Identifying and Testing Generalized Moral Hazard Models of Managerial Compensation*

George-Levi Gayle and Robert A. Miller
Tepper School of Business, Carnegie Mellon University

June 12, 2007

Abstract

This paper seeks to answer two questions: in models of executive compensation how important is hidden information relative to moral hazard, and how biased are empirical measures of moral hazard in econometric models that do not account for hidden information. An analytical stage of this paper exploit restrictions from the theory of optimal contracting to identify hidden information and differentiate its effects from moral hazard. An empirical stage uses and develops nonparametric and numerical methods to quantify the importance of the various factors identified in the first stage using a large longitudinal data set on executives.

1 Introduction

Managers are paid to organize human resources in creative ways that add value to their firm. Since their activities are hard to monitor directly, managers are rarely paid for their inputs. Rather, compensation is tied to various indicators of managerial effort, such as their firm's performance. Linking managerial compensation to the firm's performance requires the manager to hold a substantial amount of personal wealth in assets that are sensitive to the firm's performance, such as stocks and options. Thus managerial compensation schemes try to correct for moral hazard by preventing managers from diversifying their wealth as much as they would otherwise.

Implementing such a scheme becomes complicated when shareholders do not know how much wealth the manager should vest in his own firm to simultaneously minimize the cost to shareholders, meet the manager's conditions for remaining with the firm, and align his incentives with those of the shareholders. Shareholders not only rely on information from management about the firm's prospects. They also rely on managers for guidance about organizational and incentive structures that will unleash the firm's potential. The duties

*Preliminary and incomplete; please do not quote without the expressed consent of the authors.

of executives in large corporations necessarily make them privy to information about their firm's performance that is not available to the stockholders at the time.

From an empirical standpoint, trading by corporate insiders appears over time to be increasingly profitable. Seyhun (1986) finds that insiders appears tend to buy before an abnormal rise if stock prices and to sell before an abnormal decline. Earlier studies by Lorie and Niederhoffer (1968), Jaffe (1974), and Finnerty (1976) draw similar conclusions. We were able to confirm these conclusions using changes in executives holding of insider wealth using data from 1991 to 2005. More recently, Seyhun (1992a) finds compelling evidence that insider trading volume, frequency, and profitability all increases significantly during the 1980s. Over the decade, he documents that insiders earned over 5% abnormal returns on average. Seyhun (1992b) determines that insider trades predict up to 60 of the variation in year-ahead returns. According, hidden information seems to be an economically important phenomenon in executive compensation.

The two requirements, that a goodly portion of the manager's wealth should be vested in the firm to align the incentives between the firm's managers and its shareholders, and that the manager knows better than the shareholders the distribution of the firm's returns and how it varies with her own managerial activities, is at the heart of the paradox of insider information and moral hazard. This paper analyzes the quantitative importance of this paradox. We present a model of moral hazard and hidden information, characterize the optimal contract, establish necessary and sufficient conditions for identifying the model using data on the firm's returns, the compensation to managers and the economic conditions, and finally estimate the model to assess the importance of moral hazard versus insider information in executive contracts for publicly traded firms.

Section 2 presents a theoretical framework for exploring moral hazard and hidden information. In this model shareholders do not observe the prospects of the firm at the beginning of the period or manager's activities within the period. Contracts between shareholders and the executives must satisfy three conditions, a participation constraint, that assures the manager she will have higher expected utility from employment with her firm rather than another one, an incentive compatibility constraint, that induces her to maximize the value of the firm rather than using the resources of the firm to pursue some other objective, and an insider trading condition that reflects shareholders' beliefs that the manager will pursue admissible insider trading when the opportunity arises.

We characterize the set of feasible facing shareholders in Sections 3 and the solution to their cost minimization problem in Section 4 before providing measures for quantifying the relative importance of hidden information versus moral hazard. The derivation of the optimal contract exploits the fact the truth telling constraint rewards the manager in expected utility terms before the firm's output for the period is revealed to anyone.

In Section 5 we analyze identification without imposing any further parametric assumptions on our theoretical model. The first result is that, absent further restrictions, the model of pure moral hazard is identified up to the constant coefficient of absolute risk aversion. Our second result that if hidden information is an empirically significant factor in contract design, a testable null hypothesis, then the conditions that characterize truth telling about the state of the firm provide a natural experiment for estimating the executive's preferences towards risk.

We test the null hypothesis of hidden information in using the model to investigate CEO compensation in Section 6, and estimate its importance relative to moral hazard. Our empirical investigations, complied from three main sources Standard & Poor's ExecuComp and Compustat databases and Executive Compensation Reports data on firm compensation plan responses to Section 162(m), tracks 1,500 firms over an 9 year panel beginning in 1992 in the S&P 500, Midcap, and Smallcap indices and contains information on the five highest-paid executives for 1,837 unique CUSIP identifiers.

2 The Model

Our model focuses on the executive compensation when the manager is subject to moral hazard and also has private information about the firm's future returns. At the beginning of each period the manager observes the firm's prospects and provides some accounting information about them. The accounting information are based on protocols can be legally verified. This prevents the manager from overstating the firm's profitable opportunities, but does not prevent him from fully revealing their extent. The directors on the board proposes compensation plan to the manager based on the accounting information provided to them. Based on the board's proposal the manager decides whether to remain with the firm or leave it and picks real consumption expenditure for the period. If she accepts the contract offer, she then chooses a work routine, which is not observed by the directors, The return on the firm's assets are realized at the end of the period. It depends on the how well the firm was managed during the period, the private information available to the manager as well as other factors that were not anticipated by anybody. The objective of the manager is to sequentially maximize her expected lifetime utility, and the goal of the firm is expected value maximization.

2.1 Information, choices and returns

At the beginning of period t the manager is paid compensation denoted w_t for her work in period $t-1$ according to the schedule the shareholders had previously committed, and her managerial contracts is up for renewal. At that time the prospects of the firm $s_t \equiv (s_{1t}, s_{2t})$ are fully revealed to the manager but partially hidden to the shareholders. Shareholders observe $s_{1t} \in S_1$ but not $s_{2t} \in S_2$. We assume $S_1 \equiv \{1, \ldots, S\}$ is finite, an assumption made purely for expositional and purposes and notational simplicity. We also assume throughout that $S_2 \equiv \{1, 2\}$, which can also be relaxed, but would require the analysis to be extended. The board announces how the manager's compensation will be determined as a function of what she now tells them about the firm's prospects and its subsequent performance, as measured by abnormal returns x_{t+1} revealed at the beginning period $t+1$.

The manager then truthfully declares or lies about the firm's prospects by announcing $s'_t \in S$, effectively selecting one from many schedules $w(s'_t, x_{t+1})$ indexed by her announcement s'_t. She then makes her consumption and unobserved labor choices, (c_t, l_t). With regards work effort, the manager has three choices in each period t, to work diligently for the firm, to be employed by the firm but shirk, or to be engaged outside the firm, either with another firm or in retirement. Let $l_t \equiv (l_{t0}, l_{t1}, l_{t2})$ where $l_{tj} \in \{0, 1\}$ for $j \in \{0, 1, 2\}$ and

$$\sum\nolimits_{j=0}^{3} l_{tj} = 1$$

where $l_{t0} = 1$ signifies choosing another job or retirement, l_{t1} means choosing to be employed by the firm but to pursue different objectives than maximizing the firm's value, and l_{t2} means that the manager pursues the shareholders objectives of value maximization. Consumption in period t is a positive real number denoted by c_t.

At the beginning of the period $t+1$ abnormal returns x_{t+1} for the firm are drawn from a probability distribution which depends on the true state s_t and the manager's action l_t. We denote the probability distribution function for abnormal returns in period t when the manager works diligently and the state is s by $F_s(x_{t+1})$, and assume it is differentiable with density $f_s(x_{t+1})$. Similarly, let $f_s(x_{t+1})\, g_s(x_{t+1})$ denote the probability density function for abnormal returns in period t when the manager shirks. Since $f_s(x)\, g_s(x)$ is a density, $g_s(x)$ must be a positive mapping with $E_s[g_s(x)] = 1$,where the the expectation is taken with respect to $f_s(x)$. Compensation to the manager is denoted by $w_{t+1} \equiv w_s(x_{t+1})$.

We also assume there is an upper range of returns that, conditional on the state s,might be achieved with diligence, but cannot be achieved through shirking. Formally, for each $s \in \{1, 2\}$there exists some finite real number denoted $\overline{x}_s$, such that $F_s(\overline{x}_s) < 1$ and if $x > \overline{x}_s$, then $g_s(x) = 0$. For future reference we also define $\overline{x} = \max\{\overline{x}_1, \overline{x}_2\}$. This assumption has no special ramifications for the optimal contract, but is exploited in identification, providing a regularity condition. Relaxing the assumption weakens but does not overturn our results, and since it seems plausible for many kinds of enterprises, we impose it throughout.

2.2 Preferences and the budget constraint

Preferences over consumption and work are parameterized by a utility function exhibiting absolute risk aversion that is additively separable over periods and multiplicatively separable with respect to consumption and work activity within periods. In the model we estimate, lifetime utility can be expressed as:

$$-\sum\nolimits_{t=0}^{T} \sum\nolimits_{j=0}^{J} \beta^t \alpha_j l_{tj} \exp(-\rho c_t)$$

where β is the constant subjective discount factor, ρ is the constant absolute level of risk aversion, and α_j is a utility parameters with consumption equivalent $-\rho^{-1} \log(\alpha_j)$ that measures the distaste from working at level $j \in \{0, 1, 2\}$. We assume $\alpha_2 > \alpha_1 > \alpha_0$ meaning that compared to the activity called shirking, diligence is more aligned to the shareholders' interest than the managers interests.

We assume there are a complete set of markets for all publicly disclosed events, with price measure Λ_t defined on F_t and derivative λ_t. This implies that consumption by the manager is limited by a lifetime budget constraint which reflects both the opportunities she faces as an insider trader, and the expectations she has about her compensation. The lifetime wealth constraint is endogenously determined by the manager's work activities and her insider trading activity. By assuming markets exist for consumption contingent on any public event, we effectively attribute all deviations from the law of one price to the particular market imperfections under consideration. Let e_0 denote the endowment at date 0, and let p_t denote the current price of shares, denumerable in terms of forgone consumption units in period t. We also measure w_t, the manager's compensation in period t, in units of current consumption. To indicate the dependence of the consumption possibility set on the set of contingent plans determining labor supply and effort, we define $E_0[\bullet\,|l]$ as the expectations operator conditional on work and effort level choices throughout the manager's working life. The budget constraint can then be expressed as:

$$E_0\left[\sum\nolimits_{t=0}^{T}\lambda_t\left(c_t-w_t\right)|l\right]\leq e_0$$

2.3 Optimal consumption and savings

To preface the derivation of the optimal contract we derive the indirect utility function for the worker upon leaving the firm, and then solve for optimal consumption when the manager plans to work at most one period before retiring. Although there are complete markets in this model, the manager requires only two securities to attain her optimal consumption stream. Accordingly let b_t denote the price of a bond that pays of a unit of consumption from period t through to period T, relative to the price of a unit of consumption in period t.

$$b_t=E_t\left(\sum\nolimits_{s=0}^{T}\frac{\lambda_s}{\lambda_t}\right)$$

Also let a_t denote the price of a security which pays off the random quantity $(\log\lambda_s-s\log\beta)$ in periods t through T.

$$a_t=E_t\left[\sum\nolimits_{t=s}^{T}\frac{\lambda_s}{\lambda_t}\left(\log\lambda_s-s\log\beta\right)\right]$$

It is straightforward to show that maximizing the utility function upon retirement

$$-\sum\nolimits_{t=0}^{T}\alpha_0\beta^t\exp\left(-\rho c_t\right)$$

subject to the budget constraint

$$E_0\left[\sum\nolimits_{t=0}^{T}\lambda_t c_t\right]\leq e_0$$

yields the indirect utility function

$$-\alpha_0 b_t\exp\left(-\frac{a_t+\rho\lambda_t e_t}{b_t}\right)$$

Applying Bellman's principle, it now follows that conditional on choosing activity α_j the two period utility staring at t and then continuing with the indirect utility from retiring the following period is

$$-\alpha_j\beta^t \exp\left(-\rho c_t\right) - \alpha_0 E_t\left[b_{t+1}\exp\left(-\frac{a_{t+1}+\rho\lambda_{t+1}e_{t+1}+\rho\lambda_{t+1}w_{t+1}}{b_{t+1}}\right)|l_{tj}=1\right]$$

Solving for the optimal consumption subject to the the two period budget constraint

$$\lambda_t c_t + E_0\left[\lambda_{t+1}e_{t+1}\right] \leq \lambda_t e_t$$

we obtain the indirect utility function

$$-b_t\alpha_j^{\frac{\lambda_t}{b_t}}\alpha_0^{1-\frac{\lambda_t}{b_t}}\exp\left(-\frac{a_t+\rho\lambda_t e_t}{b_t}\right)E_t\left[\exp\left(-\frac{\rho\lambda_{t+1}w_{t+1}}{b_{t+1}}\right)|l_{tj}=1\right]$$

3 Contracts

Appealing to Myerson (1982), the revelation principle applies to our model, which means that, for the purpose of analyzing managerial compensation and returns to shareholders, rather than consider all bargaining games of incomplete information between manager and shareholder board, we can restrict the discussion to direct revelation games. The model is solved in stages. Using the valuation function that solves the consumption savings problem, we derive the participation, incentive compatibility and truth telling constraints that circumscribe the short term contracts. This leads to a formulation of the feasible constraints for the problem. We then analyze the cost minimization problem that shareholders solve in the direct revelation game, and last, show the optimal log term contact decentralizes to a sequence of short term contracts of the form we derive.

The constraints relate to participation, incentive compatibility and truth telling. The participation constraint states that the manager is indifferent between working one period and then leaving, versus not working for the firm at all. We show this is a necessary and sufficient condition for the worker to prefer managing the firm for a period, regardless of the choices she makes in the future. The incentive compatibility constraint restricts short term contracts to those payment schedules in which the manager prefers to work diligently rather than shirk. The truth telling condition requires shareholders to write contracts that induce the manager to select a compensation schedule that reveals the firm's prospects. Finally the contract must guard against the possibility of the manager lying about the state and also shirking, which we label as the combination constraint.

3.1 Participation and incentive compatibility constraints

The participation constraint requires that the expected lifetime utility from working one more period exceeds the expected utility from retiring immediately, or

$$b_t\alpha_0^{1-\frac{\lambda_t}{b_t}}\exp\left(-\frac{a_t+\rho\lambda_t e_t}{b_t}\right)E_{tj}\left\{\left(\alpha_j\right)^{\frac{\lambda_t}{b_t}}\exp\left[-\frac{\rho\lambda_{t+1}w_{t+1}\left(s\right)}{b_{t+1}}\right]\right\} \leq \alpha_0 b_t\exp\left(-\frac{a_t+\rho\lambda_t e_t}{b_t}\right)$$

where the second subscript on the expectations operator refers to the choice of work activity. Let $v(x,s)$ denote the his utility scaled up a factor of proportionality when the manager works diligently. Symbolically:

$$v_s(x) \equiv (\alpha_2/\alpha_0)^{\frac{\lambda_t}{b_t}} \exp\left(-\frac{\rho\lambda_{t+1}w_s(x)}{b_{t+1}}\right)$$

To induce a diligent manager to participate, his expected utility from working diligently must exceed the expected utility from choosing an outside option, such as working for another firm or retiring, instead. For example, supposing there are K discrete states, which occur with probability φ_s for $s \in \{1, 2, \ldots, K\}$ we could express the participation constraint as:

$$\sum\nolimits_{s=1}^{K} \varphi_s \int_{\underline{x}}^{\infty} v_s(x) f_s(x)\, dx \leq 1$$

Given truthful revelation of the states, the incentive compatibility constraint requires the manager to prefer working diligently to shirking or, using the definition of $v_s(x)$,

$$b_t\alpha_0 \exp\left(-\frac{a_t + \rho\lambda_t e_t}{b_t}\right)\int_{\underline{x}}^{\infty} v_s(x) f_s(x)\, dx \leq b_t(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}}\alpha_0 \exp\left(-\frac{a_t + \rho\lambda_t e_t}{b_t}\right)\int_{\underline{x}}^{\infty} v_s(x) f_s(x) g_s(x)\, dx$$

Cancelling the common terms preceding the integral, we compactly restate the incentive compatibility condition as

$$\int_{\underline{x}}^{\infty}\left(g_s(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}\right) v_s(x) f_s(x)\, dx \geq 0$$

3.2 Hidden information constraints

Information hidden from shareholders further restrict the set of contracts that can be implemented. Appealing to the revelation principle we model these restrictions as truth telling constraints. We assume throughout that legal considerations induce the manager not to overstate the firm's prospects but that incentives must be provided to persuade the manager from understating them. The expected value from lying about the second state is

$$-b_t\alpha_0 \exp\left(-\frac{a_t + \rho\lambda_t e_t}{b_t}\right)\int_{\underline{x}}^{\infty} v_1(x) f_2(x)\, dx$$

Comparing this with the expected utility from reporting honestly in the second state and working diligently, we obtain the truth telling condition

$$\int [v_2(x) - v_1(x)]\, f_2(x)\, dx \leq 0$$

An optimal contract also induces the manager not to understate and shirk. The expected utility of the manager from understating and shirking is

$$-b_t(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}}\alpha_0 \exp\left(-\frac{a_t + \rho\lambda_t e_t}{b_t}\right)\int_{\underline{x}}^{\infty} v_1(x) g_2(x) f_2(x)\, dx$$

Comparing this with the utility from reporting honestly in the second state and working diligently, the combined condition is

$$\int v_2(x) f_2(x) dx \leq \int (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} v_1(x) g_2(x) f_2(x) dx$$

where $(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} v_1(x)$ is proportional to the utility obtained from shirking and announcing the first state, and $f_2(x) g_2(x)$ is the probability density function associated with shirking when the second state occurs.

Under some conditions the manager will not understate and shirk, providing the contract satisfies the participation, incentive compatibility and truth telling conditions alone. In this case the constraint is nonbinding. For example suppose there are only two states, whose differences are hidden from shareholders, and consider two additional assumptions on the probability distributions for abnormal returns. First we assume that $F_2(x)$ first order stochastically dominates $F_1(x)$. This assumption is used to unambiguously rank the states when the manager works diligently. Second we assume the stochastic dominance ordering is reversed when the manager shirks, that is for all $y \leq \overline{x}$

$$\int_{-\infty}^{y} f_1(x) g_1(x) dx \leq \int_{-\infty}^{y} f_2(x) g_2(x) dx$$

If these assumptions are satisfied then

$$\begin{aligned}
\int v_2(x) f_2(x) dx &\leq \int v_1(x) f_2(x) dx \\
&\leq \int v_1(x) f_1(x) dx \\
&\leq \int \left(\frac{\alpha_1}{\alpha_2}\right)^{\frac{\lambda_t}{b_t}} v_1(x) f_1(x) g_1(x) dx \\
&\leq \int \left(\frac{\alpha_1}{\alpha_2}\right)^{\frac{\lambda_t}{b_t}} v_1(x) f_2(x) g_2(x) dx
\end{aligned}$$

The first inequality follows from the truth telling constraint, that is not understating the firm's prospects. The second inequality follows from the assumption that $v_1(x)$ is a decreasing function in x and that the second state stochastically dominates the first when the manager is diligent. The third inequality follows from the incentive compatibility constraint. The fourth inequality comes from the assumption that the first state stochastically dominates the second when the manager shirks.

These assumptions embody the idea that if is optimal to motivate the manager, then from the shareholders' perspective, the second state is unambiguously better than the first, but that if the manager shirks, the firm's prospects fall when such opportunities arise, through negligence. Note that the second assumption is trivially satisfied when the firm's losses from shirking does not depend on the state $f_1(x) g_1(x) \equiv f_2(x) g_2(x)$, and thus opportunities from the better state can only be realized if the manager is diligent. Because these assumptions are plausible but not entirely convincing, it behoves us to analyze when the constraint is binding, and treat both possibilities.

4 Cost Minimization

Shareholders maximize the value of the firm, inducing the manager to make choices that serve their interests. It is straightforward to show from the particatipation constraint that the cost miniminizing contract for employing the manager to shirk is a constant wage of

$$w^o = \frac{b_{t+1}\lambda_t}{\rho b_t \lambda_{t+1}} \log\left(\alpha_1/\alpha_0\right)$$

that just offsets the value of leaving the firm and consequently does not depend on the state. It pays shareholders to induce the manager to distinguish between pairs of states if and only if it is more profitable to create incentives that motivate her to work diligently in at least one of the states than to shirk in both. Denote by $w_s^o(x)$ the optimal contract that induces truth telling and diligence in the s^{th} state. Our discussion implies that the manager will choose (l_{1s}, l_{2s}) for each $s \in \{1, \ldots, S\}$ to maximize the value of the firm, namely

$$\sum\nolimits_{s=1}^{K} \varphi_s \int_{\underline{x}}^{\infty} \left\{\left[zx - w_s^o(x)\right] l_{2s} - \left[zxg_s(x) - w^o\right] l_{1s}\right\} f_s(x)\, dx$$

To complete the solution to this optimization problem, this section derives the cost minimizing contracts that induce diligence in at least one state. Our formulation satisfies the Kuhn Tucker conditions, permitting us to use Lagrangian methods to characterize of the optimal short term contracts. This is equivalent to maximizing $E_t[\log v_t]$ subject to the three constraints. Since all three constraints are convex sets, their intersection is too, and from its definition the objective function is concave. Appealing to the Kuhn Tucker theorem, we formulate the problem as maximizing a Lagrangian with respect to $v_s(x)$.

4.1 Pure moral hazard

Consider a model of pure moral hazard with a finite number of multiple states denoted by $s \in \{1, \ldots, S\}$. Whether state is revealed before or after the contract is made is immaterial to both managers and shareholders. If the contract is made before the state is revealed the maximization problem is

$$\sum_{s=1}^{S} \varphi_s \int_{\underline{x}}^{\infty} \left\{\log v_s(x) + \eta_1\left[1 - v_{st}\right] + \eta_{3s} v_s(x)\left[\left(g_s(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}\right]\right\} f_s(x)\, dx$$

where φ_s denotes the probability of the s^{th} state occurring. Denoting by

$$v_s^o(x) \equiv \exp\left(-\frac{\rho\lambda_{t+1} w_s^o(x)}{b_{t+1}}\right)$$

the first order conditions are

$$v_s^o(x)^{-1} = \eta_1 + \eta_{3s}[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_s(x)]$$

Multiplying by $v_s^o(x)$ and noting that the participation constraint is met with equality proves that $\eta_1 = 1$ and hence

$$v_s^o(x) = \left\{1 + \eta_{3s}[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_s(x)]\right\}^{-1}$$

Substituting for $v_s^o(x)$ in the incentive compatibility constraint yields the solution to η_{3s} uniquely defined by:

$$\int_{\underline{x}}^{\infty} \frac{\left[(g_s(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}\right] f_s(x)}{1 + \eta_{3s}[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_s(x)]} dx = 0$$

If contracts are made after the state is revealed then a separate participation constraint applies to each state, and the objective of the firms is to maximize

$$\int_{\underline{x}}^{\infty} \left\{\log v_s(x) + \eta_1 [1 - v_{st}] + \eta_{3s} v_s(x) \left[(g_s(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}\right]\right\} f_s(x)\, dx$$

The first order conditions are the same as above and the solutions to the Kuhn Tucker multipliers are the same. Consequently the compensation schedules are identical. It now follows that the expected utility of the manager is identical in both states, and as we shall demonstrate in the next section, this provides an identifying equation for the pure moral hazard model with multiple states.

Having derived $w_s^o(x)$ in each state $s \in \{1, \ldots, S\}$, the shareholders offers a contract of $w_s^o(x)$ to the manager if

$$\int_{\underline{x}}^{\infty} [xz - w_s^o(x)] f_s(x)\, dx > \int_{\underline{x}}^{\infty} xz g_s(x) f_s(x)\, dx - w^o$$

and otherwise pays the manager a flat wage of w^o.

4.2 Hidden information

The essentials of the model with both moral hazard and private information are easy to appreciate when there are only two states $s \in \{1, 2\}$, which occur with probability φ_s, but are hidden from the shareholders purview. With minimal notational changes, the solution derived for this two state problem can be adapted to the empirical model we estimate in Section 6, where there are an uncountable number of states $s_1 \in S_1$ the shareholders can distinguish between, where $s_2 \in \{1, 2\}$, and where $\varphi_j(s_1)$ is the probability that the state is (s_1, j) conditional on observing s_1.

Following the spirit of the pure moral hazard model the shareholders maximize:

$$\sum_{s=1}^{2} \varphi_s \int_{\underline{x}}^{\infty} \left\{\log v_s(x) + \eta_1 [1 - v_{st}] + \eta_{3s} v_s(x) \left[(g_s(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}\right]\right\} f_s(x)\, dx$$
$$+\varphi_2 \eta_2 \int_{\underline{x}}^{\infty} [v_1(x) - v_2(x)] f_2(x)\, dx + \varphi_2 \eta_4 \int \left[(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} v_1(x) g_2(x) - v_2(x)\right] f_2(x)\, dx$$

where η_1, η_2, η_{31}, η_{32} and η_4 are the shadow values assigned to the linear constraints. The first order conditions are

$$v_1(x)^{-1} = \eta_1 - \eta_2 h(x) + \eta_{31}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_1(x)\right] - \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x)\, h(x)$$

$$v_2(x)^{-1} = \eta_1 + \eta_2 + \eta_{32}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_2(x)\right] + \eta_4$$

where $h(x)$ is the weighted density

$$h(x) = \frac{\varphi_2 f_2(x)}{\varphi_1 f_1(x)}$$

Multiplying both sides by $v_s(x)^{-1}$, taking the unconditional expectation, and noting the complementary slackness conditions drop out, we obtain

$$1 = \eta_1 \sum\nolimits_{s=1}^{2} \varphi_s \int_{\underline{x}}^{\infty} (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} v_s(x) f_s(x)\, dx = \eta_1$$

the second equality following from the fact that the participation constraint itself is met with equality.

The solution to the two state case is completed by solving for the Lagrange multipliers. There are several cases to analyze. The value of the firm is computed in each case, by relaxing different combinations of constraints to see whether which ones are satisfied by inequalities. For example if neither the truth tellling nor the sincerity constraints are binding, then $\eta_2 = \eta_4 = 0$ and the optimization problem reduces to a multistate moral hazard problem, solved state by state as described above. If at least one constraint is violated in ths solution, then $\eta_2 + \eta_4 > 0$, and at least one of the constraints is binding. For example when the sincerity constraint does not bind, the truth telling constraint binds, and diligence in both states is induced, we substitute the first order condition into the incentive compatibility and truth telling constraints give use the following system of three equations in the remaining three unknowns η_2, η_{31}, and η_{32} when $\eta_4 = 0$. They are:

$$\int_{\underline{x}}^{\infty} \frac{f_2(x)\, dx}{1 - \eta_2 h(x) + \eta_{31}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_1(x)\right]} = \int_{\underline{x}}^{\infty} \frac{f_2(x)\, dx}{1 + \eta_2 + \eta_{32}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_2(x)\right]}$$

$$\int_{\underline{x}}^{\infty} \frac{g_1(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}}{1 - \eta_2 h(x) + \eta_{31}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{31} g_1(x)} f_1(x)\, dx = 0$$

$$\int_{\underline{x}}^{\infty} \frac{g_2(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}}{1 - \eta_2 + \eta_{32}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{32} g_2(x)} f_2(x)\, dx = 0$$

The Kuhn Tucker theorem guarantees there is a unique solution to this equation system. Substituting these values back into the two first order equations yields the solution to the compensation schedule as a function of the states. If the sincerity constraint is met by the unconstrained solution $(v_1(x), v_2(x))$,then we conclude that $\eta_2\eta_4 = 0$.

Alternatively if the combined condition is not satisfied by the unconstrained solution, then $\eta_4 > 0$ and the new set of conditions solving the remaining four unknowns η_2, η_{31}, η_{32} and η_4 are:

$$\begin{aligned} & \int_{\underline{x}}^{\infty} \frac{f_2(x)\, dx}{1 - \eta_2 h(x) + \eta_{31} (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{31} g_1(x) + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x) h(x)} \\ = & \int_{\underline{x}}^{\infty} \frac{f_2(x)\, dx}{1 + \eta_2 + \eta_4 + \eta_{32} (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{32} g_2(x)} \end{aligned}$$

$$\int_{\underline{x}}^{\infty} \frac{g_1(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}}{1 - \eta_2 h(x) + \eta_{31} (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{31} g_1(x) + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x) h(x)} f_1(x)\, dx = 0$$

$$\int_{\underline{x}}^{\infty} \frac{g_2(x) - (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}}{1 + \eta_2 + \eta_4 + \eta_{32} (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{32} g_2(x)} f_2(x)\, dx = 0$$

$$\begin{aligned} & \int_{\underline{x}}^{\infty} \frac{f_2(x)\, dx}{1 - \eta_2 + \eta_{32} (\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{32} g_2(x) - \eta_4} \\ = & \int \frac{(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} f_2(x)\, g_2(x)\, dx}{\eta + \eta_2 h(x) + \eta_{31} \left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_1(x) \right] + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x) h(x)} \end{aligned}$$

These equations can also be used to solve the other cases. Setting $\eta_2 = 0$, we check whether the truth telling constraint is satisfied by an inequality, in which case the other multpliers are found by ignoring the first equation in the groups of three or four, the less constratined solution yielding a lower cost than the the case in which $\eta_2 > 0$. Setting $\eta_{3s} = 0$ yields $l_{s1} = 1$, in which case $w_s(x) = w^o$. As before we check which constraints are satisfied with inequalities by relxaing the truth telling and sincerity constraints.

5 Identification

The model is characterized by $f_s(x)$ and $g_s(x)$ for each state $s \in S$, which together define the probability density functions of abnormal returns in the states, the probability distribution for the states, $(\alpha_0, \alpha_1, \alpha_2)$, the preference parameters for leaving the firm, versus shirking and working within the firm, and the risk aversion parameter ρ. For expositional purposes this section assumes that T observations on (x_t, s_t, w_t) tracking an individual firm comprises the data, that the probability distribution for $s \in S$, that $f_s(x)$ are known for each s, and that $\alpha_0 \equiv 1$. Although the states are partially hidden from shareholders exante, the nature of the optimal contract reveals the states expost, explaining why we assume s_t is observed. Setting $\alpha_0 \equiv 1$ simply normalizes the utility level from leaving the firm. Our empirical investigation demonstrates how our analysis of identification readily extends to a cross section or a panel, where $f_s(x)$ is unknown and w_t is measured with error. This is why we now focus on identifying the two mappings $g_s(x)$ plus the constants α_1, α_2 and ρ.

Our identification approach follows the work of Hurwicz (1950) and Koopmans and Reierol (1950). Specifically we use the results from Hurwicz (1950) that states that any structural characteristics, θ^* in our case, is identified relative to a structure, Θ in our case, if it can be written as a functional of the conditional distribution of the observed variables. If there does not exist such a functional, then the correspondence from the conditional distribution of the observed variables defines an equivalent class.

We begin with the caveat that moral hazard models, both pure and hybrid, can only be identified when the manager's compensation varies with the firm's abnormal return. When the solution to the model sets $w_s(x) = w^o$ for some or all states $s \in \{1, \ldots, S\}$, it is impossible to tell, in the absence of further detail, whether shareholders have used resources to implement a monitoring technology to remove moral hazard, whether it is unprofitable to implement a contract that induces diligence rather than acquiece to shirking, or whether there is no moral hazard problem, perhaps because expost settling up administrative and legal provisionis can achieve a first best outcome. Of necessity our analysis is concerned with identifying models of moral hazard when the contract calls for pay to be tied to the firm's abnormal returns.

To facilitate the discussion we partition the parameter space into Θ_1 and Θ_2, with generic elements $\rho \in \Theta_1$, the manager's coefficient of absolute risk aversion, and $\theta_2 \equiv (\alpha_1, \alpha_1, g_1(x), g_2(x))$, which characterizes the nonpecuniary benefits to the manager and the costs of shirking to the firm. We suppose the sample is generated from a model with (true) parameter $\theta^* \equiv (\rho^*, \theta_2^*) \in \Theta$.

Our discussion exploits the property that the optimal contract can be expressed in terms of $v_s(x)$. We show that if ρ^* is known, then θ_2^* is identified from the first order condition of the optimization problem augmented by some regularity conditions on the $g_s(x)$ functions. Then using the participation constraint, we fully characterize the observational equivalence class of θ^* when $S = 1$, and establish conditions for identifying ρ^*, and thence θ^*, in our multiple state model of pure moral hazard and our two state model of hidden information. If the model has three or more states we can test whether there is hidden information or not, but as we show below, the existence of hidden information is an identifying assumption in a two state model.

5.1 Identifying models of pure moral hazard

Without loss of generality, and for notational convenience, suppose $s \in \{1, \ldots, S\}$. After substituting the solution for the Lagrange multiplier on the participation constraint $\eta_1 = 1$, the the first order condition in the pure moral hazard model simplifies to

$$v_s(x)^{-1} = 1 + \eta_{3s}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_s(x)\right]$$

for each of the S states. Evaluating the equation at the point $\overline{x}$ where $g_s(\overline{x}) = 0$ for all s, we obtain

$$v_s(\overline{x})^{-1} = 1 + \eta_{3s}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}$$

Also taking the expectation with respect to x yields

$$E\left[v_s(x)^{-1}\right] = 1 + \eta_{3s}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{3s}$$

Manipulating these three equations we can solve for the $g_s(x)$ probability density ratios, the nonpecuniary benefit ratio (α_2/α_1) and the parameters the shadow prices on the incentive compatibility constraints η_1 and η_2 given ρ.

Proposition 1 *If the optimal contract in a pure moral hazard model with S states is $w_s^*(x)$ for $s \in \{1, 2, \ldots, S\}$, then the nonpecuniary benefits of working are*

$$\alpha_2^* = \left[\int \exp\left[-\frac{\rho^* \lambda_{t+1} w_s^*(x)}{b_{t+1}}\right] f_s(x)\, dx\right]^{\frac{-b_t}{\lambda_t}}$$

the ratio between diligence and shirking is

$$\frac{\alpha_2^*}{\alpha_1^*} = \left(\frac{\exp\left[\rho^* w_s^*(\overline{x})\right] - (\alpha_2^*)^{\frac{\lambda_t}{b_t}}}{\exp\left[\rho^* w_s^*(\overline{x})\right] - E\left\{\exp\left[\rho^* w_s^*(x)\right]\right\}}\right)^{\frac{b_t}{\lambda_t}}$$

while the probability density ratios are

$$g_s^*(x) = \frac{\exp\left[\rho^* w_s^*(\overline{x})\right] - \exp\left[\rho^* w_s^*(x)\right]}{\exp\left[\rho^* w_s^*(\overline{x})\right] - E\left\{\exp\left[\rho^* w_s^*(x)\right]\right\}}$$

Proposition 1 asserts that if ρ^* is identified for each $s \in \{1, 2, \ldots, S\}$, then the α_1^*, α_2^* and the S functions g_s^* are too. A natural place to investigate the identification of ρ^* is the participation constraint. We begin with the remark that when $\alpha_2^* > 1$, meaning the nonpecuniary benefits of working do not fully compensate the manager for the total benefits of his alternative, and thus expected compensation is positive, then the data imply a lower bound for the risk aversion parameter ρ. To see this we define the S mappings

$$\psi_s(\rho) \equiv E_s\left[\exp\left(-\frac{\rho \lambda_{t+1} w_s^*(x)}{b_{t+1}}\right)\right]$$

From its definition $\psi_s(0) = 1$, while the assumption above implies

$$\psi_s'(0) = \frac{\partial}{\partial \rho} E_s\left[\exp\left(-\frac{\rho \lambda_{t+1} w_s^*(x)}{b_{t+1}}\right)\right]_{\rho=0} = -E_s\left[-\frac{\lambda_{t+1} w_s^*(x)}{b_{t+1}}\right] < 0$$

Also $\psi_s(\rho)$ is convex in ρ, because

$$\frac{\partial^2}{\partial \rho^2}\left[\exp\left(-\frac{\rho w_s^*(x)}{p_{t+1}}\right)\right] = \left(\frac{w_s^*(x)}{p_{t+1}}\right)^2 \exp\left(-\frac{\rho w_s^*(x)}{p_{t+1}}\right) > 0$$

and the expectations operator preserves convexity. Assuming $\alpha_2^* > 1$, it now follows that $\psi_s(\rho)$ crosses the unit level from below once at say ρ_s, which implies $\psi_s(\rho) > 1$ for all

$\rho > \rho_s$. This rules out the possibility that $\rho^* \leq \rho_{\max} = \max\{\rho_1, \ldots, \rho_S\}$. Intuitively, the participation equation is satisfied by different (ρ, α_2) satisfying $\rho > \rho_0$and

$$\alpha_2 = \psi_s(\rho)^{-\frac{b_t}{\lambda_t}}$$

as we see in Figure 1. Along this line as ρ increases, the person becomes more risk averse, the expected utility from $w^*(x)$ declines along with its the certainty equivalent, but this is just offset by nonpecuniary amenities from the job. Consequently an observer with data on $\{w^*(x_t)\}_{t=1}^T$ cannot distinguish between someone with a high risk tolerance and unpleasant working conditions, versus a person with a lower tolerance but more nonpecuniary benefit. The remaining parameters are then identified from the value ascribed to ρ, the slope of the contract with respect to abnormal returns determining $g_s(x)$ and thence the probability distribution of abnormal returns under shirking.

This intuition on compensating variation for work with risky returns, carries over to models of pure moral hazard, leading into to our main result on observational equivalence in the pure moral hazard model. Suppose there is only a single state. Dropping the s subscripts, let $w^*(x)$ be the optimal compensation schedule for the parameterization $\theta^* \equiv (\rho^*, \alpha_1^*, \alpha_2^*, g^*(x))$. Proposition 2 below displays a class of pure moral hazard models, indexed by ρ, with the common framework described above, differing only in parameter values, that are observationally equivalent to each other.

Proposition 2 *For each $\widehat{\rho} > \rho_0^*$, respectively define the parameters $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, and the mapping $\widehat{g}(x)$, as*

$$\widehat{\alpha}_1 = \psi(\widehat{\rho})\left[\frac{\exp[\widehat{\rho}w^*(\overline{x})] - 1}{\exp[\widehat{\rho}w^*(\overline{x})] - E\{\exp[\widehat{\rho}w^*(x)]\}}\right]$$

$$\widehat{\alpha}_2 = [\varphi(\rho)]^{-\frac{\lambda_t}{b_t}}$$

$$\widehat{g}(x) = \frac{\exp[\widehat{\rho}w^*(\overline{x})] - \exp[\widehat{\rho}w^*(x)]}{\exp[\widehat{\rho}w^*(\overline{x})] - E\{\exp[\widehat{\rho}w^*(x)]\}}$$

The parameter $\widehat{\theta} \equiv (\widehat{\rho}, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{g}(x))$ is observationally equivalent to θ^*.

Proposition 1 previously established that if ρ^* is identified then so is θ_1^*. We now provide conditions for identifying ρ^* in models of pure moral hazard with at least two states. Noting first that since the participation condition holds for each state for each $s \in S$ separately, we can apply the moment conditions

$$\int \exp[-\rho w_r^*(x)] f_r(x) dx = \int \exp[-\rho w_s^*(x)] f_s(x) dx$$

which of course is solved by ρ^*. Although there may be multiple roots in ρ to the equations defined by the separate states $r \in S$ and $s \in S$, if there is a unique root common to all possible pairs, then ρ^* is identified.

Proposition 3 *A necessary and sufficient condition for identifying θ^* is that ρ^* uniquely solves*

$$\int \exp\left[-\rho w_r^*(x)\right] f_r(x)\, dx = \int \exp\left[-\rho w_s^*(x)\right] f_s(x)\, dx$$

across all possible pairs of states $r \in S$ and $s \in S$.

This condition is analogous to the principle of competitive selection. The principle is that two jobs offering the same nonpecuniary benefits but with different wages from probability distributions must equalize expected utility to attract the same applicants, and that equality provides a means for identifying ρ^*. Similarly models of pure moral hazard are identified if there is competitive selection between two states that have different compensation plans $w_r^*(x)$ and $w_s^*(x)$ but the same nonpecuniary benefits from diligent work α_2^*. In this case $w_r^*(x) \neq w_s^*(x)$ because the probability density function of abnormal returns from working diligently differs by state, that is $f_r(x) \neq f_s(x)$, or the density from shirking differs, that is $f_r(x)\, g_r(x) \neq f_s(x)\, g_s(x)$.

5.2 Identifying models of hybrid moral hazard

We now extend our analysis to models with hidden information. As in pure moral hazard models, we first investigate conditions for identifying ρ^*, and then analyze the identification of the remaining parameters $\theta_2^* \equiv (\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$. In contrast the model of pure moral hazard, we establish over-identifying restrictions in the second stage, that is conditional on identifying ρ^*. Since models of pure moral hazard are exactly identified in the absence of further functional form restrictions, a natural interpretation for rejecting over-identifying restrictions in the generalized model is that under the alternative hypothesis there is no hidden information and the true model is pure moral hazard.

In the generalized model there is only one participation constraint across states, rather than a participation constraint for each state, thus reducing the number of restrictions. Furthermore, unless the optimal contract is explicitly solved to obtain the Lagrange multipliers as a mapping of θ as indicated in the previous section, models of generalized moral hazard the optimal contract introduce new parameters, namely the shadow price of the truth telling constraint, η_2, and the shadow price of the combination constraint η_4. However the dependence of the optimal contract on η_2 and η_4 introduces other restrictions that can be exploited in identification. Similar to the approach taken in the pure moral hazard case we now show that if ρ^* is known, then the remaining structural parameters $\theta^* \equiv (\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$ can be identified too.

It is convenient to partition the discussion into two parts, depending on whether the combination constraint is binding or not, a condition that can be directly tested if ρ^* is known. If combination constraint is nonbinding, the shadow price of truth telling, η_2^*, can be identified from either first order condition $s \in \{1, 2\}$ by first multiplying through by $v_s(x)$, taking expectations over x conditional on the state s, and using the fact that the incentive compatibility constraint is met with strict equality. Making η_2^* the subject of both equations we obtain an over-identifying restriction, in the form of an equality that ρ^* must satisfy. Evaluating the first order conditions for both states at $\overline{x}$, integrating over the first order condition with respect to x, and manipulating the resulting equations in much the same way we did in the pure moral hazard case yields an expression for θ_2^* that establishes its identification, that is given the true value of ρ^*. If the combination constraint is binding, then its shadow price is positive, and consequently affects the mapping from the structural parameters to the optimal contract, but given ρ^*, we can nevertheless prove by construction there exists a mapping from $(w_1^*(x), w_2^*(x))$ to θ_2^*.

Proposition 4 *Suppose ρ^* is known, and define the mapping $v_s^*(x)$ as*

$$v_s^*(x) = \exp\left[\frac{-\rho^* \lambda_{t+1} w_s^*(x)}{b_{t+1}}\right]$$

for $s \in \{1, 2\}$. Then the remaining parameters $\theta_2^* \equiv (\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$ are identified from the optimal contract $(w_1^*(x), w_2^*(x))$, by the mappings

$$\alpha_2^* = \left\{\sum\nolimits_{s=1}^{2} \varphi_s E_s\left[v_s^*(x)\right]\right\}^{\frac{\lambda_{t+1}}{b_{t+1}}}$$

$$g_2^*(x) = \frac{v_2^*(\overline{x})^{-1} - v_2^*(x)^{-1}}{v_2^*(\overline{x})^{-1} - E\left[v_2^*(x)^{-1}\right]}$$

$$(\alpha_2^*/\alpha_1^*)^{\frac{\lambda_t}{b_t}} = \frac{v_2^*(\overline{x})^{-1} - \left\{E_2\left[v_2^*(x)\right]\right\}^{-1}}{v_2^*(\overline{x})^{-1} - E\left[v_2^*(x)^{-1}\right]}$$

$$g_1^*(x) = \frac{v_1^*(x)^{-1} - v_1^*(\overline{x})^{-1} + \left[\left\{E_2\left[v_2^*(x)\right]\right\}^{-1} - 1 - \eta_4^*\right]\left[h(x) - h(\overline{x})\right] + \eta_4^*(\alpha_1^*/\alpha_2^*)^{\frac{\lambda_t}{b_t}} g_2^*(x)\, h(x)}{E_1\left[v_1^*(x)^{-1}\right] - v_1^*(\overline{x})^{-1} + \left[\left\{E_2\left[v_2^*(x)\right]\right\}^{-1} - 1 - \eta_4^*\right]\left[h(x) - \varphi_2/\varphi_1\right] + \eta_4^*(\alpha_1^*/\alpha_2^*)^{\frac{\lambda_t}{b_t}} E_1\left[g_2^*(x)\, h(x)\right]}$$

where

$$\eta_4^* = \frac{E_1\left[v_1^*(x)\right] + E_1\left[v_1^*(x)\, h(x)\right]\left\{E_2\left[v_2^*(x)\right]\right\}^{-1} + E_1\left[v_1^*(x)\, h(x)\right] - 1}{E_1\left[v_1^*(x)\, h(x)\right] + (\alpha_1^*/\alpha_2^*)^{\frac{\lambda_t}{b_t}} E_1\left[v_1^*(x)\, g_2(x)\, h(x)\right]}$$

Proposition 4 proves that θ_2^* is identified if ρ^* is known. We now focus our attention on identifying this parameter of risk aversion. In the generalized models of moral hazard, there are by definition at least two states, and the associated truth telling constraints help identify ρ in a similar way that the participation constraint helps identify that parameter in models of risk sharing and moral hazard. The truth telling constraint in the hybrid moral hazard framework yields an equation in ρ that is similar to the competitive selection equations that exist when both states are observed in the model of pure moral hazard. Defining

$$\Psi_1(\rho) \equiv E_2\left[\exp\left(-\frac{\rho\lambda_{t+1} w_2^*(x)}{b_{t+1}}\right) - \exp\left(-\frac{\rho\lambda_{t+1} w_1^*(x)}{b_{t+1}}\right)\right]$$

the truth telling constraint in the generalized moral hazard framework yields an equation in ρ^* that is similar to the competitive selection equations that exist when both states are observed in the model of pure moral hazard, namely $\Psi_1(\rho^*) = 0$.

The combination constraint also yields a restriction on the data, regardless of whether it binds or not. To see this let

$$\begin{aligned}\Psi_2(\rho) \equiv\ & E_2\left[\exp\left(\frac{\rho\lambda_{t+1}\left[w_2^*(\overline{x})-w_1^*(x)\right]}{b_{t+1}}\right)-\exp\left(\frac{\rho\lambda_{t+1}\left[w_2^*(x)-w_1^*(x)\right]}{b_{t+1}}\right)\right] \\ & +E_2\left[\exp\left(-\frac{\rho\lambda_{t+1}w_2^*(x)}{b_{t+1}}\right)\left\{E_2\left[-\frac{\rho\lambda_{t+1}w_2^*(x)}{b_{t+1}}\right]\right\}^{-1}-\exp\left(\frac{\rho\lambda_{t+1}\left[w_2^*(\overline{x})-w_2^*(x)\right]}{b_{t+1}}\right)\right]\end{aligned}$$

The proof to the next proposition demonstrates that the combination constraint can be expressed by the inequality $\Psi_2(\rho^*) \geq 0$, which limits the values of ρ that are observationally equivalent. Furthermore if the combination constraint is not binding, that is $\Psi_2(\rho^*) > 0$ then its associated multiplier, η_4^*, is zero. Multiplying the s^{th} first order condition through by $v_s(x)$, taking conditional expectations and then rearranging we obtain two expressions for the shadow price of hidden information η_2^*, a restriction on the risk aversion parameter that we now express as $\Psi_3(\rho^*) = 0$ where

$$\Psi_3(\rho) = \{1 - E_1[v_1^*(x)]\} E_2[v_2^*(x)] - E_1[v_1^*(x)\, h(x)] \{1 - E_2[v_2^*(x)]\}$$

Proposition 5 consolidates these results in the form of sufficient conditions for identifying the triplet ρ^*.

Proposition 5 *The risk aversion parameter ρ^* is identified if there is only one value of ρ that simultaneously satisfies the truth telling contraint $\Psi_1(\rho^*) = 0$, the complementary slackness condition for the combination constraint $\Psi_2(\rho^*)\Psi_3(\rho^*) = 0$ and the inequality $\Psi_2(\rho^*) \geq 0$.*

This only leaves open the question whether the solutions to the Lagrange multipliers yield additional restrictions that can be exploited in identification or testing. More generally, are there other restrictions implied by the optimal contract that can further narrow the subset of $\rho \in \Theta_1$ satisfying the conditions in the previous proposition? As in the pure moral hazard model we answer this question in the negative. Every $\widehat{\rho} \in \Theta_1$ satisfying the conditions in the previous proposition induces a $\widehat{\theta}_2 \in \Theta_2$ for which $(w_1^*(x), w_2^*(x))$ is the optimal contract to $\widehat{\theta} \equiv \left(\widehat{\rho}, \widehat{\theta}_2\right)$. Consequently θ^* and $\widehat{\theta}$ are observationally equivalent. Thus Propositions 4 through 6 fully characterize identification in this class of moral hazard and hidden information models.

Proposition 6 *Define the mapping $\theta_2(\rho)$ from Proposition 4. Every $\widehat{\rho} \in \Theta_1$ satisfying the conditions of Proposition 5, induces a $\widehat{\theta} = (\widehat{\rho}, \theta_2(\widehat{\rho})) \in \Theta$ such that $\widehat{\theta}$ is observationally equivalent to (the true) $\theta^* \in \Theta$.*

6 Estimation and Testing

We first analyze the basis for testing models of pure and hybrid moral hazard, and then explain our approach to identification, testing and estimation.

6.1 The empirical content of moral hazard models

There are many ways to test whether data from a vector sequence $\{s_t, x_t, w_t\}_{t=1}^{\infty}$ are generated by the class of moral hazard models examined in this paper. For example one simple specification test is whether the functional relations satisfy certain exclusion restrictions, such as whether w_t depends on x_{t-1}. While we do conduct several of these omnibus tests in our empirical application on executive compensation below, we have nothing to contribute at an abstract level. Our point of departure is to focus on regular data generating processes. We call a data generating process regular, and denote it by $\{p(s\,|s_{t-1}, x_{t-1}), f_s(x), w_s^*(x)\}_{s=1}^{S}$ for $(s, x) = (s_t, x_t)$, if s_t is sequentially generated a Markov probability transition for the states denoted by $p(s\,|s_{t-1}, x_{t-1})$, if x_t is an independently distributed random variable with conditional density $f_s(x)$, and if realized compensation $w_t = w_s^*(x)$ for some real valued mapping $w_s^*(x)$ defined on the space of states and abnormal returns. We analyze whether a regular data generating process comes from a model of pure moral hazard, a model of hybrid moral hazard, both, or neither. The constraints on the shareholders cost minimization problem ensuring participation, incentive compatibility, truth telling and sincerity, provide a set of restrictions on the risk aversion parameter ρ that apply to the models of moral hazard considered in this paper, but need not apply more generally to regular data generating processes. This set of restrictions provides the basis for the test statistics we derive below.

Consider first the model of pure moral hazard with S states. Defining

$$\Psi_{4s}(\rho) \equiv E_s\left[\exp\left(-\frac{\rho\lambda_{t+1}w_s^*(x)}{b_{t+1}}\right)\right] - E_1\left[\exp\left(-\frac{\rho\lambda_{t+1}w_1^*(x)}{b_{t+1}}\right)\right]$$

for each $s \in \{1, \ldots, S\}$, the participation constraint implies that in a pure moral hazard with risk aversion parameter ρ^*, we have $\Psi_{4s}(\rho^*) = 0$. We might consider rejecting the pure moral hazard if there does not exist some $\rho > 0$ solving $\Psi_{4s}(\rho) = 0$ for all $s \in \{1, \ldots, S\}$. Even in this case, however, one could trivially extend the pure moral hazard framework to replace α_2 with state specific nonpecuniary benefits α_{2s} that dissolve the identifying restrictions, and with it the possibility of testing the pure moral hazard framework. In other words models of pure moral hazard can only be tested by placing restrictions on the nonpecuniary benefits from working diligently in the different states, which may or may not be plausible depending on the particular application.

Second we note that in a two state hybrid model $\Psi_1(\rho^*) = 0$. In models of pure moral hazard $\Psi_1(\rho^*) \neq 0$ generically. If $\Psi_1(\rho^*) > 0$ there is no incentive for the manager to lie about the state of the world even though it is unobserved. On the other hand if $\Psi_1(\rho^*) < 0$, the manager prefers shareholders to be mistaken about the state, so the fact that shareholders do nothing to deter him from truthfully announcing it might merely indicate that they directly observe the state. It follows that if $\Psi_1(\rho^*) \neq 0$ for all $\rho > 0$, the hybrid model is rejected by the alternative of a pure moral hazard model.

Third we note $\Psi_2(\rho^*) \geq 0$ in the hybrid model, and if $\Psi_2(\rho^*) = 0$ then the combination constraint binds. Our preceding analysis also implies $\Psi_3(\rho^*) = 0$ under the hybrid model if the combination constraint does not bind. Hence $\Psi_2(\rho^*)\,\Psi_3(\rho^*) = 0$ in the hybrid model. Generically $\Psi_2(\rho^*)\,\Psi_3(\rho^*) \neq 0$ in a model of pure moral hazard with risk aversion parameter ρ^*.

We remark that one could construct a pure moral hazard model where $\Psi_2(\rho^*)\Psi_3(\rho^*) = 0$. However the constructed model is not generic. In other words there exist small perturbations in the structural parameters that would break the equality $\Psi_2(\rho^*)\Psi_3(\rho^*) = 0$. Similarly pure moral hazard models can be constructed so that $\Psi_1(\rho^*) = 0$, but they are not generic. For this reason we interpret as evidence against the null hypothesis of the hybrid model, in favor of the alternative hypthesis of the pure moral hazard model, a finding that there is no $\rho^* > 0$ simultaneously satisifying the two equalities $\Psi_1(\rho^*) = \Psi_2(\rho^*)\Psi_3(\rho^*) = 0$ along with the weak inequality $\Psi_2(\rho^*) \geq 0$.

The crucial role the $\Psi_1(\rho)$ through $\Psi_{4s}(\rho)$ mappings in constructing test statistics is evident from the next proposition. It shows that within the class of regular data generating processes, there are no restrictions to base specification tests of pure and hybrid models apart from the sort considered here. More precisely we show that every regular data generating structure can be interpreted as a model of pure moral hazard, and also a model of hybrid model of moral hazard, providing there is at least one value of ρ satisfying the conditions given above. To prove this proposition we extend our previous results on observational equivalence beyond the class of moral hazard models.

Proposition 7 *Consider any regular data generating process* $\{p(s'|s,x), f_s(x), w_s^*(x)\}_{s=1}^{S}$.

1. *Given S states denoted by $s \in \{1,\ldots,S\}$, assume there exists a real number $\gamma > 0$ satisfying the S restrictions $\Psi_{4s}(\gamma) = 0$ for each s. Then the pure moral hazard model with risk parameter γ and parameter vector $\theta_1(\gamma)$ is observationally equivalent to $\{p(s'|s,x), f_s(x), w_s^*(x)\}_{s=1}^{S}$.*

2. *Given 2 states, suppose there exists a real number $\gamma > 0$ simultaneously satisfying the two equations $\Psi_1(\gamma) = \Psi_2(\gamma)\Psi_3(\gamma) = 0$ together with the weak inequality $\Psi_2(\gamma) \geq 0$. Then a hybrid moral hazard model with risk parameter γ and parameter vector $\theta_2(\gamma)$ is observationally equivalent to $\{p(s'|s,x), f_s(x), w_s^*(x)\}_{s=1}^{S}$.*

6.2 Empirical implementation

As indicated earlier, we assume that data on compensation is measured with error that is orthogonal to the firms abnormal returns. We first estimate the wage regression functions $w_s^*(x)$ for $s \in \{1,\ldots,S\}$ point-wise with a Kernel estimator taking the form

$$w_j^{(N)}(x) = \frac{\sum_{n=1}^{N} w_n 1\{s_n = j\} K\left(\frac{x-x_n}{\delta_N}\right)}{\sum_{n=1}^{N} 1\{s = j\} K\left(\frac{x-x_n}{\delta_N}\right)}$$

where $K(\cdot)$ is a univariate probability density function with full support and δ_N is the bandwidth satisfying the convergence property $\delta_N \to 0$ as $N \to \infty$. To estimate $w_j(\overline{x})$, we use the fact that although $\overline{x}$ is unknown, $w_j(x)$ is a locally nondecreasing function in x. Following Brunk (1958), for each state $s \in \{1,\ldots,S\}$ we rank the observations on returns in decreasing order by $x_{s1},\ x_{s2},\ldots$ and so on, denoting by $w_{s1},\ w_{s2},\ldots$ the corresponding compensations, and estimate $w_j(\overline{x})$ with $w_j^{(N)}$ defined as

$$w_s^{(N)} \equiv \max_q \sum\nolimits_{r=1}^{q} \frac{w_{sr}}{q}$$

Then we form the empirical counterparts to $\Psi_{1s}(\rho)$ through $\Psi_4(\rho)$, which we denote by $\Psi_{1s}^{(N)}(\rho)$ through $\Psi_4^{(N)}(\rho)$ by substituting estimates of $w_s^*(x)$ into the sample moments of real valued functions defined on the space of abnormal returns $\tau(x)$ as

$$E_j^{(N)}[v(x)] = \frac{\sum_{n=1}^{N} 1\{s_n = j\}\,\tau(x_n)}{\sum_{n=1}^{N} 1\{s_n = j\}}$$

Thus our empirical analogue to $\Psi_2(\rho)$ is defined as

$$\Psi_2^{(N)}(\rho) \equiv E_2^{(N)}\left[\exp\left(-\frac{\rho\lambda_{t+1}w_2^{(N)}(x)}{b_{t+1}}\right) - \exp\left(-\frac{\rho\lambda_{t+1}w_1^{(N)}(x)}{b_{t+1}}\right)\right]$$

and the other analogues $\Psi_{1s}^{(N)}(\rho)$ through $\Psi_4^{(N)}(\rho)$ are formed in similar fashion, by substituting Kernel estimators for the wage regression functions into sample moments that approximate their corresponding population moments.

Since each of these tests only involve estimating a function that depends on parameter we can graph their results and investigate their roots by constructing confidence intervals.

7 Appendix

Proof of Proposition 1. After substituting the solution for the Lagrange multiplier on the participation constraint $\eta_1 = 1$, the the first order condition in the pure moral hazard model simplifies to

$$v_s(x)^{-1} = 1 + \eta_{3s}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_s(x)\right]$$

for each of the S states. Evaluating the equation at the boundary point $\overline{x}$ where $g(\overline{x}) = 0$ we obtain

$$v_s(\overline{x})^{-1} = 1 + \eta_{3s}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}$$

Also taking the expectation with respect to x yields

$$E\left[v_s(x)^{-1}\right] = 1 + \eta_{3s}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - \eta_{3s}$$

From the first two equations

$$\eta_{3s}g_s(x) = v_s(\overline{x})^{-1} - v_s(x)^{-1}$$

From the second two equations

$$\eta_{3s} = v_s(\overline{x})^{-1} - E\left[v_s(x)^{-1}\right]$$

Solving for $g_s(x)$ and A_1 in terms of $v_s(x)^{-1}$, and then substituting the definition of $v_s(x)^{-1}$ into the resulting expressions we obtain:

$$g_s(x) = \frac{\exp[\rho w_s(\overline{x})] - \exp[\rho w_s(x)]}{\exp[\rho w_s(\overline{x})] - E\{\exp[\rho w_s(x)]\}}$$

Also from the top equation and the equation for η_{3s} we obtain

$$\begin{aligned}
(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} &= \frac{v_s(\overline{x})^{-1} - 1}{\eta_{3s}} \\
&= \frac{v_s(\overline{x})^{-1} - 1}{v_s(\overline{x})^{-1} - E\left[v_s(x)^{-1}\right]} \\
&= \frac{\exp[\rho w_s(\overline{x})] - (\alpha_2/\alpha_0)^{\frac{\lambda_t}{b_t}}}{\exp[\rho w_s(\overline{x})] - E\{\exp[\rho w_s(x)]\}}
\end{aligned}$$

as required. ■

Proof of Proposition 2. Since $\exp[\rho w^*(\overline{x})] \geq \exp[\rho w^*(x)]$ for all x and $\exp[\rho w^*(\overline{x})] > E\{\exp[\rho w^*(x)]\}$, it immediately follows that $g(x) \geq 0$. Integrating over x, establishes $E[g(x)] = 1$, and evaluating $g(x)$ at $\overline{x}$ demonstrates $g(\overline{x}) = 0$. Therefore $g(x)$ satisfies the properties imposed upon $g(x) \in G$.

Next we define the Lagrange multiplier associated with the incentive compatibility condition for the model defined by $\widehat{\theta}$ as

$$\eta(\rho) = [\alpha_0(\rho)/\alpha_2(\rho)]^{\frac{\lambda_t}{b_t}} (\exp[\rho w^*(\overline{x})] - E\{\exp[\rho w^*(x)]\})$$

From the definitions of $\widehat{\alpha}_1 \equiv \alpha_1(\widehat{\rho})$ and $\widehat{\alpha}_2 \equiv \alpha_2(\widehat{\rho})$:

$$\begin{aligned}
(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}(x) &= \frac{\exp[\rho w^*(\overline{x})] - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}}{\exp[\rho w^*(\overline{x})] - E\{\exp[\rho w^*(x)]\}} - \frac{\exp[\rho w^*(\overline{x})] - \exp[\rho w^*(x)]}{\exp[\rho w^*(\overline{x})] - E\{\exp[\rho w^*(x)]\}} \\
&= \frac{\exp[\rho w^*(x)] - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}}{\exp[\rho w^*(\overline{x})] - E\{\exp[\rho w^*(x)]\}}
\end{aligned}$$

Multiply through by $\widehat{\eta} = \eta(\widehat{\rho})$ and rearrange to obtain

$$\begin{aligned}
1 + \widehat{\eta}\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}(x)\right] &= 1 + (\widehat{\alpha}_0/\widehat{\alpha}_2)^{\frac{\lambda_t}{b_t}}\left[\exp[\rho w^*(x)] - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}\right] \\
&= (\widehat{\alpha}_0/\widehat{\alpha}_2)^{\frac{\lambda_t}{b_t}} \exp[\widehat{\rho} w^*(x)]
\end{aligned}$$

which proves that $w^*(x)$ satisfies the first order condition for the $\widehat{\theta}$ parameterization.

Turning to the incentive compatibility condition, and substituting for the definitions of

$(\widehat{\alpha}_2/\widehat{\alpha}_0)$ and $\widehat{g}(x)$ we see that

$$\begin{aligned}
& E\left\{\left[\widehat{g}(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}\right](\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}\exp\left[-\widehat{\rho}w^*(x)\right]\right\} \\
= \; & (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}} E\left\{\left[\frac{\exp\left[\widehat{\rho}w^*(x)\right] - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}}{\exp\left[\widehat{\rho}w^*(\overline{x})\right] - E\left\{\exp\left[\widehat{\rho}w^*(x)\right]\right\}}\right]\exp\left[-\widehat{\rho}w^*(x)\right]\right\} \\
= \; & \frac{(\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}}{\exp\left[\widehat{\rho}w^*(\overline{x})\right] - E\left\{\exp\left[\widehat{\rho}w^*(x)\right]\right\}}\left\{1 - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}} E\left[\exp\left(-\widehat{\rho}w^*(x)\right)\right]\right\} \\
= \; & 0
\end{aligned}$$

the fourth line following from the participation constraint.

Note that $\eta(\rho) > 0$, as required by the definition of the multiplier in the minimization problem, since $\exp\left[\rho w^*(\overline{x})\right] > E\left\{\exp\left[\rho w^*(x)\right]\right\}$. Also note that

$$\begin{aligned}
E\left[\frac{\widehat{g}(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}}{1 + \widehat{\eta}\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}(x)\right]}\right] & \equiv \widehat{\eta}^{-1} E\left[\frac{\widehat{\eta}\left[\widehat{g}(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}\right]}{1 + \widehat{\eta}\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}(x)\right]}\right] \\
& = \widehat{\eta}^{-1} E\left[\frac{\exp\left[\rho w^*(x)\right] - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}}{(\widehat{\alpha}_0/\widehat{\alpha}_2)^{\frac{\lambda_t}{b_t}}\exp\left[\widehat{\rho}w^*(x)\right]}\right] \\
& = \widehat{\eta}^{-1}(\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}}\left[1 - (\widehat{\alpha}_2/\widehat{\alpha}_0)^{\frac{\lambda_t}{b_t}} E\left\{\exp\left[-\widehat{\rho}w^*(x)\right]\right\}\right] \\
& = 0
\end{aligned}$$

It now follows that this definition of $\widehat{\eta} = \eta(\widehat{\rho})$ satisfies the definition of a the Lagrange multiplier for the optimization problem.

It now follows that $\widehat{\theta}$ and $\eta(\rho)$ jointly satisfy the first order condition for every realization (x_t, w_t) based on the original parameterization θ^*, and also satisfy the participation constraint as and incentive. By construction $\widehat{\theta}$, specifically $\widehat{\rho}$ and $\widehat{\alpha}_2$ satisfy the participation constraint. The objective function for optimization function is globally concave and the constraints are linear, so there is a unique optimum determined by the first order condition and the constraints. As $w^*(x)$ satisfies the first order condition and the three constraints then it uniquely solves the optimal contract. Since these conditions uniquely define the solution to this strictly convex minimization problem, it follows that the compensation schedule for $\widehat{\theta}$ is $w^*(x)$. This completes the proof. ■

Proof of Proposition 4. Recall that the first order condition for the second state is

$$v_2(x)^{-1} = 1 + \eta_2 + \eta_{32}[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_2(x)] + \eta_4$$

Taking expectations we obtain

$$E\left[v_2(x)^{-1}\right] = 1 + \eta_2 + \eta_{32}[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - 1] + \eta_4$$

Also

$$v_2(\overline{x})^{-1} = 1 + \eta_2 + \eta_{32}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} + \eta_4$$

Differencing the second two equations

$$v_2(\overline{x})^{-1} - E\left[v_2(x)^{-1}\right] = \eta_{32}$$

Differencing the first and the third gives

$$v_2(\overline{x})^{-1} - v_2(x)^{-1} = \eta_{32} g_2(x)$$

Taking the quotient yields the solution

$$g_2(x) = \frac{v_2(\overline{x})^{-1} - v_2(x)^{-1}}{v_2(\overline{x})^{-1} - E\left[v_2(x)^{-1}\right]}$$

Substituting the solution for η_{32} into the first order condition evaluated at $\overline{x}$ shows that

$$(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} = \frac{v_2(\overline{x})^{-1} - 1 - \eta_2 - \eta_4}{v_2(\overline{x})^{-1} - E\left[v_2(x)^{-1}\right]}$$

Now rewrite the first order condition for the first state as

$$v_1(x)^{-1} - 1 + \eta_2 h(x) + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x) h(x) = \eta_{31}\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_1(x)\right]$$

At $\overline{x} = x$ we have

$$v_1(\overline{x})^{-1} - 1 + \eta_2 h(\overline{x}) = \eta_{31}(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}}$$

Differencing the expressions yields

$$v_1(x)^{-1} - v_1(\overline{x})^{-1} + \eta_2\left[h(x) - h(\overline{x})\right] + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x) h(x) = \eta_{31} g_1(x)$$

and we obtain by taking expectations

$$E_1\left[v_1(x)^{-1}\right] - v_1(\overline{x})^{-1} + \eta_2\left[h(x) - \phi_2/\phi_1\right] + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} E_1\left[g_2(x) h(x)\right] = \eta_{31}$$

since $E_1[h(x)] = \phi_2/\phi_1$. Taking the quotient we obtain the solution for $g_1(x)$, namely

$$g_1(x) = \frac{v_1(x)^{-1} - v_1(\overline{x})^{-1} + \eta_2\left[h(x) - h(\overline{x})\right] + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} g_2(x) h(x)}{E_1\left[v_1(x)^{-1}\right] - v_1(\overline{x})^{-1} + \eta_2\left[h(x) - \phi_2/\phi_1\right] + \eta_4 (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} E_1\left[g_2(x) h(x)\right]}$$

The equations above give $g_1(x)$ as a mapping of ρ and x, and the remaining parameters as a function of η_2 and η_4 as well.

We now solve for η_2 in terms of η_4. First multiply the first order conditions for the second state by $v_2(x) f_2(x)$, after solving for $\eta_1 = 1$ to obtain

$$f_2(x) = v_2(x) f_2(x) + \eta_2 v_2(x) f_2(x) + \eta_{32} v_2(x)\left[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_2(x)\right] f_2(x) + \eta_4 v_2(x) f_2(x)$$

Integrating over x, and noting the incentive compatibility constraint is satisfied with equality in both states, yields

$$1 = E_2\left[v_2\left(x\right)\right] + \eta_2 E_2\left[v_2\left(x\right)\right] + \eta_4 E_2\left[v_2\left(x\right)\right]$$

We divide the result by $E_2\left[v_2\left(x\right)\right]$ and make η_2 the subject and obtain

$$\eta_2 = \left\{E_2\left[v_2\left(x\right)\right]\right\}^{-1} - 1 - \eta_4$$

Substituting $\left\{E_2\left[v_2\left(x\right)\right]\right\}^{-1}$ for $1+\eta_2+\eta_4$ in the expression derived for $\left(\alpha_2/\alpha_1\right)^{\frac{\lambda_t}{b_t}}$ now yields the equation for $\left(\alpha_2/\alpha_1\right)^{\frac{\lambda_t}{b_t}}$ given in the statement of the proposition. Similarly with regards to $g_1\left(x\right)$ we have

$$g_1\left(x\right) = \frac{v_1\left(x\right)^{-1} - v_1\left(\overline{x}\right)^{-1} + \left[\left\{E_2\left[v_2\left(x\right)\right]\right\}^{-1} - 1 - \eta_4\right]\left[h(x) - h\left(\overline{x}\right)\right] + \eta_4\left(\alpha_1/\alpha_2\right)^{\frac{\lambda_t}{b_t}} g_2\left(x\right)h\left(x\right)}{E_1\left[v_1\left(x\right)^{-1}\right] - v_1\left(\overline{x}\right)^{-1} + \left[\left\{E_2\left[v_2\left(x\right)\right]\right\}^{-1} - 1 - \eta_4\right]\left[h(x) - \phi_2/\phi_1\right] + \eta_4\left(\alpha_1/\alpha_2\right)^{\frac{\lambda_t}{b_t}} E_1\left[g_2\left(x\right)h\left(x\right)\right]}$$

A similar procedure can be applied to the first state to solve for η_4. First multiply the first order conditions by $v_1\left(x\right)f_1\left(x\right)$, after solving for $\eta_1 = 1$ to obtain

$$\begin{aligned} f_1\left(x\right) &= v_1\left(x\right)f_1\left(x\right) - \eta_2 v_1\left(x\right)h(x)f_1\left(x\right) + \eta_{31}v_1\left(x\right)\left[\left(\alpha_2/\alpha_1\right)^{\frac{\lambda_t}{b_t}} - g_1\left(x\right)\right]f_1\left(x\right) \\ &\quad -\eta_4(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} v_1\left(x\right)g_2\left(x\right)h\left(x\right)f_1\left(x\right) \end{aligned}$$

Then

$$1 = E_1\left[v_1\left(x\right)\right] - \eta_2 E_1\left[v_1\left(x\right)h(x)\right] - \eta_4(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} E_1\left[v_1\left(x\right)g_2\left(x\right)h\left(x\right)\right]$$

Substituting out the solution for η_2 we obtained from manipulating the first order condition from the second state, we get

$$\begin{aligned} 1 &= E_1\left[v_1\left(x\right)\right] - E_1\left[v_1\left(x\right)h(x)\right]\left\{E_2\left[v_2\left(x\right)\right]\right\}^{-1} - E_1\left[v_1\left(x\right)h(x)\right] \\ &\quad -\eta_4 E_1\left[v_1\left(x\right)h(x)\right] - \eta_4(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} E_1\left[v_1\left(x\right)g_2\left(x\right)h\left(x\right)\right] \end{aligned}$$

Solving for η_4 we now have

$$\eta_4 = \frac{E_1\left[v_1\left(x\right)\right] + E_1\left[v_1\left(x\right)h(x)\right]\left\{E_2\left[v_2\left(x\right)\right]\right\}^{-1} + E_1\left[v_1\left(x\right)h(x)\right] - 1}{E_1\left[v_1\left(x\right)h(x)\right] + \left(\alpha_1/\alpha_2\right)^{\frac{\lambda_t}{b_t}} E_1\left[v_1\left(x\right)g_2\left(x\right)h\left(x\right)\right]}$$

■

Proof of Proposition 5. First we show that $C\left(\rho^*\right) \geq 0$. In the proof of Proposition 4 we showed that

$$1 = E_2\left[v_2\left(x\right)\right] + \left(\eta_2^* + \eta_4^*\right)E_2\left[v_2\left(x\right)\right]$$

Substituting the expression implied for $\{E_2[v_2(x)]\}^{-1}$ into the definition of $C(\rho)$ we obtain

$$\begin{aligned}
C(\rho) &\equiv \int \left[v_1(x)\left[v_2^{-1}(\overline{x}) - v_2^{-1}(x)\right] - v_2(x)\left[v_2^{-1}(\overline{x}) - \{E_2[v_2(x)]\}^{-1}\right]\right] f_2(x)\,dx \\
&= \int \left[v_1(x)\left[v_2^{-1}(\overline{x}) - v_2^{-1}(x)\right] - v_2(x)\left[v_2^{-1}(\overline{x}) - 1 - \eta_2^* - \eta_4^*\right]\right] f_2(x)\,dx \\
&= \left\{v_2^{-1}(\overline{x}) - E\left[v_2^{-1}(x)\right]\right\} \int \left[\frac{v_2^{-1}(\overline{x}) - v_2^{-1}(x)}{v_2^{-1}(\overline{x}) - E\left[v_2^{-1}(x)\right]}\right] v_1(x) f_2(x)\,dx \\
&\quad - \int \left[\frac{v_2^{-1}(\overline{x}) - 1 - \eta_2 - \eta_4}{v_2^{-1}(\overline{x}) - E\left[v_2^{-1}(x)\right]}\right] v_2(x) f_2(x)\,dx \\
&= \left\{v_2^{-1}(\overline{x}) - E\left[v_2^{-1}(x)\right]\right\} \int \left[g_2^*(x) v_1(x) - (\alpha_2^*/\alpha_1^*)^{\frac{\lambda_t}{b_t}} v_2(x)\right] f_2(x)\,dx
\end{aligned}$$

where the last line uses the derivation of α_1^*, α_2^* and $g_2^*(x)$. Comparing this expression for $C(\rho)$ with the combination constraint, we see that $C(\rho^*) \geq 0$ if and only if the combination constraint it satisfied, and from the Kuhn Tucker theorem, if $C(\rho^*) > 0$, then $\eta_4^* = 0$.

The first equation in the proposition, the truth telling constraint, is satisfied with equality by the optimal contract $(w_1^*(x), w_2^*(x))$ at ρ^* the true value of ρ. The other condition follows directly from Proposition 4 and the first part of this proposition, which we just proved. ■

Proof of Proposition 6. By the definition of $\widehat{\rho}$ and $\widehat{\alpha}_2$, the participation constraint is satisfied. Next we show that $\widehat{\theta}$ satisfies the first order conditions, the incentive compatibility, the truth telling and the combination constraints, and that $(\widehat{\eta}_2, \widehat{\eta}_{31}, \widehat{\eta}_{32}, \widehat{\eta}_4)$ solves the equations for the Lagrange multipliers. For notational convenience, we define

$$\widehat{v}_s(x) \equiv \exp\left[-\widehat{\rho} w^*(x)\right]$$

From the first order condition for the second state

$$v_2(x)^{-1} = 1 + \eta_2 + \eta_{32}[(\alpha_2/\alpha_1)^{\frac{\lambda_t}{b_t}} - g_2(x)] + \eta_4$$

From the definitions of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\eta}_{32}$ and $\widehat{g}_2(x)$ we see that

$$\begin{aligned}
(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x) &= \left[\frac{\widehat{v}_2(\overline{x})^{-1} - 1 - \widehat{\eta}_2 - \widehat{\eta}_4}{\widehat{v}_2(\overline{x})^{-1} - E_2\left[\widehat{v}_2(x)^{-1}\right]}\right] - \left[\frac{\widehat{v}_2(\overline{x})^{-1} - \widehat{v}_2(x)^{-1}}{\widehat{v}_2(\overline{x})^{-1} - E_2\left[\widehat{v}_2(\overline{x})^{-1}\right]}\right] \\
&= \frac{\widehat{v}_2(x)^{-1} - 1 - \widehat{\eta}_2 - \widehat{\eta}_4}{\widehat{v}_2(\overline{x})^{-1} - E_2\left[\widehat{v}_2(x)^{-1}\right]} \\
&= \widehat{\eta}_{32}^{-1}\left[\widehat{v}_2(x)^{-1} - 1 - \widehat{\eta}_2 - \widehat{\eta}_4\right]
\end{aligned}$$

Rearranging to make $\widehat{v}_2(x)^{-1}$ the subject of the equation yields

$$\widehat{v}_2^{-1}(x) = 1 + \widehat{\eta}_2 + \widehat{\eta}_4 + \widehat{\eta}_{32}\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)\right]$$

which satisfies the first order condition if $\widehat{\eta}$ is the vector of Lagrange multipliers, yet to be shown.

Turning to the first state, the definitions of $\widehat{\eta}_{31}$ and $\widehat{g}_1(x)$ imply

$$\begin{aligned}\widehat{\eta}_{31}\widehat{g}_1(x) &= \widehat{v}_1(x)^{-1} - \widehat{v}_1(\overline{x})^{-1} + \widehat{\eta}_2\left[h(x) - h(\overline{x})\right] + \widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{\frac{\lambda_t}{b_t}}\widehat{g}_2(x)\,h(x)\\ &= \widehat{v}_1(x)^{-1} - \widehat{v}_1(\overline{x})^{-1} - \widehat{\eta}_2 h(\overline{x}) + \widehat{\eta}_2 h(x) + \widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{\frac{\lambda_t}{b_t}}\widehat{g}_2(x)\,h(x)\\ &= \widehat{v}_1(x)^{-1} - 1 - \widehat{\eta}_{31}(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} + \widehat{\eta}_2 h(x) + \widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{\frac{\lambda_t}{b_t}}\widehat{g}_2(x)\,h(x)\end{aligned}$$

where the last line follows from the definition of $\widehat{\eta}_{31}$. Rearranging to make $\widehat{v}_1(x)^{-1}$ the subject of the equation

$$\widehat{v}_1(x)^{-1} = 1 + \widehat{\eta}_{31}(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{\eta}_2 h(x) - \widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{\frac{\lambda_t}{b_t}}\widehat{g}_2(x)\,h(x) + \widehat{\eta}_{31}\widehat{g}_1(x)$$

Thus $w^*(x)$ satisfies the first order condition for $\widehat{\theta}$ if $\widehat{\eta}$ is the vector of Lagrange multipliers.

From the definition of $\widehat{\rho}$, the $w^*(x)$ contract satisfies the truth telling and the combination constraints. Similarly the construction of $\widehat{\alpha}_2$ ensures that $w^*(x)$ satisfies the participation constraint. Next we show that $w^*(x)$ satisfies the incentive compatibility constraints under $\widehat{\theta}$. In the second state, note that

$$\begin{aligned}\widehat{v}_2(x)\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)\right] &= \widehat{v}_2(x)\left\{\left[\frac{\widehat{v}_2(\overline{x})^{-1} - 1 - \widehat{\eta}_2 - \widehat{\eta}_4}{\widehat{v}_2(\overline{x})^{-1} - E_2\left[\widehat{v}_2(x)^{-1}\right]}\right] - \left[\frac{\widehat{v}_2^{-1}(\overline{x}) - \widehat{v}_2(x)^{-1}}{v_2^{-1}(\overline{x}) - E_2\left[\widehat{v}_2(x)^{-1}\right]}\right]\right\}\\ &= \widehat{\eta}_{32}^{-1}\left[1 - \widehat{v}_2(x) - \widehat{\eta}_2\widehat{v}_2(x) - \widehat{\eta}_4\widehat{v}_2(x)\right]\end{aligned}$$

Taking expectations of the bracketed term with respect to the second state then gives

$$1 - E_2\left[\widehat{v}_2(x)\right] - \widehat{\eta}_2 E_2\left[\widehat{v}_2(x)\right] - \widehat{\eta}_4 E_2\left[\widehat{v}_2(x)\right] = 0$$

the result following directly from the definition of $\widehat{\eta}_2$ and $\widehat{\eta}_4$. Therefore

$$E\left\{\widehat{v}_2(x)\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)\right]\right\} = 0$$

proving the incentive compatibility constraint holds with equality in the second state. In the first state

$$\widehat{v}_1(x)\,\widehat{\eta}_{31}\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_1(x)\right] = 1 - \widehat{v}_1(x) + \widehat{\eta}_2 h(x)\,\widehat{v}_1(x) + \widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{\frac{\lambda_t}{b_t}}\widehat{g}_2(x)\,h(x)\,\widehat{v}_1(x)$$

But the definitions. Taking expectations with respect to the first state we appeal to the definitions of $\widehat{\eta}_2$ and $\widehat{\eta}_4$ to obtain

$$1 - E_1\left[\widehat{v}_1(x)\right] + \widehat{\eta}_2 E_1\left[h(x)\,\widehat{v}_1(x)\right] + \widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{\frac{\lambda_t}{b_t}} E_1\left[\widehat{g}_2(x)\,h(x)\,\widehat{v}_1(x)\right] = 0$$

Therefore

$$E\left\{\widehat{v}_1(x)\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_1(x)\right]\right\} = 0$$

thus proving the incentive compatibility constraint also holds with equality in the first state.

Turning now to the four equations defining Lagrange multipliers, It follows from the definition of the mappings $\widehat{v}_s(x)$, $\widehat{g}_1(x)$ and $\widehat{g}_2(x)$ and the parameters $\widehat{\eta}_2, \widehat{\eta}_{31}$, $\widehat{\eta}_{31}$, and $\widehat{\eta}_4$ that

$$\widehat{v}_1(x)^{-1} = 1 - \widehat{\eta}_2 h(x) + \widehat{\eta}_{31} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{\eta}_{31}\widehat{g}_1(x) + \widehat{\eta}_4 (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} \widehat{g}_2(x) h(x)$$

$$\widehat{v}_2(x)^{-1} = 1 + \widehat{\eta}_2 + \widehat{\eta}_{32}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)] + \widehat{\eta}_4$$

Since the truth telling constraint is satisfied of the we see that

$$\begin{aligned}
& \int_{\underline{x}}^{\infty} \frac{f_2(x)\,dx}{1 - \widehat{\eta}_2 h(x) + \widehat{\eta}_{31} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{\eta}_{31}\widehat{g}_1(x) + \widehat{\eta}_4 (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} \widehat{g}_2(x) h(x)} \\
\equiv\ & E_2\left[\widehat{v}_1(x)^{-1}\right] \\
=\ & E_2\left[\widehat{v}_2(x)^{-1}\right] \\
\equiv\ & \int_{\underline{x}}^{\infty} \frac{f_2(x)\,dx}{1 + \widehat{\eta}_2 + \widehat{\eta}_{32}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)] + \widehat{\eta}_4}
\end{aligned}$$

Similarly the incentive compatibility conditions are satisfied with equality in each state by $\widehat{\theta}$ and $\widehat{\eta}$, which implies:

$$\begin{aligned}
0 &= E_1\left\{\widehat{v}_1(x)\left[\widehat{g}_1(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}\right]\right\} \\
&= \int_{\underline{x}}^{\infty} \frac{\widehat{g}_1(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}}{1 - \widehat{\eta}_2 h(x) + \widehat{\eta}_{31} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{\eta}_{31}\widehat{g}_1(x) + \widehat{\eta}_4 (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} \widehat{g}_2(x) h(x)} f_1(x)\,dx
\end{aligned}$$

and

$$\begin{aligned}
0 &= E_2\left\{\widehat{v}_2(x)\left[\widehat{g}_2(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}\right]\right\} \\
&= \int_{\underline{x}}^{\infty} \frac{\widehat{g}_2(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}}}{1 + \widehat{\eta}_2 + \widehat{\eta}_{32}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)] + \widehat{\eta}_4} f_2(x)\,dx
\end{aligned}$$

From its definition $\widehat{\eta}_4 = 0$ when $C(\widehat{\rho}) > 0$, and when $C(\widehat{\rho}) = 0$

$$\begin{aligned}
& \int_{\underline{x}}^{\infty} \frac{f_2(x)\,dx}{1 + \widehat{\eta}_2 + \widehat{\eta}_{32}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{g}_2(x)] + \widehat{\eta}_4} \\
=\ & (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} E_2\{\widehat{v}_2(x)\} \\
=\ & (\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} E_2\{\widehat{v}_2(x)\} \\
=\ & \int \frac{(\alpha_1/\alpha_2)^{\frac{\lambda_t}{b_t}} f_2(x)\, g_2(x)\,dx}{1 - \widehat{\eta}_2 h(x) + \widehat{\eta}_{31} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} - \widehat{\eta}_{31}\widehat{g}_1(x) + \widehat{\eta}_4 (\widehat{\alpha}_2/\widehat{\alpha}_1)^{\frac{\lambda_t}{b_t}} \widehat{g}_2(x) h(x)}
\end{aligned}$$

This demonstrates that given $\widehat{\theta}$ and $(w_1^*(x), w_2^*(x))$ the equations defining the Lagrange multipliers are solved by $\widehat{\eta}$. Therefore $\widehat{\eta}$ is a vector of Lagrange multipliers for the optimization problem under $\widehat{\theta}$. Appealing to the Kuhn Tucker theorem, we conclude that under $\widehat{\theta}$ the contract $(w_1^*(x), w_2^*(x))$ solves the generalized moral hazard model for $\widehat{\theta}$. Therefore $\widehat{\theta}$ is observationally equivalent to θ^* as claimed. ■

References

[1] **Abowd, John M. and Kaplan, David S.** "Executive Compensation: Six Questions That Need Answering" *The Journal of Economic Perspectives,* 1999 13 (4) pp.145-168.

[2] **Aggarwal, Rajesh K. and Samwick, Andrew A.** "The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation" *Journal of Political Economy, 1999,* 107 (1), pp.65-105.

[3] **Antle, Rick and Smith, Abbie.** "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives" *Journal of Accounting Research, 1986,* 24 pp.1-39.

[4] **Antle, Rick and Smith, Abbie.** "Measuring Executive Compensation: Methods and an Application" J*ournal of Accounting Research, 1985,* 23 pp.296-325..

[5] **Bertrand, Marianne and Mullainathan, Sendhil.** "Are CEOS Rewarded for Luck? The Ones Without Principals Are." *The Quarterly Journal of Economics, August 2001,* pp.901-932.

[6] **Bertrand, Marianne and Schoar, Antoinette.** "Managing with Style: The Effect of Managers on Firm Policies", *The Quarterly Journal of Economics,* November, 2003, CXVII (4) pp.1169-1208.

[7] **Bolton, Patrick and Dewatripont, Mathias**. *Contract Theory*, Cambridge, Massachusetts: The MIT Press, 2005.

[8] **Chiappori, Pierre-André and Salanié, Bernard.** "Testing Contract Theory: A Survey of Some Recent Work", *in Advances in Economics and Econometrics - Theory and Applications, Eighth World Congress*, M. Dewatripont, L. Hansen and P. Turnovsky, ed., Econometric Society Monographs, Cambridge University Press, Cambridge, 2003, pp.115-149.

[9] **Finnerty, J.E**. " Insiders and Market Efficiency," *Journal of Business*, 1976,Vol. 31, pp.1141-1148.

[10] **Garen, John E.** "Executive Compensation and Principal-Agent Theory" *Journal of Political Economy,* 1994 102 (6), pp.1175-1199.

[11] **Gayle, George-Levi and Miller, Robert A.** "Has Moral Hazard become a more important Factor in Managerial Compensation?", Tepper school of Business, Carnegie Mellon University, November 2005.

[12] **Grossman, Sanford and Hart, Oliver.** "An Analysis of the Principal-Agent Problem" *Econometrica, 1983,* 51 pp. 7-46.

[13] **Hall, Brian J. and Liebman, Jeffrey B.** "Are CEOS Really Paid Like Bureaucrats?" *The Quarterly Journal of Economics,* August 1998, CXIII pp. 653-680.

[14] **Huddart, Steven, John S. Hughes, and Carolyn B. Levine**, " Public Disclosure and Dissimulation of Insider Trades", *Econometrica*, Vol. 69, No. 3. pp. 665-681.

[15] **Hurwicz, Leonid,** " Generalization of the concept of Identification" *In Statistical inference in dynamic economic models.* Cowles Commission Monograph, 1950, 10, New York, John Wiley.

[16] **Koopmans, Tjalling C., and Olaf Reierol** : " The Identification of Structural Characteristics", *Annals of Mathematical Statistics*, 1950, 21, pp165-181.

[17] **Jaffe, J.F**. " Special information and Insider Trading", *Journal of Business*, 1974,Vol. 47, pp. 410-428

[18] **Laffont, Jean-Jacques. and Martimont, David.** *The Theory of Incentives: The Prinicipal-Agent Model.* Princeton University Press, Princeton, 2002

[19] **Lorie, J.H. and V. Niederhoffer "** Predictive and Statistical Properties of Insider Trading**,"** *Journal of Law and Economics*, 1968, vol. 11, pp.35-51.

[20] **Margiotta, Marry M. and Miller, Robert A.** "Managerial Compensation and The Cost of Moral Hazard" *International Economic Review, August 2000,* 41 (3) pp. 669-719.

[21] **Masson, Robert.** "Executive Motivations, Earnings, and Consequent Equity Performance" *Journal of Political Economy, 1971,* 79 pp. 1278-1292.

[22] **Murphy, Kevin J.** "Executive Compensation", *in Handbook of Labor Economics*, Volume 3, O. Ashenfelter and D. Card, ed., Elsevier Science B.V., 1999. Chapter 38, pp. 2485-2563.

[23] **Myerson, R.** "Optimal coordination mechanisms in generalized principalagent problem". *Journal of Mathematical Economics*, 10:67–81, 1982.

[24] **Newey, Whitney** "A Methods of Moments Interpretation of Sequential Estimators" *Economic Letters* 14 pp. 201-206 (1984)

[25] **Prendergast, Canice.** "The Provision of Incentives in Firms" *Journal of Economic Literature* XXXVII pp. 7-63 (1999).

[26] **Salanié, Bernard.** *The Economics of Contracts: A Primer*, MIT Press, Cambridge, 2005.

[27] **Seyhun, H. Neiat**. " Insiders' Profits, Cost of Trading, and Market Efficiency," J*ournal of Financial Economics*, 1986, vol.16, pp. 189-212.

[28] ————-. " The Effectiveness of Insider Trading Sanctions," J*ournal of Law and Economics*, 1992a, vol. 35, pp. 149-182.

_ _ _ _ _ _ _ " Why Does Aggregate Insider Trading Predict Future Stock Returns?" *The Quarterly Journal of Economics*, Vol. 107, pp. 1303-1331.

Table 1

CROSS-SECTIONAL INFORMATION ON SECTORS ALL CURRENCY IN MILLION OF $US (2000)

(STANDARD DEVIATIONS IN PARENTHESIS)

Variables	
Sales	4,168 (109,000)
Value of Equity	1,868 (4,648)
Total Assets	9,926 (40,300)
Number of Employees	18,341 (46,960)
Abnormal returns 1	11.484 (47.091)
Abnormal returns 2	8.63E-8 (45.903)
Number of Obsevations	282,768
Number of Firms	2,557

TABLE 2
CROSS-SECTION INFORMATION ON COMPONENT OF TOTAL COMPENSATION IN THOUSANDS OF $US (2000)
(STANDARD DEVIATIONS IN PARENTHESIS)

Components	All	CEO	Non-CEO
Total Compensation	2,319 (12,121)	5,320 (19,369)	1,562 (9,303)
Salary and Bonus	667 (905)	1,127 (1,282)	552 (738)
Value of Options Granted	903 (3,753)	1,782 (7,169)	681 (2,106)
Value of Restricted Stock Granted	152 (936)	298 (1,464)	115 (743)
Change in Wealth from Options Held	281 (8,710)	1,474 (13,567)	-18 (6,939)
Change in Wealth from Stock Held	125 (4,350)	264 (6,791)	90 (3,473)

TABLE 3
COEFFICIENTS FROM REGRESSION OF CHANGES IN MANAGERS STOCK HOLDINGS ON
(STANDARD ERRORS IN PARENTTHESIS)

VARIABLES	COEFFICIENT
salary/total compensation	-0.768 (2.13)
lead abnormal return	2.304 (1.108)