)	0.00189 (0.0058)	-0.00119 (0.00099)			0.00008 (0.00028)	0.00020 (0.00012)			

¹ The variable change in shareholders' wealth from abnormal returns was constructed by multiplying abnormal returns by common equity.

² The variable total pay is after-tax compensation minus return on stock held and return on options held.

³ The variable PV (salary and bonus) is the present value of an annuity equal to the change in pretax salary and bonus starting the following year and ending the year in which the executive turns 65, computed using the interest rates from Table 3.

define the observed compensation, denoted $\tilde{w}_{n,t+1}$, as

$$\tilde{w}_{n,t+1} \equiv w_{n,t+1} + \varepsilon_{n,t+1}$$

where $\varepsilon_{n,t+1}$ is an independent and identically distributed normal random variable with mean 0 and variance $2\lambda_{t+1}^{-1} p_{t+1, \bar{n}} \rho^{-2} \xi$, and ξ is a normalizing parameter to be estimated.

In this application it is useful to incorporate differences between the three sectors that affect the probability distributions for abnormal returns and to distinguish between the preferences of the top manager and the other two executives. To this end, let $s \in \{1, \dots, S\}$ enumerate the industrial sector and $k \in \{1, \dots, K\}$ label the executive's position within the firm's hierarchy. For each observation $n \in \{1, \dots, N\}$, the indicator variables d_{1ns} and d_{2nk} are now, respectively, defined as

$$d_{1ns} = \begin{cases} 1 & \text{if the } n\text{th observation occurs in the } s\text{th industrial sector} \\ 0 & \text{otherwise} \end{cases}$$

and

$$d_{2nk} = \begin{cases} 1 & \text{if the } n\text{th observation occurs in the } k\text{th executive position} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, denote by $f_{2s}(x)$ the probability density function for the s th sector if all S executives in the firm work diligently, let $f_{1s}(x)$ denote the probability density function for the s th sector if all but one of the K executives in the firm work diligently, and add another subscript to the preference parameters so that the vector $(\alpha_{0k}, \alpha_{1k}, \alpha_{2k}, \rho, \beta)$ now characterizes the (time-additive exponential) preferences of the k th executive. This parameterization is not fully interactive because not all the SK combinations induced by partitioning the observations by a graph of the coordinate pairs (d_{1ns}, d_{2nk}) are treated in a distinctive manner. For example, we are assuming that if only one executive shirks, the probability distribution of abnormal returns does not depend on which executive shirks. Similarly, the preferences of executives are assumed to be unrelated to their industry of employment. While other forms of observed heterogeneity could be added without complicating the estimation procedure, our preliminary empirical investigations pursuing this line of enquiry did not appear promising.

The parameter estimates were obtained in two steps. First, $f_{2s}(x)$ was estimated for $s \in \{1, \dots, S\}$ using data on abnormal returns to the firm. Then estimates of the other parameters were found using data on managerial compensation and firm returns by constructing orthogonality conditions from the participation and incentive compatibility constraints, as well as the managerial compensation schedule. These steps are now discussed in detail.

9.1. *The Distribution of Abnormal Returns.* Our empirical application parameterizes $f_{js}(x)$ for $s \in \{1, \dots, S\}$ and $j \in \{1, 2\}$. More specifically, we assume that for each $j \in \{1, 2\}$, the firm's abnormal returns are distributed as a truncated normal random

variable with support bounded below by ψ_s . Thus in each sector $s \in \{1, \dots, S\}$,

$$(22) \quad f_{js}(x) = \left[\Phi \left(\frac{\mu_{js} - \psi_s}{\sigma_s} \right) \sigma_s \sqrt{2\pi} \right]^{-1} \exp \left[\frac{-(x - \mu_{js})^2}{2\sigma_s^2} \right]$$

where Φ is the standard normal distribution function, and (μ_{js}, σ_s) denotes the mean and variance of the parent normal distribution associated with $l_{ji} = 1$. Under this parameterization, $g_s(x)$ takes the functional form

$$(23) \quad g_s(x) = \frac{\Phi[(\mu_{2s} - \psi_s)/\sigma_s]}{\Phi[(\mu_{1s} - \psi_s)/\sigma_s]} \exp \left[\frac{\mu_{2s}^2 - \mu_{1s}^2 - 2x(\mu_{2s} - \mu_{1s})}{2\sigma_s^2} \right]$$

There are four reasons for choosing this flexible three-parameter probability distribution. The first is empirical. Figures 1 through 3 suggest that a probability distribution with a monotonic density function would not capture some of the main features and that the distribution is skewed to the right. In principle, extending this parameterization to permit dependence of the standard deviation of the parent normal distributions on the managers' effort levels poses no conceptual problems, but in practice, it proved fruitless exercise with this data set. Second, it is well known that if $f_{js}(x)$ are normal distributions for each $j \in \{1, 2\}$ that differ only in mean, there are feasible contracts arbitrarily close to the first best allocation that almost eliminate the moral hazard problem. The third reason is the flexibility that a truncated normal gives in modeling the shirking distribution. Depending on where the mean of its parent normal distribution lies, the probability density function of the shirking distribution could be either monotonic or almost bell shaped. Fourth, this parameterization implies that $g(x)$ is decreasing in x throughout its whole range, since

$$\frac{\partial g(x)}{\partial x} = \frac{(\mu_1 - \mu_2)g(x)}{\sigma^2} < 0$$

This inequality follows from the fact that if $E(x_{n,t+1} | l_{1nt} = 1)$ is less than $E(x_{n,t+1} | l_{2nt} = 1)$, then $\mu_1 \leq \mu_2$. Thus the model satisfies a monotone likelihood ratio property that guarantees that compensation is increasing in abnormal returns, a feature that, as we have shown, receives some limited support from the pay-performance regressions.⁶

It is tempting to interpret the threshold ψ_s and the asymmetry of the probability density function characterizing abnormal returns as stemming from a notion of bankruptcy or takeover. Presumably, firms cannot incur large losses and remain solvent without undergoing drastic changes to their administrative structure. However, while this interpretation may seem appealing, it should be treated cautiously. Bankruptcy is not analyzed within our theoretical model, and to the extent that managerial actions push the firm into bankruptcy, our model is incomplete and

⁶ Having justified the truncated normal parameterization, we nevertheless acknowledge that one could estimate the distribution of abnormal returns nonparametrically.

underestimates the cost of moral hazard. Furthermore, the survivor firms comprising our data set would constitute a poor empirical foundation for estimating and testing a theory that formally addressed the relationship between managerial behavior and bankruptcy.

The preceding section showed that our data cannot reject the simple hypothesis that

$$(24) \quad E(x_{n,t+1} | I_{2nt} = 1) = 0$$

We exploit this condition in estimation below. Following Maddala (1983:365), for example, the mean of a normal distribution truncated from below at ψ is

$$(25) \quad E(x_{n,t+1} | I_{2nt} = 1) = \mu_{2s} + \frac{\sigma_s \Phi'[(\psi_s - \mu_{2s})/\sigma_s]}{\Phi[(\mu_{2s} - \psi_s)/\sigma_s]}$$

Combining Equations (24) and (25), we have

$$(26) \quad \mu_{2s} = - \frac{\sigma_s \Phi'[(\psi_s - \mu_{2s})/\sigma_s]}{\Phi[(\mu_{2s} - \psi_s)/\sigma_s]} \equiv \mu(\psi_s, \sigma_s)$$

where the mapping $\mu(\psi_s, \sigma_s)$ from $\Re \times \Re^+$ to \Re is implicitly defined by the first line in Equation (26). Therefore, the distribution of abnormal returns for the three-sector case is characterized by two vectors $\theta_1 \equiv (\psi_1, \psi_2, \psi_3)'$ and $\theta_2 \equiv (\sigma_1, \sigma_2, \sigma_3)'$.

We now turn to the estimation $\theta_1^{(0)}$, the true value of θ_1 . For each $s \in \{1, \dots, S\}$, a consistent estimator for $\theta_1^{(0)}$ is $\theta_1^{(N)} \equiv (\psi_1^{(N)}, \psi_2^{(N)}, \psi_3^{(N)})'$, where

$$(27) \quad \psi_s^{(N)} \equiv \min_{n \in \{1, \dots, N\}} \{d_{1ns} x_{nt}\}$$

Admittedly this estimator is very sensitive to misspecification from unaccounted measurement error. For example, instead of observing x_{nt} for each $n \in \{1, \dots, N\}$, suppose the econometrician only observes the sum $\tilde{x}_{n,t+1} \equiv x_{n,t+1} + \varepsilon'_{n,t+1}$, where $\varepsilon'_{n,t+1}$ denotes an independent and identically distributed random variable with a probability density function defined on the closed interval $[\underline{\varepsilon}, \bar{\varepsilon}]$ attributable to measurement error in the series on abnormal returns. Then it is easy to prove that $\psi_s^{(N)}$ converges in probability to $\psi_s^{(0)} + \underline{\varepsilon}'$. However, while the use of this estimator is therefore hard to justify in household panels, where sample respondents typically have limited incentives to respond truthfully and reliably to questions asked by the interviewer, data on the financial returns are, by comparison, subject to audit procedures that all but guarantee their accuracy. For this reason, using the estimator for $\psi_s^{(0)}$ defined in Equation (27), which ignores the possibility of measurement error in abnormal returns, seems reasonable in this context.

The estimation of $f_2(x)$ is completed by setting $j = 2$ in Equation (22) and maximizing the log-likelihood function for N observations in θ_2 . Conditional on $\theta_1^{(0)}$,

the ML estimator $\theta_2^{(ML)}$ is thus defined as the minimizer of

$$\begin{aligned}
 L_N(\theta_1^{(0)}, \theta_2) &= \sum_{n=1}^N \sum_{s=1}^3 d_{1ns} \\
 (28) \quad &\times \left\{ \ln \sigma_s + \ln \Phi \left[\frac{\mu(\psi_s^{(0)}, \sigma_s) - \psi_s^{(0)}}{\sigma_s} \right] + \frac{[x_{n,t+1} - \mu(\psi_s^{(0)}, \sigma_s)]^2}{2\sigma_s^2} \right\}
 \end{aligned}$$

in θ_2 . Since the true value $\theta_1^{(0)}$ is unknown, obtaining $\theta_2^{(ML)}$ is infeasible, so we estimate the true value $\theta_2^{(0)}$ by $\theta_2^{(N)}$, found by minimizing Equation (28) after substituting in $\theta_1^{(N)}$ for $\theta_1^{(0)}$. Because convergence in probability of $\theta_1^{(N)}$ to $\theta_1^{(0)}$ occurs more rapidly than \sqrt{N} , substituting $\theta_1^{(N)}$ for $\theta_1^{(0)}$ throughout the other parts of the empirical analysis does not affect the asymptotic properties of the remaining parameter estimates. In particular, it is straightforward to show that

$$\sqrt{N}(\theta_2^{(N)} - \theta_2^{(ML)}) = o_p(1)$$

which immediately implies that the asymptotic distributional properties of $\theta_2^{(N)}$ and $\theta_2^{(ML)}$ are identical.⁷ Finally, estimates for $(\theta_1^{(0)}, \theta_2^{(0)})$ are used to obtain an estimate of $\mu_{2s}^{(N)} \equiv \mu(\psi_s^{(N)}, \sigma_s^{(N)})$ for each $s \in \{1, \dots, S\}$ from Equation (26).

9.2. *The Remaining Parameters.* This leaves the parameters characterizing managerial preferences and the firm’s distribution of abnormal returns from shirking to estimate. Because utility levels are unobserved, α_0 , α_1 , and α_2 are only identified up to a factor of proportionality. For this reason, α_0 was normalized to unity. Because the optimal contract is independent of the subjective discount factor, β cannot be identified either. Consequently, the only remaining parameters to be estimated are

$$\theta_3 \equiv (\mu_{11}, \mu_{12}, \mu_{13}, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \rho, \xi)'$$

Denote the true value of θ_3 by $\theta_3^{(0)}$. It was estimated using a generalized methods of moments (GMM) procedure by constructing orthogonality conditions derived from the compensation Equation (21), in which η_{nt} is treated as a function $p_{t\bar{n}}$ implicitly given by Equation (20); from the participation constraint Equation (16); and from the incentive compatibility condition Equation (18).

The first vector of orthogonality conditions is formed from

$$h_{1n}(\theta) = \exp \left[\frac{-\lambda_{t+1}(\rho \tilde{w}_{n,t+1} + \xi)}{p_{t+1,\bar{n}}} \right] - \alpha_2^{\lambda_k / (\lambda_t - p_{t\bar{n}})}$$

⁷ Donald and Paarsch (1993) recently have examined the properties of such estimators in the context of estimating auction models.

where $\theta \equiv (\theta_1, \theta_2, \theta_3)$ denotes all the unknown parameters in the model. The distributional assumptions about $\varepsilon_{n,t+1}$ imply

$$E \left\{ \exp \left[\frac{-\lambda_{t+1}(\rho \tilde{w}_{n,t+1} + \xi)}{p_{t+1, \bar{n}}} \right] \middle| w_{n,t+1}, p_{t+1, \bar{n}} \right\} = \exp \left(\frac{-\rho \lambda_{t+1} w_{n,t+1}}{p_{t+1, \bar{n}}} \right) = U_{n,t+1}$$

and hence, from the participation constraint Equation (16) (which is met with equality under the optimal contract),

$$E \left[\sum_{k=1}^2 d_{2nk} h_{1n}(\theta^{(0)}) \right] = 0$$

where $\theta^{(0)} \equiv (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)})$ denotes the true value of θ . The incentive compatibility condition (Equation 18) provides the basis for a second vector of orthogonality conditions. Let

$$h_{2n}(\theta) = \exp \left[\frac{-\lambda_{t+1}(\rho \tilde{w}_{n,t+1} + \xi)}{p_{t+1, \bar{n}}} \right] \left[g_s(x_{n,t+1}; \theta) - \left(\frac{\alpha_{2k}}{\alpha_{1k}} \right)^{\lambda_t / (\rho_{t\bar{n}} - \lambda_t)} \right]$$

where $g_s(x; \theta) \equiv g_s(x)$, defined in Equation (23). Again note that if Equation (18) is met with equality under the optimal contract (which is easy to show), then

$$E \left[\sum_{s=1}^3 \sum_{k=1}^2 d_{1ns} d_{2nk} h_{2n}(\theta^{(0)}) \right] = 0$$

Last, the orthogonality conditions based on the compensation equation are derived from Equation (21). Defining

$$h_{3n}(\theta) = \alpha_{2k}^{\lambda_t / (\lambda_t - \rho_{t\bar{n}})} \exp \left[\frac{\lambda_{t+1}(\rho \tilde{w}_{n,t+1} + \xi)}{p_{t+1, \bar{n}}} \right] - 1 - \eta_{nt} \left[\left(\frac{\alpha_{2k}}{\alpha_{1k}} \right)^{\lambda_t / (\rho_{t\bar{n}} - \lambda_t)} - g(x_{n,t+1}; \theta) \right]$$

the same reasoning yields

$$E \left[\sum_{s=1}^3 \sum_{k=1}^2 d_{1ns} d_{2nk} h_{3n}(\theta^{(0)}) \right] = 0$$

Notice that in contrast to the other two equations, measurement error plays a critical role in avoiding stochastic singularities in this equation.⁸ Accordingly, define

⁸ Absent measurement error $\sum_{s=1}^3 \sum_{k=1}^2 d_{1ns} d_{2nk} h_{3n}(\theta^{(0)}) = 0$ for all $n \in \{1, 2, \dots, N\}$.

the 3×1 vector $h_n(\theta)$ as

$$(29) \quad h_n(\theta) = \begin{pmatrix} \sum_{k=1}^2 d_{2nk} h_{1n}(\theta) \\ \sum_{s=1}^3 \sum_{k=1}^2 d_{1ns} d_{2nk} h_{2n}(\theta) \\ \sum_{s=1}^3 \sum_{k=1}^2 d_{1ns} d_{2nk} h_{3n}(\theta) \end{pmatrix}$$

Let y_n denote a $q \times 1$ instrument vector formed from variables orthogonal to $h_n(\theta^{(0)})$, and A_N is a $q \times q$ matrix that converges to some constant nonsingular matrix A . An estimator for $\theta_3^{(0)}$, denoted $\theta_3^{(N)}$, is found by minimizing

$$(30) \quad M_N(\theta_1^{(N)}, \theta_2^{(N)}, \theta_3) \equiv \left[\frac{1}{N} \sum_{n=1}^N y_n \otimes h_n(\theta_1^{(N)}, \theta_2^{(N)}, \theta_3) \right]' A_N \left[\frac{1}{N} \sum_{n=1}^N y_n \otimes h_n(\theta_1^{(N)}, \theta_2^{(N)}, \theta_3) \right]$$

with respect to θ_3 subject to Equation (20), which defines η_{nt} . Note that since η_{nt} is the solution to a fixed-point problem, which must be solved for each value of the parameter vector θ_3 to evaluate the econometric criterion function, this estimator is an example of a nested fixed-point algorithm; in the literature on managerial compensation, Ferrall and Shearer (1999) also use a nested fixed-point algorithm to obtain their estimates.⁹

If, instead of $(\theta_1^{(N)}, \theta_2^{(N)})$, the true value $(\theta_1^{(0)}, \theta_2^{(0)})$ had been substituted into Equation (30) and minimized with respect to θ_3 subject to Equation (20), the resulting estimator, denoted $\theta_3^{(GMM)}$, would be \sqrt{N} -consistent and asymptotically normal, as discussed in Hansen (1982), for example. Given the instrument set y_n , the most efficient choice of A_N in this case is any matrix that converges in probability to

$$A_0 \equiv \left\{ E \left[(y_n \otimes h_n(\theta^{(0)}))(y_n \otimes h_n(\theta^{(0)}))' \right] \right\}^{-1}$$

Proving that $\theta_3^{(N)}$ is consistent amounts to demonstrating that since the differences $(\theta_1^{(N)} - \theta_1^{(0)})$ and $(\theta_2^{(N)} - \theta_2^{(0)})$ are both $o_p(1)$, then $(\theta_2^{(N)} - \theta_2^{(GMM)})$ is too, using the fact that the criterion function (Equation 30) is continuous and the parameter space is assumed compact. Moreover, $\theta_1^{(N)}$ converges at a rate exceeding \sqrt{N} , so preestimating $\theta_1^{(0)}$ with $\theta_1^{(N)}$ affects neither the rate of convergence of $\theta_2^{(N)}$ nor its asymptotic covariance matrix. Consequently, the asymptotic error induced from the prior estimation of the distribution for abnormal returns is solely due to $\sqrt{N}(\theta_2^{(ML)} - \theta_2^{(0)})$. Given a weighting matrix A_N converging in probability to A_0 , and noting that the score of the likelihood function is itself a moment condition, it then follows [e.g., from Newey (1984) and using the fact that the information matrix equals the expected value of the outer product of the scores] that the asymptotic covariance matrix for $\theta_3^{(N)}$ is

$$B_1^{-1} \left[B_2 + B_3(B_4^{-1} + B_4^{-1}B_5 + B_5'B_4^{-1})B_3' \right] B_1^{-1'}$$

⁹ An earlier literature on the structural estimation of dynamic discrete choice models also uses this estimation method. See Miller (1997) for a recent survey.

where

$$\begin{aligned}
 B_1 &= E \left[\frac{\partial^2 M_1(\theta^{(0)})}{\partial \theta_3 \partial \theta'_3} \right] \\
 B_2 &= E \left[\frac{\partial M_1(\theta^{(0)})}{\partial \theta_3} \frac{\partial M_1(\theta^{(0)})'}{\partial \theta'_3} \right] \\
 B_3 &= E \left[\frac{\partial^2 M_1(\theta^{(0)})}{\partial \theta_3 \partial \theta'_2} \right] \\
 B_4 &= -E \left[\frac{\partial L_1^2(\theta_1^{(0)}, \theta_2^{(0)})}{\partial \theta_2 \partial \theta'_2} \right] \\
 B_5 &= E \left[\frac{\partial L_1(\theta_1^{(0)}, \theta_2^{(0)})}{\partial \theta_2} \frac{\partial M_1(\theta^{(0)})'}{\partial \theta_3} \right]
 \end{aligned}$$

In this study, the instruments were constructed from a combination of financial and accounting measures of firm performance and executive compensation. Specifically, the return on total assets, the accounting return on equity, the stock return, earnings per share, the debt-to-equity ratio, and executive compensation itself, all lagged one period (and defined in Appendix B), as well as a constant, were used to form the seven-dimensional y_n vector to yield a total of 21 orthogonality conditions on multiplication by $h_n(\theta)$.

10. STRUCTURAL PARAMETER ESTIMATES

A summary of the main findings appears in the Introduction. We note at the outset that the overidentifying restrictions associated with the orthogonality conditions formed from Equation (29) are not rejected. The test statistic for the sample criterion function (Equation 30) is 2.149. Under the null hypothesis, the statistic is asymptotically distributed χ^2 with 12 degrees of freedom (since we use 21 sample moments to estimate 9 parameters), implying a significance level of 98 percent.

Table 7 reports the parameter estimates and their estimated asymptotic errors. The top half of the table relates to the production side. From Equation (26), our estimates of μ_{2s} are negative because $E(x_{n,t+1} | l_{2nt} = 1) = 0$, and the parent normal distribution is truncated from below by the estimated values of $\psi_s^{(0)}$. Consistent with the model, our estimates of μ_{2s} exceed those for μ_{1s} in all three industries, although not all the differences are significant. In the aerospace and electronics industries, the estimated mean of the parent normal distributions for shirking $\mu_{11}^{(1)}$ and $\mu_{12}^{(1)}$ are less than their respective truncation points $\psi_1^{(1)}$ and $\psi_2^{(1)}$. Therefore, the estimated probability density function associated with the shirking distribution monotonically declines in output in these two industries, in contrast to the chemicals industry. The estimated standard deviations of the parent normal distributions are quite precise, with the chemicals industry displaying significantly less variability in returns than the other two. In a supplementary analysis not reported here, we could not obtain significant differences between the variances of the parent shirking distributions and their analogues for working diligently.

Figures 1 through 3 illustrate the estimated probability density functions for each industry conditional on effort level. We calculated $E(x_{n,t+1} | I_{1nt} = 1)$ from the parent normal distribution and obtained estimates of -0.197 , -0.348 and -0.306 for the aerospace, chemicals, and electronics industries, respectively. These figures imply that $E(x_{n,t+1} | I_{2nt} = 1) \geq E(x_{n,t+1} | I_{1nt} = 1)$, as required by our framework, and show that output would fall quite substantially if managers pursued their own interests to the detriment of their firms. Moreover, the standard deviations of the estimated truncated distributions for hard work match perfectly to the sample standard deviations listed in the tables to four significant figures. This constitutes evidence that imposing the truncated normal parameterization does not seriously bias estimated functions of the first two moments for the probability distributions of abnormal returns under hard work, whatever their true functional form.

Estimates of the managers' preference parameters are in the bottom half of Table 7. None of the signs contradict the underlying premises of our model. Our estimates of α_2/α_0 , the parameter measuring preference for diligence relative to retirement, are significantly greater than one for both types of executive officer, implying that working diligently is more distasteful than retiring. Noting that our estimate of the ratio α_2/α_0 is less for the CEO position than for the non-CEO position, this result shows that the job of CEO is more desirable than that of subordinate. Our estimates of α_2/α_1 , the parameter measuring preference for diligence relative to retirement, are greater than one for both classes of executives, implying that managers prefer to pursue their own goals rather than work in shareholders' interests. This difference is only significant for the CEO. Our finding that α_2/α_1 is significantly higher for the CEO than for the non-CEO suggests that the greater responsibilities of the top

TABLE 7
ESTIMATES OF STRUCTURAL PARAMETERS

Parameter	Description	Industry/Executive	Estimate	Standard Error
μ_1	Percent mean return from shirking	Aerospace	-1.756	0.004
		Chemicals	-0.349	0.035
		Electronics	-0.789	0.006
μ_2	Percent mean return from diligence	Aerospace	-0.235	
		Chemicals	-0.015	
		Electronics	-0.114	
σ	Percent standard deviation of return	Aerospace	0.471	0.040
		Chemicals	0.247	0.010
		Electronics	0.450	0.036
ψ	Treshold minimum return	Aerospace	-0.480	
		Chemicals	-0.497	
		Electronics	-0.606	
ρ	Risk tolerance parameter		0.107	0.080
ξ	Variance associated with measurement error		1.481	0.018
α_2/α_0	Preference for diligence relative to retiring	CEO	1.071	0.016
		Non CEO	1.137	0.630
α_2/α_1	Preference for diligence relative to shirking	CEO	1.172	0.014
		Non CEO	1.002	0.010

position provide more tempting opportunities to act against the firm's interests than subordinate positions in the organizational hierarchy, which apparently permit their occupants less discretion over the choice of work activities. Indeed, the nonpecuniary benefits to the CEO from shirking are so large that he or she would prefer to pursue his or her own goals within the firm to retirement. (Our estimate of the ratio α_1/α_0 is 0.914.) However, the discretionary opportunities to non-CEOs are much less attractive. (The corresponding number is 1.135, implying that they prefer retirement.) Finally, the estimated concavity parameter ρ implies that marginal utility is declining in consumption (as required by the theory), although it is not precisely estimated. By way of comparison, the findings from structural estimates obtained from panel studies of household consumption data yield are ambiguous. For example, the estimates of Altug and Miller (1990) imply that utility is concave increasing in male leisure but convex increasing in food consumption, whereas those of Miller and Sieg (1997) imply that household utility is concave increasing in both male leisure and housing consumption.

To interpret the economic implications of the structural parameter estimates, Table 8 reports our estimates of the several measures of moral hazard discussed at the end of Section 3 by industry and executive type (as appropriate) where the security price p_{nt} (which appears in all three measures) is calculated for a manager at the mean age of 57 in 1967, the last year all firms are in the sample. Recall that Δ_{1t} is the expected difference between the optimal contract motivating high effort and the fixed wage to be paid for high effort if it could be observed, a compensating differential for bearing risk to satisfy the incentive compatibility constraint. This is reported by industry because variation in compensation ultimately depends on $g(x_{nt})$ (the ratio of the industry-specific probability density functions for the respective effort levels), as well as by executive type, since the incentive compatibility condition depends on α_2/α_1 (the value of working diligently versus shirking that is specific to the position of the executive). As the table shows, the differences between executive type dominate differences attributable to industry specifics. While the shadow value for directly observing activities of the CEO is in the neighborhood of \$200,000 (in 1967 prices), shareholders would not even be willing to pay \$3000 to have his or her subordinates perfectly monitored.

The second measure of the importance of moral hazard is Δ_{2t} , the additional compensation needed to motivate high effort in the absence of private information.

TABLE 8
COST OF MORAL HAZARD (MEAN AGED EXECUTIVE IN 1967 US\$)

Measure	Industry	Executive	Cost
Δ_1	Aerospace	CEO	186,689
		Non-CEO	2,370
	Chemicals	CEO	232,966
		Non-CEO	2,680
	Electronics	CEO	173,643
		Non-CEO	2,327
Δ_2		CEO	259,181
		Non-CEO	3,272
Δ_3	Aerospace		263,283,500
	Chemicals		85,355,000
	Electronics		104,222,000

Since this measure only depends on the position of the manager, not on the distribution of firm returns, only two estimates are reported for our parameterization. Reflecting the parameter estimates reported in Table 7, illegitimate perks to the CEO would be substantial (but for the offsetting effects of the compensation package that induces his or her loyalty to the firm). In contrast, the non-CEO is almost indifferent to working diligently versus shirking, so the value of the nonpecuniary benefits from shirking is very modest to him or her. We note that the magnitudes of Δ_{1t} and Δ_{2t} are quite similar. In other words, a manager's benefit from shirking roughly equals the premium a firm must pay to align incentives absent monitoring yet given managerial risk aversion by varying compensation with output.

Summing Δ_{1t} and Δ_{2t} , we obtain the extra cost of getting managers to work hard rather than shirk. Even in the chemical industry, where the divergence between managerial and shareholder goals is most pronounced, this cost still barely exceeds \$0.5 million (to properly motivate all three managers). The benefits to the firm from such a move are given by Δ_{3t} , the expected increased output from superior managerial performance. The numbers in the bottom panel of Table 8 are computed from the estimated expected returns under the shirking distribution (we reported in our discussion of Table 6) for the median firm in each industry by market value as of December 31, 1967. Ranging between \$83 million and \$263 million, these benefits dwarf the costs in each of the three industries studied. From the shareholders' perspective, motivating the manager to act in the interests of the firm, through the asset position he or she is required to take, is cheap compared with the substantial losses we estimate a firm would bear from an implicit contract that encourages managers to pursue their own goals on company time.

11. CONCLUSION

This article was motivated by the observation that managers appear to hold financial assets in their own firms that are not warranted by diversification arguments. From there, we showed how a model of private information based on moral hazard could be embedded within an otherwise competitive economy to demonstrate that the optimal contract would produce this qualitative result. Parameterizing the utility function of managers and the probability distribution of returns to firms (which depend on the manager's actions), the model was then estimated from data on managerial compensation (appropriately measured to take account of their asset positions) and stock returns of the firms they manage (i.e., net of the movement in the market portfolio). Our empirical findings imply that if these series are generated by the principal-agent model we develop, then moral hazard is indeed quite an important empirical phenomenon. Although managers would not make huge utility gains from privately acting against the firm's interests, shareholders would not be prepared to pay huge sums to eliminate the moral hazard. Despite the substantial losses incurred if managers deviate from the actions shareholders prefer, the incentive compatibility constraints are not that costly for shareholders to impose. These results are qualitatively similar to Haubrich (1994), which is somewhat surprising because his model has a different structure than ours and his empirical methodology, calibration, contrasts with our statistical approach. Our results are also comparable with those of Ferrall and Shearer (1999), who find that there are

substantial gains from implementing an optimal contract over a linear bonus system. We find that shareholders would incur huge losses from ignoring moral hazard when setting the manager's compensation scheme.

The sharp interpretation our structural model offers should be balanced against some of the features it ignores. First of all, there is an important sense in which our framework understates the divergence of interests between management and shareholders. Many actions managers would like to take, such as absconding with the firm's assets, are outlawed by criminal and civil law; these actions are not discouraged by the form of the contract but by the penal code. Our model focuses only on those actions which are not deterred by the law itself or, for that matter, anything that might be *ex post* observable.

Second, the wealth of managers only plays a limited role in this model because our parameterization of current utility assumes absolute risk aversion. This simplification is driven by two concerns: Panel data on the manager's wealth that is not tied to his or her own firm are, so far as we know, simply not available, and it is unclear to us whether stockholders have this kind of information either. If the latter point is correct, relaxing the assumption of exponential utility over current consumption would compound the moral hazard problem with private information about the type of manager, i.e., his or her unobserved wealth (at each period). As we have seen, under exponential utility, the optimal contract does not require consumption or total wealth to be observed, is not feasible unless there are public disclosure laws that allow shareholders to see what bets the manager is taking against the firm's idiosyncratic income (effectively prohibiting the manager from doing just that), and cannot be improved on even if total wealth is observed. In reality, the scope of public disclosure laws is limited, not requiring the managers to disclose everything about their consumption and savings patterns, an institutional feature that is compatible with the dual assumptions of exponential utility and pure moral hazard. Thus, if the assumption of exponential utility were relaxed, and if the model avoided the complex issues raised by incomplete information about player type, then the optimal contract would require shareholders to observe total wealth at the beginning of each period (in addition to securities whose payoffs were tied to the abnormal returns of the firm). Not that the limited scope of public disclosure laws provides empirical support for the assumption of exponential utility. After all, in a world where theft occurs, there are other unrelated reasons for privacy.

There are other simplifying features of the model that are used in deriving the result that the optimal long-term contract can be implemented via a sequence of short-term contracts. While the return during the period the action is taken is stochastically determined by the manager's actions that period, actions taken in previous periods do not directly affect it. Moreover, because the principal, i.e., shareholders, are assumed to be value-maximizers (in this case because shareholders face complete markets and the firm is infinitesimal relative to the size of the value-weighted portfolio of all firms), there is only a limited role for dynamic considerations. Other motivations for dynamic behavior, such as human capital accumulation and job matching, are also ignored.¹⁰ In defense of our approach, it is

¹⁰ Gibbons and Murphy (1992) analyze lifetime managerial compensation schemes that arise from recognizing these factors.

worth noting our findings that the overidentifying restrictions are not rejected. These tests provide some confidence that ignoring the factors mentioned above (as well as others we have not raised) is not seriously biasing our estimates (many of which are significant and show that the considerations dealt with here are quantitatively important), although they are not guaranteed to be powerful against all alternative hypotheses of interest. Acknowledging their limited value, we therefore conclude that our study provides a benchmark for further structural estimation work in this area and simultaneously complements the many empirical studies of managerial compensation that, unlike ours, are not explicitly based on economic models of private information.

APPENDIX A: PROOFS

PROPOSITION 1. *The value function obtained from choosing the consumption sequence to maximize Equation (4) subject to Equation (5) is*

$$-\alpha_0 p_{t\bar{n}} \exp\left(-\frac{q_{t\bar{n}} + \rho \lambda_t e_{nt}}{p_{t\bar{n}}}\right)$$

PROOF. By admitting negative consumption choices, we guarantee an interior solution. The first-order condition is

$$\rho \alpha_0 \beta^s \exp(-\rho c_{ns}^o) = \lambda_s \delta$$

where δ is the Lagrange multiplier associated with Equation (5). Substituting the implied Frisch demands into the utility function yields

$$-\rho^{-1} \delta E_t \left(\sum_{s=t}^{\bar{n}} \lambda_s \right) = -\rho^{-1} \delta p_{t\bar{n}}$$

The multiplier δ is solved by using the fact that the budget constraint is binding. Making c_{ns}^o the subject of the first-order condition, we obtain

$$(A.1) \quad c_{ns}^o = \rho^{-1} (\ln \rho \alpha_0 \beta^s - \ln \lambda_s - \ln \delta)$$

Multiplying by the respective Frisch demands (Equation A.1) by the factor λ_s , integrating over future histories, and then summing over $s \in \{t, \dots, \bar{n}\}$, Equation (5) is rewritten

$$\begin{aligned} \lambda_t e_{nt} &= \rho^{-1} E_t \left[\sum_{s=t}^{\bar{n}} \lambda_s (\ln \rho \alpha_0 \beta^s - \ln \lambda_s - \ln \delta) \right] \\ &= \rho^{-1} E_t \left(\sum_{s=t}^{\bar{n}} \lambda_s \right) \ln \rho \alpha_0 - E_t \\ &\quad \times \left[\sum_{s=t+1}^{\bar{n}} \lambda_s (\ln \lambda_s - s \ln \beta) \right] - E_t \left(\sum_{s=t}^{\bar{n}} \lambda_s \right) \ln \delta \\ &= \rho^{-1} (p_{t\bar{n}} \ln \rho \alpha_0 - q_{t\bar{n}} - p_{t\bar{n}} \ln \delta) \end{aligned}$$

Rearranging,

$$\ln \delta = \ln \rho + \ln \alpha_0 - \frac{q_{t\bar{n}} + \rho \lambda_t e_{nt}}{p_{t\bar{n}}}$$

Therefore, the indirect utility is

$$-\rho^{-1} \delta p_{t\bar{n}} = -\alpha_0 p_{t\bar{n}} \exp\left(-\frac{q_{t\bar{n}} + \rho \lambda_t e_{nt}}{p_{t\bar{n}}}\right)$$

as required. □

PROPOSITION 2. *The value function obtained from choosing $(c_{nt}, e_{n,t+1})$ either to maximize Equation (9) subject to Equation (10) or to maximize Equation (11) subject to Equation (12) is*

$$(A.2) \quad -p_{t\bar{n}} \alpha_j^{\lambda_t / p_{t\bar{n}}} \alpha_0^{(p_{t\bar{n}} - \lambda_t) / p_{t\bar{n}}} \exp\left[-\frac{\rho \lambda_t e_{nt} + \rho E_t(\lambda_{t+1} w_{n,t+1}) + q_{t\bar{n}}}{p_{t\bar{n}}}\right]$$

PROOF. Preliminary to solving this problem, we remark from the definitions of $p_{t+1, \bar{n}}$ and $q_{t+1, \bar{n}}$, given by Equations (6) and (7), respectively, by the law of iterated expectations:

$$\begin{aligned} p_{t, \bar{n}} &= E_t\left(\sum_{s=t}^{\bar{n}} \lambda_s\right) \\ &= 1 + E_t\left(\sum_{s=t+1}^{\bar{n}} \lambda_s\right) \\ &= \lambda_t + E_t\left[E_{t+1}\left(\sum_{s=t+1}^{\bar{n}} \lambda_s\right)\right] \\ &= \lambda_t + E_t(p_{t+1, \bar{n}}) \\ q_{t, \bar{n}} &= E_t\left(\sum_{s=t}^{\bar{n}} \lambda_s \ln \lambda_s \beta^{-s}\right) \\ &= \lambda_t \ln \lambda_t \beta^{-t} + E_t\left(\sum_{s=t+1}^{\bar{n}} \lambda_s \ln \lambda_s \beta^{-s}\right) \\ &= \lambda_t \ln \lambda_t \beta^{-t} + E_t\left[E_{t+1}\left(\sum_{s=t+1}^{\bar{n}} \lambda_s \ln \lambda_s \beta^{-s}\right)\right] \\ &= \lambda_t \ln \lambda_t \beta^{-t} + E_t(q_{t+1, \bar{n}}) \end{aligned}$$

The first-order conditions for c_{nt} and $e_{n,t+1}$ are

$$\begin{aligned} \rho \alpha_j \beta^t \exp(-\rho c_{nt}^o) &= \lambda_t \eta \\ \alpha_0 \rho \exp \left[-\frac{q_{t+1, \bar{n}} + \rho \lambda_{t+1} (e_{n,t+1}^o + w_{n,t+1})}{p_{t+1, \bar{n}}} \right] &= \eta \end{aligned}$$

where η is the Lagrange multiplier associated with Equation (10). Substituting the implied Frisch demands for c_{nt}^o and each $e_{n,t+1}^o$ into the utility function yields

$$-\rho^{-1} \lambda_t \eta - \rho^{-1} E_t(p_{t+1, \bar{n}}) \eta = -\rho^{-1} p_{t\bar{n}} \eta$$

To solve for η in terms of wealth and security prices, we take the natural logarithm of the first-order conditions to obtain

$$\begin{aligned} \ln \rho \beta^t + \ln \alpha_j - \rho c_{nt}^o &= \ln \lambda_t + \ln \eta \\ \ln \alpha_0 \rho - \left[\frac{q_{t+1, \bar{n}} + \rho \lambda_{t+1} (e_{n,t+1}^o + w_{n,t+1})}{p_{t+1, \bar{n}}} \right] &= \ln \eta \end{aligned}$$

Multiplying the first equation by λ_t and the second equation by $p_{t+1, \bar{n}}$ and then making $\rho \lambda_t c_{nt}^o$ and $\rho \lambda_{t+1} e_{n,t+1}^o$ the subjects of their respective equations yields

$$\begin{aligned} \rho \lambda_t c_{nt}^o &= \lambda_t (\ln \rho \alpha_j \beta^t - \ln \lambda_t - \ln \eta) \\ \rho \lambda_{t+1} e_{n,t+1}^o &= p_{t+1, \bar{n}} \ln \alpha_0 \rho - q_{t+1, \bar{n}} - \rho \lambda_{t+1} w_{n,t+1} - p_{t+1, \bar{n}} \ln \eta \end{aligned}$$

Now integrating the second equation over the states that can occur in period $t+1$ and then summing over $s \in \{t+1, \dots, \bar{n}\}$, Equation (10), scaled up by the factor ρ , is rewritten as

$$\begin{aligned} \rho \lambda_t e_{nt} &= \lambda_t (\ln \rho \alpha_j \beta^t - \ln \lambda_t - \ln \eta) \\ &+ E_t(p_{t+1, \bar{n}} \ln \alpha_0 \rho - q_{t+1, \bar{n}} - \rho \lambda_{t+1} w_{n,t+1} - p_{t+1, \bar{n}} \ln \eta) \\ &= -[\lambda_t + E_t(p_{t+1, \bar{n}})] \ln \eta + [\lambda_t + E_t(p_{t+1, \bar{n}})] \ln \rho \\ &+ [\lambda_t \ln \beta^t - \lambda_t \ln \lambda_t - E_t(q_{t+1, \bar{n}})] \\ &- \lambda_t \ln \alpha_j + E_t(p_{t+1, \bar{n}}) \ln \alpha_0 - E_t(\rho \lambda_{t+1} w_{n,t+1}) \\ &= -p_{t\bar{n}} \ln \eta + p_{t\bar{n}} \ln \rho + \lambda_t \ln \alpha_j + (p_{t\bar{n}} - \lambda_t) \ln \alpha_0 - q_{t\bar{n}} - E_t(\rho \lambda_{t+1} w_{n,t+1}) \end{aligned}$$

Rearranging to make $\ln \eta$ the subject of the equation, we get

$$\ln \eta = p_{t\bar{n}}^{-1} [\lambda_t \ln \alpha_j + (p_{t\bar{n}} - \lambda_t) \ln \alpha_0 - \rho \lambda_t e_{nt} - q_{t\bar{n}} - E_t(\rho \lambda_{t+1} w_{n,t+1})] + \ln \rho$$

Therefore, the indirect utility for this problem is the expression in Equation (A.2) as claimed. \square

PROPOSITION 3. *If $l_{1nt} = 1$, then the cost-minimizing feasible contract is*

$$w_{1n,t+1} = \rho^{-1} \left(\frac{p_{t+1, \bar{n}}}{\lambda_{t+1}} \right) \left(\frac{\lambda_t}{p_{t\bar{n}} - \lambda_t} \right) \ln(\alpha_1/\alpha_0)$$

PROOF. The shareholders' objective is to minimize the discounted value of expected compensation $E_t(w_{n,t+1})$ or, equivalently, from Equation (14), maximize $E_t(\ln v_{n,t+1})$ subject to the participation constraint (Equation 16). The Lagrangian for this constrained optimization problem is

$$E_t(\ln v_{n,t+1}) + \kappa \left[(\alpha_0/\alpha_1)^{\lambda_t/(p_{t\bar{n}} - \lambda_t)} - E_t(v_{n,t+1} | l_{ntj} = 1) \right]$$

where κ , the Lagrange multiplier, is an F_t -measurable function. The associated first-order condition is

$$(A.3) \quad v_{n,t+1}^{-1} = \kappa$$

Equation (A.3) demonstrates that $v_{n,t+1}$ is also F_t -measurable. Therefore,

$$v_{n,t+1} = E_t(v_{n,t+1} | l_{ntj} = 1)$$

Noting that a necessary condition for cost minimization is that the participation constraint be met with equality, it now follows from Equation (16) that

$$\begin{aligned} (\alpha_0/\alpha_1)^{\lambda_t/(p_{t\bar{n}} - \lambda_t)} &= v_{n,t+1} \\ &= \exp\left(\frac{-\rho\lambda_{t+1}w_{n,t+1}}{p_{t+1, \bar{n}}}\right) \end{aligned}$$

the second equality following from the definition of $v_{n,t+1}$ given in Equation (14). On making $w_{n,t+1}$ the subject of the equation, the proposition is proved. \square

PROPOSITION 4. *There is a unique, strictly positive solution for η to*

$$\int \frac{f_2(x)}{\eta(\alpha_2/\alpha_1)^{\lambda_t/(p_{t\bar{n}} - \lambda_t)} - \eta g(x) + 1} dx = 1$$

Denote this unique solution by η_{nt} . If $l_{2nt} = 1$, then the cost-minimizing feasible contract is

$$\begin{aligned} w_{2n,t+1} &= \rho^{-1} \left(\frac{p_{t+1, \bar{n}}}{\lambda_{t+1}} \right) \left(\frac{\lambda_t}{p_{t\bar{n}} - \lambda_t} \right) \ln(\alpha_2/\alpha_0) \\ &+ \rho^{-1} \left(\frac{p_{t+1, \bar{n}}}{\lambda_{t+1}} \right) \ln \left[1 + \eta_{nt} (\alpha_2/\alpha_1)^{\lambda_t/(p_{t\bar{n}} - \lambda_t)} - \eta_{nt} g(x_{n,t+1}) \right] \end{aligned}$$

PROOF. Noting that the objective function $E_t(\ln v_{n,t+1})$ is strictly concave and the feasibility constraints (Equations 16 and 18) are linear in $v_{n,t+1}$, it follows from

the Kuhn Tucker theorem that the solution to this optimization problem can be found by maximizing

$$E_t[\ln(v_{n,t+1}) | I_{nt2} = 1] + \eta_{1nt} E_t[(\alpha_2/\alpha_0)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} - v_{n,t+1} | I_{nt2} = 1] + \eta_{2nt} E_t[v_{n,t+1} g(x_{n,t+1}) - v_{n,t+1} (\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} | I_{nt2} = 1]$$

with respect to $v_{n,t+1}$, where η_{1nt} and η_{2nt} are the respective Kuhn Tucker multipliers associated with the feasibility constraints. The first-order condition is

$$(A.4) \quad v_{n,t+1}^{-1} = \eta_{1nt} + \eta_{2nt} [(\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} - g(x_{n,t+1})]$$

Multiplying both sides of the first-order condition (Equation A.4) by $v_{n,t+1}$, adding and subtracting $\eta_{1nt}(\alpha_0/\alpha_2)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)}$, and then taking expectations yields

$$1 = \eta_{2nt} E_t\{[g(x_{n,t+1}) - (\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)}] v_{n,t+1} | I_{nt2} = 1\} + \eta_{1nt} E_t[(\alpha_0/\alpha_2)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} - v_{n,t+1} | I_{nt2} = 1] + \eta_{1nt} (\alpha_0/\alpha_2)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)}$$

Notice that the first and second terms of this equation are the complementary slackness conditions for the Kuhn Tucker problem, which are zero. Therefore, the preceding equation implies that

$$(A.5) \quad \eta_{1nt} = (\alpha_2/\alpha_0)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)}$$

Using the first-order condition (Equation A.4) to substitute for $v_{n,t+1}$ in the incentive compatibility constraint (Equation 18) yields the ratio of the Lagrange multipliers $\eta_{nt} \equiv \eta_{2nt}/\eta_{1nt}$ as a solution in η to

$$(A.6) \quad \int \left[\frac{g(x) - (\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)}}{(\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} - g(x) + \eta^{-1}} \right] f_2(x) dx = 0$$

which can be expressed as:

$$\int \frac{f_2(x)}{\eta (\alpha_2/\alpha_1)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)} - \eta g(x) + 1} dx = 1$$

Let $\tilde{\eta}_{nt}$ denote any positive solution to Equation (A.6). Noting that η_{1nt} has a closed-form solution given by Equation (A.5), it follows that

$$\tilde{\eta}_{nt} \equiv \frac{\tilde{\eta}_{2nt}}{(\alpha_2/\alpha_0)^{\lambda_t/(p_{i\bar{n}} - \lambda_t)}}$$

for some positive $\tilde{\eta}_{2nt}$. We now define

$$\tilde{v}_{n,t+1}^{-1} \equiv (\alpha_2/\alpha_0)^{\lambda_t/(p_{t\bar{n}}-\lambda_t)} + \tilde{\eta}_{2nt} \left[(\alpha_2/\alpha_1)^{\lambda_t/(p_{t\bar{n}}-\lambda_t)} - g(x_{n,t+1}) \right]$$

By construction, $\tilde{v}_{n,t+1}$ satisfies the first-order conditions for the Lagrangian multipliers η_{1nt} and $\tilde{\eta}_{2nt}$. Since the criterion function for the transformed maximization problem is strictly concave and the constraints are linear, there is at most one stationary point, which implies that $\tilde{v}_{n,t+1} = v_{n,t+1}$ and hence $\tilde{\eta}_{2nt} = \eta_{2nt}$, thus establishing that $\tilde{\eta}_{nt} = \eta_{nt}$ is uniquely defined by the solution to Equation (A.6).

Combining the first-order condition (Equation A.4) with the solution to η_{1nt} (Equation A.5), we thus obtain

$$v_{n,t+1}^{-1} = (\alpha_2/\alpha_0)^{\lambda_t/(p_{t\bar{n}}-\lambda_t)} \left\{ 1 + \eta_{nt} \left[(\alpha_2/\alpha_1)^{\lambda_t/(p_{t\bar{n}}-\lambda_t)} - g(x_{n,t+1}) \right] \right\}$$

Substituting for $v_{n,t+1}$ using Equation (14) and making $w_{n,t+1}$ the subject of the resulting equation, the proposition is proved. □

PROPOSITION 5. The optimal long-term contract can be written as a sequence of short-term contracts in which the shareholders' expected profit every period is zero; the actions and payment plans are identical to those which would be offered in a one-period problem and where the available technology is the same and a manager of the same age retires at the end of each period.

PROOF. The proposition can be proved directly. To proceed, the existence of an optimal long-term contract is first established by writing down the principal's optimization problem and showing that a maximum exists (that the supremum over feasible policies is itself feasible). Then time additivity in preferences and technology is exploited to show that the problem has a recursive structure. This step demonstrates that the optimal contract can be written as a sequence of one-period contracts. Finally, the assumption of the exponential utility is used to show that monitoring the agent's wealth is redundant. However, in the interests of brevity, we sketch the modifications required in order to appeal to the proofs of Theorems 4 and 5 on pages 22 through 25 of Fudenberg et al. (1990), hereafter abbreviated FHM. First we note that their Assumptions 1 (verifiability), 2 (finite contract term), and 8 (history independent technology), listed on pages 6 and 23, are satisfied by our model. Their Assumption 3 (equal access to banking) is relaxed by our assumption of competitive, complete markets coupled to monitoring of the agent's financial transactions, while their Assumption 7 (exponential utility) is modified slightly to (2.2), which assumes a finite age of death.

To verify Assumptions 5 (common knowledge of preferences over action-payment streams) and 6 (decreasing utility frontier), for any date t , define c'_{ns} as consumption at date s , net of an annuity payable each year to death (at the end of period \bar{n}) that e_{nt} , wealth at date t , would support.

$$c'_{ns} \equiv c_{ns} - \frac{e_{nt}}{p_{t\bar{n}}}$$

Noting that $p_{t\bar{n}} \equiv E_t(\sum_{s=t}^{\bar{n}} \lambda_s)$, the period t budget constraint for the agent's intertemporal consumption problem can be expressed as

$$\begin{aligned}
 & 0 \leq \lambda_t e_{nt} + E_t \left[\sum_{s=t}^{\bar{n}} \lambda_s (w_{ns} - c_{ns}) \right] \\
 & = \lambda_t E_t \left(\sum_{s=t}^{\bar{n}} \lambda_s \right) \frac{e_{nt}}{p_{t\bar{n}}} + E_t \left[\sum_{s=t}^{\bar{n}} \lambda_s (w_{ns} - c_{ns}) \right] \\
 \text{(A.7)} \quad & = E_t \left[\sum_{s=t}^{\bar{n}} \lambda_s \left(w_{ns} - c_{ns} + \frac{e_{nt}}{p_{t\bar{n}}} \right) \right] \\
 & = E_t \left[\sum_{s=t}^{\bar{n}} \lambda_s (w_{ns} - c'_{ns}) \right]
 \end{aligned}$$

while the utility stream received between t and \bar{n} is

$$\begin{aligned}
 & - \sum_{s=t}^{\tau} \sum_{j=1}^{\bar{n}} \beta^s l_{nsj} \alpha_j \exp(-\rho c_{ns}) \\
 \text{(A.8)} \quad & = - \sum_{s=t}^{\bar{n}} \sum_{j=1}^2 \beta^s l_{nsj} \alpha_j \exp \left[-\rho \left(c'_{ns} + \frac{e_{nt}}{p_{t\bar{n}}} \right) \right] \\
 & = - \exp(-\rho e_{nt}/p_{t\bar{n}}) \sum_{s=t}^{\bar{n}} \sum_{j=1}^2 \beta^s l_{nsj} \alpha_j \exp(-\rho c'_{ns})
 \end{aligned}$$

Therefore, the problem of choosing the F_s -measurable consumption functions c_{ns} for all $s \in \{t, \dots, \bar{n}\}$ to maximize the expected value of the top line of Equation (A.8) subject to the top line of Equation (A.7) is equivalent to optimizing the expected value of the bottom line of Equation (A.8) subject to the bottom line of Equation (A.7) in the F_s -measurable consumption functions c'_{ns} for all $s \in \{t, \dots, \bar{n}\}$. Since e_{nt} only enters the problem through the transformed utility (Equation A.8) within the factor of proportionality $\exp(-\rho e_{nt}/p_{t\bar{n}})$, it effectively drops out of the agent's consumption allocation problem, thus establishing Assumption 5. Similarly, since the utility function is increasing in e_{nt} , it follows that Assumption 6 (decreasing utility frontier) is also met.

Second, Theorem 3 of FHM (on their page 21) extends to our environment. Given the preceding discussion, the only point to check is that replacing their Assumption 3 (of equal access to banks) with our assumption of complete markets with public disclosure leaves the results of the theorem unchanged. The proof in FHM only uses Assumption 3 to justify rescheduling payments, which, with the richer set of securities available to both parties in our environment, is still possible here.

The last step is to show that the results of Theorem 5 in FHM apply. Parts (i) and (ii) (on page 24 of FHM) follow directly from the preceding discussion. Finally, part (iii) follows after modifying Equation (6.7) in FHM along the lines they suggest in footnote 13 (on page 25). □

APPENDIX B: CONSTRUCTING THE DATA SET

This appendix describes the construction of the variables used in the estimation.

B.1. Compensation. The Antle and Smith (1985, 1986) current income equivalent was used as the measure of total compensation. Their work contains a detailed description of its construction. Briefly, a current income equivalent is the cash payment that makes an individual indifferent between receiving that sum or the after-tax value of the compensation package. It is calculated as the after-tax sum of salary and bonus, changes in expected retirement benefits, market value of stock grants, value of stock option grants, the present value of deferred bonuses, stock grants and option grants value of dividend units, and the change in wealth due to change in the value of options and stocks held.

B.2. Position. The position variable was created from the original Antle and Smith data set using the pretax salary and bonus information. We defined the CEO indicator variable d_{2n1} by setting $d_{2n1} = 1$ if the salary and bonus for that observation was the top salary and bonus paid to the executives by the firm in that year and setting $d_{2n1} = 0$ otherwise. Similarly, the (common) indicator variable for the two non-CEO positions was defined as $d_{2n2} = 1 - d_{2n1}$.

B.3. Age and Tenure. Executive age in year t was calculated from the year that the executive turned 65, and unless otherwise specified, we assumed that the year when the executive turned 65 was the executive's last year in the sample. Tenure in the current job was defined as the difference between the executive's last year in the sample and his or her first year in sample or 1947, whichever came later.

B.4. Financial Market Measures. Data on market performance was computed from raw data taken from the Center for Research in Security Prices (CRSP) monthly stock returns tape and the consumer price index. Denote by φ_{mnt} the nominal monthly rate of return for firm n in month m of year t , and let ζ_t denote the inflation rate for year t computed from the consumer price index. Then we defined the real annual rate of return for each stock as the compounded monthly returns divided by the inflation rate:

$$\pi_{nt} = \frac{\prod_{m=1}^{12} (1 + \varphi_{mnt})}{1 + \zeta_t}$$

The value-weighted real annual return, denoted π_t , was computed using the same formula as Equation (B.1) by substituting the monthly nominal value weighted return φ_{mt} for φ_{mnt} .

B.5. Accounting Measures. Data on firm accounting performance were collected from the Annual Compustat Industrial files. Return on assets, return on equity, sales, and the debt-to-equity ratio were used as instruments in the analysis. Return on assets was computed as the ratio of income before extraordinary items to

total assets. Return on equity was computed as the ratio of income before extraordinary items to total shareholders equity. The debt-to-equity ratio was computed as the ratio of total liabilities to total shareholders' equity.

B.6. Console Bond Prices. The prices of the console bonds were computed by first computing the forward rates for a risk-free security from the present until the year of the expected death and then pricing the bond. The expected life span for each year was taken from actuarial tables on life expectancy. The Solomon Brothers data on yields and yield spreads were used to compute the forward rates. For years for which no yield was available, linear yield curves were fitted, and the missing data were interpolated. The yield curve is assumed to flatten after 30 years, so the yield for subsequent years is assumed to be the same as the yield in year 30. Because no data are available prior to 1950, we assumed that the yields in those years were the same as the yields in 1950.

REFERENCES

- ALTUG, S., AND R. MILLER, "Household Choices in Equilibrium," *Econometrica* 58 (1990), 543–70.
- ANTLE, R., AND A. SMITH, "Measuring Executive Compensation: Methods and an Application," *Journal of Accounting Research* 23 (1985), 296–325.
- AND ———, "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives," *Journal of Accounting Research* 24 (1986), 1–39.
- BELLMAN, R., *Dynamic Programming* (Princeton, NJ: Princeton University Press, 1957).
- DEBREU, G., *Theory of Value* (New Haven, CN: Yale University Press, 1959).
- DONALD, S., AND H. PAARSCH, "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review* 34 (1993), 121–48.
- FERRALL, C., AND B. SHEARER, "Incentives and Transactions Costs within the Firm: Estimating an Agency Model using Payroll Records," *Review of Economic Studies* 66 (1999), 309–338.
- FUDENBERG, D., B. HOLMSTROM, AND P. MILGROM, "Short-Term Contracts and Long-Term Agency Relationships," *Journal of Economic Theory* 50 (1990), 1–31.
- GIBBONS, R., AND K. MURPHY, "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy* 100 (1992), 468–505.
- GROSSMAN, S., AND O. HART, "A Theory of Competitive Equilibrium in Stock Market Economies," *Econometrica* 47 (1979), 293–329.
- AND ———, "An Analysis of the Principal-Agent Problem," *Econometrica* 51 (1983), 7–46.
- HALL, B., AND J. LIEBMAN, "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* 113 (1998), 653–92.
- HANSEN, L., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50 (1982), 1029–54.
- HAUBRICH, J., "Risk Aversion, Performance Pay, and the Principal-Agent Problem," *Journal of Political Economy* 102 (1994), 258–76.
- HOLMSTROM, B., "Moral and Hazard and Observability," *Bell Journal of Economics* 10 (1979), 74–91.
- , "Moral Hazard in Teams," *Bell Journal of Economics* 13 (1982), 234–40.
- JENSEN, M., AND K. MURPHY, "Performance Pay and Top-Management Incentives," *Journal of Political Economy* 98 (1990), 225–64.
- LAMBERT, R., AND D. LARCKER, "An Analysis of the Use of Accounting and Market Measures of Performance in Executive Compensation," *Journal of Accounting Research* 25 (1987), 85–125.
- MADDALA, G., *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge, England: Cambridge University Press, 1983).
- MALCOLMSON, T., AND F. SPINNEWYN, "The Multiperiod Principal-Agent Problem," *Review of Economic Studies* 55 (1988), 391–408.

- MASSON, R., "Executive Motivations, Earnings, and Consequent Equity Performance," *Journal of Political Economy* 79 (1971), 1278-92.
- MERTON, R., "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory* 3 (1971), 373-413.
- MILLER, R., "Job Matching and Occupational Choice," *Journal of Political Economy* 92 (1984), 1086-1120.
- , "Estimating Models of Dynamic Optimization with Microeconomic Data," in M. Pesaran and P. Schmidt, eds., *Handbook of Applied Econometrics*, Vol. 2 (Oxford, England: Basil Blackwell, 1997), pp. 246-99.
- AND H. SIEG, "A Microeconometric Comparison of Household Behavior Between Countries," *Journal of Business and Economic Statistics* 15 (1997), 237-54.
- MIRRELES, J., "The Optimal Structure of Incentives and Authority within an Organization," *The Bell Journal of Economics* 7 (1976), 105-31.
- MURPHY, K., "Corporate Performance and Managerial Remuneration: An Empirical Analysis," *Journal of Accounting Economics* 7 (1985), 11-42.
- , "Incentives, Learning and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts," *Rand Journal of Economics* 17 (1986), 59-76.
- NEWBY, W., "A Methods of Moments Interpretation of Sequential Estimators," *Economic Letters* 14 (1984), 201-6.
- REY, P., AND B. SALANIE, "Long-Term, Short-Term and Renegotiation," *Econometrica* 58 (1990), 597-620.
- RUBINSTEIN, M., "A Discrete-Time Synthesis of Financial Theory," *Research in Finance* 3 (1981), 53-102.