

Identifying and Testing Models of Managerial Compensation

George-Levi Gayle and Robert A. Miller*

Tepper School of Business, Carnegie Mellon University

June 8, 2009

Abstract

We develop a pure moral hazard model, and a closely related hybrid one, where there are both hidden actions and hidden information, to derive the restrictions from optimal contract theory that characterize set identification. In pure moral hazard models, the expected utility of managers is equalized across states, whereas in a hybrid model the optimal contract equates the expected utility of truth telling with the expected utility of lying. These restrictions are testable. Our identification analysis establishes sharp and tight bounds on the identified set. Our tests and estimators are based on these bounds. We apply semiparametric methods to test the models, estimate the structural parameters, and quantify the effects of hidden actions versus hidden information. The pure moral hazard model is rejected on a large panel data set measuring the compensation of chief executive officers and the financial and accounting returns of the publicly traded firms they manage. We do not, however, reject the restrictions of the hybrid model, and our structural estimates for that model show the degree of private information varies considerably across sectors and over firm size.

1 Introduction

Managerial compensation, and their year to year change in wealth, varies significantly with the excess return on their firm's stock.¹ The positive correlation is primarily driven by the

*We would like to thank Maher Said, Melissa Tartari and Steve Tadelis for their comments. We have also benefitted from presentations at the Econometric Society 2007 Summer Meetings, SITE Theory-Based Micro-Econometric Modelling Workshop 2008, Society of Economic Dynamics 2008, Toulouse Econometrics of Industrial Organization Workshop 2008, 2nd Annual CAPCP Conference at Pennsylvania State University, CRES Empirical Microeconomics Conference at Washington University and seminars at Carnegie Mellon University, Cornell University, Duke University, University of Essex, Georgetown University, London School of Economics, Northwestern University, University of Pennsylvania, University of Pittsburgh, Princeton University, Queens University and the University of Toronto. This research was supported by National Science Foundation Grant SES0721098.

¹See Antle and Smith (1985, 1986), Hall and Liebman (1998) and Gayle and Miller (2008a), who find that about half the total variation in compensation can be explained by a nonlinear regression on excess

fact that managers hold stocks and options in their own firms. A large body of nonstructural empirical work has investigated how well managerial compensation can be rationalized by moral hazard.² As the dominant paradigm for explaining executive compensation, the theory of moral hazard postulates that risk averse managers are paid compensation that fluctuates with signals shareholders observe about decisions their managers make, notably excess returns of the firm, in order to align the incentives of the managers when their nonpecuniary goals differ from maximizing shareholder wealth and the actions and decision of management are not monitored by shareholders. A growing number of papers on structural estimation have sort to quantify the economic significance of moral hazard in the labor market.³

Applying the theory of moral hazard to explain managerial compensation has been challenged on several fronts. First, managers are paid for luck, risk factors beyond executive control that increase the volatility of their income,⁴ which is inconsistent with the notion of mitigating uncertainty in compensation to risk averse agents. Second, several empirical studies find that trading by corporate insiders appears profitable,⁵ but in models of pure moral hazard, managers do not have private information about the firm's future prospects. Third, managerial compensation not only depends on the financial returns of the firm, but also its accounting returns.⁶ In models of pure moral hazard, shareholders might use signals other than financial returns to determine optimal compensation, but the reporting of accounting income is subject to considerable discretion by the manager.

The findings in the second paragraph cannot be easily explained by a pure moral hazard model where shareholders maximize the expected value of the firm. Yet compensation boards representing shareholders can and do exercise considerable discretion in setting executive pay. Thus the boards could ignore accounting income when compensating the manager, and they could retrospectively neutralize any returns managers receive from owning or trading financial securities related to the firm.⁷ Is it reasonable to assume that the only two constraints facing compensation boards arise from possibly losing the manager to another firm, or not being able to monitor the manager's actions?

We argue in this paper that compensation boards are also constrained because they have

returns, industry effects and bond prices.

²Murphy (1999), Prendergast (1999) and Chiappori and Salanie (2000) survey this literature.

³See Haubrich (1994), Ferrall and Shearer (1999), Margiotta and Miller (2000), Shearer (2004), Dubois and Vukina (2005), Dufflo, Hann and Ryan (2007), Gayle and Miller (2008a, 2008b) and Golan, Gayle and Miller (2008b).

⁴See Bertrand and Mullainathan (2001) and Bebchuk and Fried (2003).

⁵See Lorie and Niederhoffer (1968), Jaffe (1974), and Finnerty (1976) and Seyhun (1986), who finds that insiders tend to buy before an abnormal rise in stock prices and sell before an abnormal decline. Seyhun (1992a, 1992b) presents evidence showing that insiders earn over 5 percent abnormal returns on average, and determines that insider trades predict up to 60 percent of the total variation in one-year-ahead returns. Gayle and Miller (2008b) construct a simple self-financing dynamic portfolio strategy based on changes in asset holdings by managers that significantly outperforms the market portfolio, realizing over 90 percent of the gains that could have been achieved with perfect foresight.

⁶We document this stylized fact in Section 4 of this paper when presenting the data on managerial compensation and firm performance.

⁷Corporation executives are legally obligated to report their financial assets holdings in firms they manage to the Securities and Exchange Commission in a timely manner, and therefore cannot trade anonymously in their firm's shares.

less information about the state of the firm as the manager. We reconcile the correlations between managerial compensation and accounting and financial measures of firm returns with an optimal contracting model, where there is both hidden action and hidden information. We derive the restrictions from the model on data of periodic reports by the manager about the firm's future prospects, excess returns by firms and managerial compensation. We demonstrate which data generating processes can be rationalized by the models we consider, and in that way establish set identification. Taking our framework to a large panel data set on the compensation of chief executive officers and the firms they manage, we test whether hidden information is present or not, and assess its importance relative to pure moral hazard.

Section 2 presents a theoretical framework for exploring moral hazard and hidden information, characterizes the set of feasible contracts facing shareholders, and presents the solution to their cost minimization problem. In our model there are two ways of hiding the truth, by lying about the state but working in shareholder interests, or by lying about the state and working against shareholder interests. The first is prevented by contracts satisfying a truth telling constraint, the second by contracts satisfying a sincerity constraint. Our least cost approach to optimal contracting to moral hazard models with hidden information, or hybrid models, extends the two step procedure for solving pure moral hazard models pioneered by Grossman and Hart (1980).

We analyze identification in Section 3, establishing necessary and sufficient conditions for identifying and testing the theory with data on compensation, abnormal returns and states revealed by the optimal contract. From the equations and inequalities defining the optimal contract, we derive sharp and tight bounds on the parameter space of the theoretical structure, without imposing strong parametric assumptions on the conditional distributions of abnormal returns, or the functional form of the contract. The closest papers to our work on nonparametric identification are the independent analyses of Perrigne and Vuong (2007), who also exploit predictions from principal agent theory to analyze nonparametric identification in models of moral hazard and adverse selection, and Huang, Perrigne and Vuong (2007), who nonparametrically identify and estimate a nonlinear pricing model of advertising. Nevertheless our work differs from theirs in many respects. The differences range from the theoretical structures considered, to the results obtained, and the nature of empirical problems to which the identification analysis applies.⁸

Section 4 describe the sources of data on chief executive officers of publicly traded companies, explain how the data were compiled, and summarize its main features. The panel of roughly 27,000 observations covering the period 1993 to 2005 contains data on compensation to about 4,700 chief executive officers compiled from Standard and Poor's ExecuComp data base, financial and accounting information of the 2,600 publicly trade firms they manage, taken from the Center for Securities Research (CRSP), as well as background characteristics on the sector and size of the firms. Each point in our sample essentially consists of background information about the size and sector of the firm, the compensation of the CEO,

⁸For example their theoretical framework models adverse selection with continuous choices by the agent, whereas we analyze a model with hidden information and discrete choices. Their work considers point identification, ours set identification and testing. Their application investigates nonlinear pricing, as opposed to executive compensation.

the financial returns of the firm, and the firm's accounting returns, which we treat as an announcement by the CEO about the state of his firm.

We derive test statistics for both the pure and hybrid models of moral hazard in Section 5 and present our results for the CEO compensation data. We develop nonparametric methods and moment inequalities to test the model, and provide the first structural estimates quantifying the importance of hidden information relative to pure moral hazard in executive compensation packages. The methods we develop to test the specification draw from Chernozhukov, Hong and Tamer (2007) and Romano and Shaikh (2006). Our tests reject the class of pure moral hazard model we consider, because the expected utilities managers receive from announcing different states cannot be equated. In other words, our data also show that managers are paid for luck, contradicting a strong prediction of the pure moral hazard model. However we cannot reject the hybrid model, because the moment restrictions implied by the truth telling and sincerity constraints, not relevant for the pure moral hazard model, are satisfied by the data.

In Section 6 we develop a semi-parametric estimator for structural parameters of the hybrid model by exploiting the moment inequalities implied by the restrictions derived from our analysis of identification.⁹ We report our estimates, including measures of the costs of hidden information and moral hazard in managerial compensation issues. The risk aversion parameter is plausible and precisely estimated. The cost of ridding the firm of a moral hazard problem is of the same order as managerial compensation but the cost to the firm of not incentivizing the manager is greater by orders of magnitude. Finally the importance of hidden information varies significantly by firm and industry characteristic, and within our framework this finding implies that the value of accounting information depends on firm type.

2 The Models

We analyze models of pure and hybrid moral hazard. In the hybrid model the manager of a firm is subject to moral hazard, but also has private information about his firm's future returns. Shareholders do not observe the state of the firm or manager's activities within the period. Contracts between shareholders and the manager must satisfy three conditions, a participation constraint, that assures the manager he will have higher expected utility from employment with the firm rather than another one, an incentive compatibility constraint, that induces him to maximize the value of the firm rather than using the resources of the firm to pursue some other objective, and two other conditions that induce the manager to truthfully reveal his private information. At the beginning of each period the board of directors proposes a compensation plan to the manager, which depends on the realization of the firms abnormal returns as well as accounting information to be provided by the manager. Based on the board's proposal the manager decides whether to remain with the firm or leave

⁹For applications of moment inequalities in the parametric estimation of structural models, see Andrews, Berry and Jia (2004), Bajari, Bankard and Levin (2007), Ciliberto and Tamer (2007), Ho (2008), Ho, Ho and Mortimer (2008), Levine (2007), and Pakes, Porter, Ho, and Ishii (2006).

and picks real consumption expenditure for the period. Having accepted the contract offer, the manager observes the firm's state, provides some accounting information, and chooses a work routine that is not observed by the directors. The return on the firm's equity in the current period is realized at the beginning of the next period. It partly depends on how well the firm was managed and the private information available to the manager. The manager is paid at the beginning of the next period. The objective of the manager is to sequentially maximize his expected lifetime utility, and the goal of the firm is expected value maximization.

2.1 Assumptions and Notation

At the beginning of period t the manager is paid compensation denoted by w_t for his work the previous period, denominated in terms of period t consumption units. He makes his consumption choice, a positive real number denoted by c_t , and the board proposes a new contract. The board announces how managerial compensation will be determined as a function of what he will disclose about the firm's prospects, denoted by $r_t \in \{1, 2\}$, and its subsequent performance, measured by abnormal returns x_{t+1} revealed at the beginning of the next period. We denote this mapping by $w_{rt}(x)$, the subscript t designating that the optimal compensation schedule may depend on current economic conditions, such as a bond prices. Then the manager chooses whether to be engaged by the firm or be engaged outside the firm, either with another firm or in retirement. Denote this decision by the indicator $l_{t0} \in \{0, 1\}$, where $l_{t0} = 1$ if the manager chooses to be engaged outside the firm and $l_{t0} = 0$ if he chooses to be engaged inside the firm.

In the hybrid model if the manager accepts employment with the firm, so $l_{t0} = 0$, the prospects of the firm are now fully revealed to the manager but partially hidden to the shareholders. There are two states, and the probability the first state occurs is identically and independently distributed with probability $\varphi_1 \in (0, 1)$. For convenience we denote the probability of the second state occurring by $\varphi_2 \equiv 1 - \varphi_1$. We assume that managers privately observe the true state $s_t \in \{1, 2\}$ in period t , information that affects the distribution of the firm's abnormal returns next period, and reports the state $r_t \in \{1, 2\}$ to the board.

If the manager discloses the second state, meaning $r_t = 2$, then the board can independently confirm or refute it. (For example imagine shareholders can review geological surveys of new oil fields, but that managers exercise some discretion about when to disclose them.) This prevents the manager from lying when the first state occurs, and models the idea 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.¹⁰ If $s_t = 2$ the manager then truthfully declares or lies about the firm's prospects by announcing

¹⁰Thus the managers at Enron, for example, were prosecuted as criminals, not penalized internally by shareholders, for overstating the firm's prospects. Dye and Sridhar (2005) make a similar assumption about information disclosure in their theoretical analysis of the severity of moral hazard. The agent can fail to disclose a new project without fear of recrimination, but cannot lie about a nonexistent project. Gayle, Li and Miller (2009) analyze the following natural extension of this two state disclosure framework: All but one state, called normal, are considered extraordinary, and the principal cannot distinguish any extraordinary state from the normal state unless the agent discloses it.

$r_t \in \{1, 2\}$, effectively selecting one of two schedules, $w_{1t}(x)$ or $w_{2t}(x)$ in that case, but if $s_t = 1$ he reports $r_t = 1$. By way of contrast, in the pure moral hazard model we analyze, the true state $s_t \in \{1, 2\}$ is observed by shareholders in period t , eliminating any role for managerial reporting. This is the only difference between the pure and hybrid frameworks.¹¹

The manager then makes his unobserved labor effort choice, denoted by $l_{stj} \in \{0, 1\}$ for $j \in \{1, 2\}$ for period t which may depend on his private information about the state. There are two possibilities, to diligently pursue the shareholders objectives of value maximization by setting $l_{st2} = 1$, or to accept employment with the firm but follow the objectives he would pursue if he was paid a fixed by setting $l_{st1} = 1$, and called shirking. Let $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$. Since leaving the firm, working diligently and shirking are mutually exclusive activities, $l_{t0} + l_{st1} + l_{st2} = 1$.¹²

At the beginning of period $t + 1$ abnormal returns for the firm, x_{t+1} , are drawn from a probability distribution which depends on the true state s_t in period t and the manager's action then, l_{st} . We denote the probability density function for abnormal returns when the manager works diligently and the state is s by $f_s(x)$. Similarly, let $f_s(x)g_s(x)$ denote the probability density function for abnormal returns in period t when the manager shirks. Thus for both states $s_t \in \{1, 2\}$:

$$\int x f_s(x) g_s(x) dx \equiv E_s [x g_s(x)] < E_s [x] \equiv \int x f_s(x) dx$$

the inequality reflecting the preference of shareholders for diligent work over shirking. Since $f_s(x)g_s(x)$ is a density, $g_s(x)$ must be a positive mapping; integrating $f_s(x)g_s(x)$ with respect to x demonstrates $E_s [g_s(x)] = 1$, where the expectation is taken with respect to $f_s(x)$. We assume there is an upper range of returns that, conditional on the state s , might be achieved with diligence, but is extremely unlikely to occur if the manager shirks. Formally $\lim_{x \rightarrow \infty} [g_s(x)] = 0$ for each $s \in \{1, 2\}$. Intuitively this assumption states that a truly extraordinary performance by the firm can only be attained if the manager works hard. We show below that under optimal contracting this assumption guarantees compensation is bounded at the limit of $x \rightarrow \infty$. We assume that $g_s(x)$ is bounded, an assumption that rules out the possibility of setting a contract that is arbitrarily close to the first best resource allocation, first noted by Mirrlees (1975), by severely punishing the manager when $g_s(x)$ takes an extremely high value.

¹¹Our model should be distinguished from the mixed models reviewed for example in Chapter 7 of Laffont and Martimort (2002) and Chapter 6 of Bolton and Dewatripont (2005), where there is adverse selection of agents. In our framework this would occur if managers had different abilities or types, unobserved to shareholders, and competed with each other for a position with the firm after their type is revealed.

¹²Data with variation in the number of hours executives work is much harder to obtain, and more problematic to interpret, than comparable data from longitudinal surveys of the general workforce (such as the PSID and the NLSY). However data on executive position within the firm, and job turnover at the upper levels, is readily accessible, and serves as a useful job descriptor. See Baker, Gibbs and Holmstrom (1994a, 1994b) and Gayle, Golan and Miller (2008a). Seeking to explain the executive career profile in their empirical analysis of firm specific and general managerial human capital, Gayle, Golan and Miller (2008b) extend our moral hazard framework, by increasing the number of choices, considering external employment opportunities to managers and different ranks within each firm.

We make similar assumptions about the function $h(x) \equiv \varphi_2 f_2(x) / \varphi_1 f_1(x)$, the weighted likelihood ratio of the second state occurring relative to the first given any observed value of excess returns $x \in R$. If $h(x)$ is unbounded for some value of x , then truth telling about the second state can be enforced without cost. To rule out this possibility we assume $h(x)$ is bounded throughout. This assumption implies shareholders cannot be sure that the second state occurred simply by observing excess returns. To capture the idea that the manager can, without fear of legal recrimination, withhold information that might increase the value of the firm, we merely assume that the second state is most likely to have occurred relative to the first state at the limit $x \rightarrow \infty$, which is much weaker than assuming $f_2(x)$ first order stochastically dominates $f_1(x)$. Summarizing our assumptions about $h(x)$ mathematically:

$$\lim_{x \rightarrow \infty} [h(x)] = \sup_{x \in R} [h(x)] \equiv \bar{h} < \infty \quad (1)$$

The manager's wealth is endogenously determined by his consumption and compensation. We assume there are a complete set of markets for all publicly disclosed events, effectively attributes all deviations from the law of one price to the particular market imperfections under consideration. Let b_t denote the price of a bond that pays of a unit of consumption each period from period t onwards, relative to the price of a unit of consumption in period t ; to simplify the exposition we assume b_{t+1} is known at period t .¹³ 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:

$$- \sum_{t=0}^{\infty} \sum_{j=0}^J \beta^t \alpha_j l_{tj} \exp(-\gamma c_t) \quad (2)$$

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

2.2 Feasible Short Term Contracts

At the end of the section we prove that the optimal long term contract can be implemented by a sequence of short term contracts, which explains why our discussion focuses on the optimal one period contract. First we derive the indirect utility function for a manager who, upon reaching period t , works at most one period before retiring, as a function of $w_{rt}(x)$, the compensation contract he anticipates receiving from the firm, l_{st} , his labor supply choices, and r_t , his report about the firm's prospects which may depend on the true state of the firm s_t , which he observes after making his employment decision but before

¹³Our empirical implementation is based on a short panel. Assuming b_{t+1} is known one period in advance frees us from estimating the conditional probability distribution of interest rates next period. Our procedures and results are easy to generalize to economies where b_{t+1} is stochastic if its distribution is known or can be estimated.

his effort decision. Then we define the set of feasible contracts to those that respect five constraints we define. A participation constraint ensures that the manager is indifferent between working one period and then leaving, versus not working for the firm at all. Two incentive compatibility constraints restrict short term contracts to those payment schedules in which the manager prefers to work diligently rather than shirk in each state. A truth telling condition requires shareholders to write contracts that induce the manager to select a compensation schedule that reveals the firm's prospects in the second state. Finally the contract guards against the possibility of the manager lying about the second state and also shirking, by satisfying a sincerity constraint.¹⁴

The cornerstone of the constraint formulation that circumscribe the minimization problem shareholders solve is the indirect utility function for a manager choosing between immediate retirement versus retirement one period hence. Lemma 2.1 states this indirect utility function in terms of the utility he would receive from returning immediately. To state the lemma, let $r_t(s)$ denote the manager's disclosure rule about the state when the true state is $s \in \{1, 2\}$.

Lemma 1 *If the manager, offered a contract of $w_{rt}(x)$ for announcing r , retires in period t or period $t + 1$ by setting $(1 - l_{t0})(1 - l_{t+1,0}) = 0$, upon observing the state s and reporting $r_t(s)$ he optimally chooses $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$ to minimize:*

$$\sum_{s=1}^2 \varphi_s \left\{ \left(\frac{\alpha_0}{\alpha_1} l_{st1} + \frac{\alpha_0}{\alpha_2} l_{st2} \right)^{1/(b_t-1)} + E_s \left[\exp \left(-\frac{\gamma w_{r_t(s),t}(x)}{b_{t+1}} \right) [g_s(x) l_{t1} + l_{t2}] \right] \right\} \quad (3)$$

Had he truthfully disclosed the true state s_t in period t , the manager would actually receive $w_{st}(x)$ as compensation if abnormal returns x are realized at the end of the next period $t + 1$. Suppressing for expositional convenience the bond price b_{t+1} , and recalling our assumption that b_{t+1} is known at period t , we let $v_{st}(x)$ measure how (the negative of) utility is scaled up by $w_{st}(x)$:

$$v_{st}(x) \equiv \exp \left(-\frac{\gamma w_{st}(x)}{b_{t+1}} \right) \quad (4)$$

To induce an honest, diligent manager to participate, his expected utility from employment must exceed the utility he would obtain from retirement. Setting $(l_{t2}, r_t) = (1, s_t)$ in (3) and substituting in $v_{st}(x)$, the participation constraint is thus:

$$\left[\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s v_{st}(x) f_s(x) dx \right] \equiv E[v_{st}(x)] \leq (\alpha_0/\alpha_2)^{1/(b_t-1)} \quad (5)$$

Given his decision to stay with the firm one more period, and truthfully reveal the state, the incentive compatibility constraint induces the manager to prefer working diligently to

¹⁴This framework does not quite fit into the generalized principal agent problem for which Myerson (1982) and Green and Laffont (1986) formulated the revelation principle, because in our model shareholders can independently verify whether the second state has occurred if the manager discloses it. In Gayle and Miller (2009) we show that the revelation principle also extends to the hybrid model. Hence there is no loss of generality from restricting our theoretical and empirical analysis to direct mechanisms.

shirking. Substituting the definition of $v_{st}(x)$ into (3) and comparing the expected utility obtained from setting $l_{t1} = 1$ with the expected utility obtained from setting $l_{t2} = 1$ for any given state, we obtain the incentive compatibility constraint for diligence as:

$$0 \leq \int_{\underline{x}}^{\infty} \left(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)} \right) v_{st}(x) f_s(x) dx \equiv E_s \left[\left(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)} \right) v_{st}(x) \right] \quad (6)$$

for $s \in \{1, 2\}$. In the pure moral hazard model, shareholders minimize expected compensation subject to the incentive compatibility and participation constraints.

In the hybrid model information hidden from shareholders further restricts the set of contracts that can be implemented. Comparing the expected value from lying about the second state and working diligently with the expected utility from reporting honestly in the second state and working diligently, we obtain the truth telling constraint:

$$0 \leq \int [v_{1t}(x) - v_{2t}(x)] f_2(x) dx \equiv E_2 [v_{1t}(x) - v_{2t}(x)] \quad (7)$$

An optimal contract also induces the manager not to understate and shirk in the second state, behavior we describe as sincere. Comparing the manager's expected utility from lying and shirking with the utility from reporting honestly and working diligently, the sincerity condition reduces to:

$$0 \leq \int \left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t}(x) g_2(x) - v_{2t}(x) \right] f_2(x) dx \equiv E_2 \left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t} g_2(x) - v_{2t}(x) \right] \quad (8)$$

where $(\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t}(x)$ 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.

2.3 Cost Minimization

Having defined the constraints, we formulate the cost minimization problem shareholders solve to induce diligence and truth telling in both frameworks, starting with the hybrid case. Then we discuss cost minimizing when diligence is called for in at most one state.

2.3.1 Diligence and Truth Telling in Both States

Recall from its definition that $\log [v_{st}(x)]$ is proportional to $-w_{st}(x)$. Therefore deriving $w_{st}(x)$ to minimize expected compensation of inducing diligent work in both states subject to the five constraints is equivalent to choosing $v_{st}(x)$ to maximize:

$$\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s \log [v_{st}(x)] f_s(x) dx \equiv E [\log v_{st}(x)] \quad (9)$$

subject to the same five constraints. To achieve diligent work and truth telling, shareholders maximize:

$$\begin{aligned}
& \sum_{s=1}^2 \varphi_\rho \int_{\underline{x}}^{\infty} \left\{ \log [v_{st}(x)] + \eta_{0t} \left[(\alpha_0/\alpha_2)^{1/(b_t-1)} - v_{st}(x) \right] \right\} f_s(x) dx \quad (10) \\
& + \sum_{s=1}^2 \varphi_s \eta_{st} \int_{\underline{x}}^{\infty} v_{st}(x) \left[(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}) \right] f_s(x) dx \\
& + \varphi_2 \eta_{3t} \int_{\underline{x}}^{\infty} [v_{1t}(x) - v_{2t}(x)] f_2(x) dx \\
& + \varphi_2 \eta_{4t} \int_{\underline{x}}^{\infty} \left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t}(x) g_2(x) - v_{2t}(x) \right] f_2(x) dx
\end{aligned}$$

with respect to $v_{st}(x)$, where η_{0t} through η_{4t} are the shadow values assigned to the linear constraints. Since each constraint is a convex set, their intersection is too. Also $\log v$ is concave increasing in v , the expectations operator preserves concavity, so the objective function is concave in $v_{st}(x)$ for each x . Hence the Kuhn Tucker theorem guarantees there is a unique positive solution to the equation system formed from the first order conditions augmented by the complementary slackness conditions.

The first order conditions for this problem are:

$$\begin{aligned}
v_{1t}(x)^{-1} &= \eta_{0t} + \eta_{1t} \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x) \right] - \eta_{3t} h(x) \quad (11) \\
&\quad - \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x) h(x) \\
v_{2t}(x)^{-1} &= \eta_{0t} + \eta_{2t} \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x) \right] + \eta_{3t} + \eta_{4t}
\end{aligned}$$

The following lemma is helpful for interpreting the first order conditions.

Lemma 2 *The Lagrange multipliers satisfy:*

1. $\eta_{0t} = (\alpha_2/\alpha_0)^{1/(b_t-1)}$
2. $\eta_{3t} + \eta_{4t} = E_2 [v_{2t}(x)]^{-1} - E [v_{st}(x)]^{-1}$

By inspection, the first equality in Lemma (2) demonstrates that (α_2/α_0) only enters through the definition of η_{0t} , the shadow value of participation, and conversely the value of η_{0t} does not depend on which other constraints are binding. It follows that if a monitoring technology was introduced to align managerial and shareholder interests by fiat, the manager could be directed to work diligently, the reason for seeking information from the manager about the state would disappear, the incentive compatibility, truth telling and sincerity constraints would be relaxed. In that case the manager would receive a fixed wage of:

$$w_t^{(2)} \equiv b_{t+1} \log (\alpha_2/\alpha_0) / \gamma (b_t - 1) \quad (12)$$

in both states to compensate him for the nonpecuniary attributes of the job relative to the value of alternative employment.

2.3.2 A Specialization

The cost minimizing solution in the hybrid case is found by substituting the first order conditions into the constraints to solve the remaining four Lagrange multipliers, successively imposing different combinations of constraints to check which, if any, are satisfied by strict inequalities, rather than equalities. Simple yet general conditions on the primitives that determine which combination of Lagrange multipliers hold with equality do not exist. However there are sufficient conditions on the primitives for the sincerity constraint not to bind. Suppose the firm's losses from shirking do not depend on the state, meaning that opportunities afforded by the second state can only be realized if the manager is diligent. By Lemma (3) below this ensures the sincerity constraint does not bind and $\eta_{4t} = 0$. In this case we substitute the first order condition into the incentive compatibility and truth telling constraints yielding the following three equations in the remaining three unknowns η_{1t} , η_{2t} , and η_{3t} . They are:

$$\begin{aligned} & \int_{\underline{x}}^{\infty} \frac{1}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_{3t}h(x) + \eta_{1t}[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x)]} f_2(x) dx \\ = & \int_{\underline{x}}^{\infty} \frac{1}{(\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_{3t} + \eta_{2t}[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)]} f_2(x) dx \end{aligned} \quad (13)$$

$$\begin{aligned} 0 &= \int_{\underline{x}}^{\infty} \frac{g_1(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_{3t}h(x) + \eta_{1t}(\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_{1t}g_1(x)} f_1(x) dx \\ &= \int_{\underline{x}}^{\infty} \frac{g_2(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_{3t} + \eta_{2t}(\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_{2t}g_2(x)} f_2(x) dx \end{aligned} \quad (14)$$

Lemma 3 *If $f_1(x)g_1(x) \equiv f_2(x)g_2(x)$, then $\eta_{4t} = 0$.*

2.3.3 Pure Moral Hazard

To solve the pure moral hazard problem for diligent work, shareholders maximize (9) subject to the participation constraint (5) and incentive compatibility constraints (6). It is immaterial to both managers and shareholders whether the state is revealed before or after the contract is made in models of pure moral hazard. If the contract is made before the state is revealed, there is only one participation constraint, and the shareholders' problem reduces to maximizing (10) and the associated first order conditions (11), but setting $\eta_{3t} = \eta_{4t} = 0$ to nullify the truth telling and sincerity constraints which are not relevant in the pure moral hazard case. Similarly two of the Kuhn Tucker equations drop out, leaving the solution to the remaining two Lagrange multipliers η_{st} for $s \in \{1, 2\}$ uniquely defined by the solution in η to:

$$\int_{\underline{x}}^{\infty} \frac{[(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)})] f_s(x)}{(\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_s(x)]} dx = 0 \quad (15)$$

Substituting the solution, denoted by $\eta_{st}^{(M)}$, into Equation (15) into the first order condition state s and making compensation the subject of the resulting equation then yields the optimal compensation schedule $w_{st}^{(M)}(x)$ for the pure moral hazard model when diligence in both states is demanded by shareholders, defined:¹⁵

$$w_{st}^{(M)}(x) = w_t^{(2)} + \frac{b_{t+1}}{\gamma} \log \left[1 + \eta_{st}^{(M)} (\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_{st}^{(M)} g_s(x) \right] \quad (16)$$

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 separately maximize:

$$\int_{\underline{x}}^{\infty} \left\{ \log v_{st}(x) + \eta_{0st} \left[(\alpha_0/\alpha_2)^{1/(b_t-1)} - v_{st}(x) \right] + \eta_{st} v_{st}(x) \left[(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}) \right] \right\} f_s(x) dx \quad (17)$$

for $s \in \{1, 2\}$. In this case, however, the first order conditions simplify to the other case where there is only one participation constraint, because following the same logic as the proof to Lemma (2) it is straightforward to show that $\eta_{01t} = \eta_{02t}$.

From the second equality in Lemma (2) we infer that if $\eta_{3t} = \eta_{4t} = 0$, then:

$$E_2[v_{2t}(x)] = E[v_{st}(x)] = E_1[v_{1t}(x)]$$

In words, if neither the truth telling nor the sincerity constraints bind, or if the state is directly observed by shareholders, then the pure moral hazard case applies, and expected utility is equalized across states.

Otherwise $(\eta_{3t} + \eta_{4t})$ is strictly positive implying expected utility from the pure moral hazard case straddles the expected utility attained in the two states of the hybrid model:

$$E_2[v_{2t}(x)] < E[v_{st}(x)] < E_1[v_{1t}(x)]$$

When the manager has private information he is rewarded for the firm's good prospects and penalized for the firm's bad prospects; in other words, the optimal contract pays him for luck.

2.3.4 Shirking

Shareholders might not want the manager to work diligently in both states, given the extra compensation required to motivate him. To cover the other three possibilities, let $L_{st} \in \{0, 1\}$ indicate a directive to work diligently in period t when the reported state is $s \in \{1, 2\}$, where $L_{st} = 0$ means the shareholders anticipate the manager to shirk in state s . Irrespective of whether there is hidden information or not, the wage for employing a manager to shirk in both states is independent of the firm's abnormal return and the state; it just offsets the value of leaving the firm. The cost minimizing way of achieving $(L_{1t}, L_{2t}) = (0, 0)$ to the fixed wage:

$$w_t^{(1)} \equiv b_{t+1} \log(\alpha_1/\alpha_0) / \gamma (b_t - 1) \quad (18)$$

¹⁵See Margiotta and Miller (2000) for a derivation of this specialization.

In pure moral hazard models, the optimal compensation in each state does not depend on the effort level induced in the other state because the expected utility in each is equalized to the expected value from leaving the firm. Hence $w_t^{(1)}$ is paid in the first state and $w_{2t}^{(M)}(x)$ is paid in the second state to minimize, subject to truth telling, the expected costs of achieving $(L_{1t}, L_{2t}) = (0, 1)$, while $w_{1t}^{(M)}(x)$ is paid in the first state and $w_t^{(1)}$ is paid in the second state if $(L_{1t}, L_{2t}) = (1, 0)$ is selected by shareholders.

This only leaves the two cases $(L_{1t}, L_{2t}) = (1, 0)$ and $(L_{1t}, L_{2t}) = (0, 1)$ to solve for the hybrid model. We denote by $w_{st}^{(L_1, L_2)}(x)$ the cost minimizing way of achieving the various (L_{1t}, L_{2t}) combinations of effort in the hybrid model subject to truth telling, where $w_t^{(1)} \equiv w_{st}^{(0,0)}(x)$ and $w_{st}(x) \equiv w_{st}^{(1,1)}(x)$ for $s \in \{1, 2\}$. After solving for $\eta_{0t}^{(1,0)}$, the Lagrange multiplier on the participation constraint in the cost minimization problem associated with $(L_{1t}, L_{2t}) = (1, 0)$, using a similar notation for the other Lagrange multipliers, and defining:

$$v_{st}^{(L_1, L_2)}(x) \equiv \exp \left[-\gamma w_{st}^{(L_1, L_2)}(x) / b_{t+1} \right]$$

the first order conditions for inducing effort in the first state only reduce to:

$$\begin{aligned} v_{1t}^{(1,0)}(x)^{-1} &= \sum_{s=1}^2 (1 - \varphi_s) (\alpha_s / \alpha_0)^{1/(b_t-1)} + \eta_{1t}^{(1,0)} [(\alpha_2 / \alpha_1)^{1/(b_t-1)} - g_1(x)] \quad (19) \\ &\quad - \eta_{3t}^{(1,0)} h(x) - \eta_{4t}^{(1,0)} (\alpha_1 / \alpha_2)^{1/(b_t-1)} g_2(x) h(x) \\ v_{2t}^{(1,0)}(x)^{-1} &= \sum_{s=1}^2 (1 - \varphi_s) (\alpha_s / \alpha_0)^{1/(b_t-1)} + \eta_{3t}^{(1,0)} + \eta_{4t}^{(1,0)} \end{aligned}$$

Managers are paid a constant wage to shirk in the second state, but more than they would be in a pure moral hazard model, so that they will reveal the state, as indicated by the fact that $(\eta_{3t}^{(1,0)} + \eta_{4t}^{(1,0)})$ is positive in the hybrid model but zero when there is no private information. Substituting Equations (19) into the incentive compatibility, truth telling and sincerity constraints, we solve the three equations in three unknowns $\eta_{1t}^{(1,0)}$, $\eta_{3t}^{(1,0)}$ and $\eta_{4t}^{(1,0)}$ for the case $(L_1, L_2) = (1, 0)$. The Kuhn Tucker theorem ensures the positive solution to the three multipliers is unique. Substituting the solution back into Equation (19), optimal compensation $w_{st}^{(1,0)}(x)$ is obtained for the $(L_1, L_2) = (1, 0)$ case as a function of x .

The remaining case, $(L_1, L_2) = (0, 1)$ in the hybrid model, inducing effort in the second state only, is solved in a similar fashion. The first order conditions reduce to:

$$\begin{aligned} v_{1t}^{(0,1)}(x)^{-1} &= \sum_{s=1}^2 \varphi_s (\alpha_s / \alpha_0)^{1/(b_t-1)} - \eta_{3t}^{(0,1)} h(x) - \eta_{4t}^{(0,1)} (\alpha_1 / \alpha_2)^{1/(b_t-1)} g_2(x) h(x) \quad (20) \\ v_{2t}^{(0,1)}(x)^{-1} &= \sum_{s=1}^2 \varphi_s (\alpha_s / \alpha_0)^{1/(b_t-1)} + \eta_{2t}^{(0,1)} [(\alpha_2 / \alpha_1)^{1/(b_t-1)} - g_2(x)] + \eta_{3t}^{(0,1)} + \eta_{4t}^{(0,1)} \end{aligned}$$

Though shirking is a default action requiring no performance incentives, here the optimal compensation for shirking in the first state nevertheless depends on the firm's returns, in order to induce the manager at the beginning of the period to truthfully reveal the second state, when he would work diligently should it occur. Following an analogous approach to the $(L_1, L_2) = (1, 0)$ case, we obtain $w_{st}^{(0,1)}(x)$ for $s \in \{1, 2\}$.

2.4 The Optimal Contract

The equations above demonstrate that the cost minimizing schedule selected for each state depends on the effort level the shareholders wish to induce in both states. Anticipating his reaction to any proposed compensation package, shareholders demand the pair of effort choices from the manager that maximize the expected value of the firm, by selecting one of the four cost minimizing contracts. Upon observing the state $s \in \{1, 2\}$ at the beginning of period t shareholders in the pure moral hazard model choose (L_{1t}, L_{2t}) to maximize:

$$\int_{\underline{x}}^{\infty} \left\{ L_{st} \left[V_t x - w_{st}^{(M)}(x) \right] + (1 - L_{st}) \left[V_t x g_s(x) - w_t^{(1)} \right] \right\} f_s(x) dx \quad (21)$$

where V_t is the equity value of the firm at the beginning of the period. In the hybrid model their optimization problem is not separable across the unobserved states. Then shareholders maximize:

$$\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s \left\{ L_{st} V_t x + (1 - L_{st}) V_t x g_s(x) - w_{st}^{(L_{1t}, L_{2t})}(x) \right\} f_s(x) dx \quad (22)$$

in the hybrid model. Denoting by (L_{1t}^o, L_{2t}^o) the maximizing choice of (L_{1t}, L_{2t}) in Expressions (21) and (22) completes our characterization of the solution to the short term contracting problem.

In this framework there are no gains from a long term arrangement between shareholders and the manager. Lemma 4 verifies the assumptions of Fudenberg, Holmstrom and Milgrom (1990) are met, thus establishing that the long term optimal contract decentralizes to a sequence of short term contracts satisfying the first order conditions described above.

Lemma 4 *Denote by $\bar{\tau}$ the manager's date of retirement. The optimal long term contract can be implemented by a $\bar{\tau}$ period replication of the optimal short term contract.*

The optimality of short term contracts follows from four main assumptions.¹⁶ First, the firm is assumed to have no better access to financial markets than its manager, so there is no reason for the manager to use the firm as a bank to help smooth his consumption stream. In our model there are complete markets in financial assets, and the manager cannot trade anonymously in the financial assets of her own firm. In practice chief executive officers are quite wealthy, can select from a rich array of financial instruments for managing their personal assets, and are legally obligated to report trading in the securities of their firms. Second, we assume there is absolute risk aversion and additive separability in consumption in work and effort choices. Since the manager's degree of risk aversion is not affected by his wealth, tracking consumption and wealth is not necessary to form the optimal compensation contract, an assumption of convenience because consumption data is not observed. Although

¹⁶Malcomson and Spinnewyn (1988), Fudenberg, Holmstrom and Milgrom (1990) and Rey and Salanie (1990) have independently established conditions under which long term optimal contracts can be implemented via a sequence of one period contracts in dynamic models of generalized moral hazard, and the proof of Lemma 4 in the Appendix draws extensively upon their results.

there is little reason to believe managers exhibit absolute risk aversion, short term contracting would remain optimal if the other three assumptions hold, and if in addition the executive's consumption is observed by shareholders. In practice networking between executives and their boards would provide a source of information about the managerial consumption and how their leisure time is spent. Third, the signal the manager receives about the state is independently distributed over time. In a competitive market shareholders efficiently price the firm when they are truthfully updated by what the manager knows. Consequently changes brought about by new states are fully unanticipated, so as a first approximation the value of the firm follows a random walk. For this reason we view the signal independence assumption as innocuous.

Arguably the most problematic assumption is the fourth, additive separability in production with respect to effort and information. Supposing hidden choices the manager selects have unforeseen long term repercussions, or the manager holds private information about the firm or his own aptitude for the job that carries over several periods, under such circumstances shareholders would engage their manager in a long term contract. As an empirical matter, chief executive officers of major firms are well known in the business community, at the peak of their lifecycle earnings, and have an average tenure of less than five years. In their structural empirical analysis of promotion and executive turnover, Gayle, Golan and Miller (2008b) find that human capital accumulation is not an important motivator for these individuals as for lower ranked managers. We conclude that while dynamic contracting is a promising topic to explore in the context of managerial compensation, particularly amongst the lower ranks, our model of short term optimal contracting for chief executive officers is a useful benchmark.

3 Identification

The parameters of the model are characterized by $f_s(x)$ and $g_s(x)$ for each state $s \in \{1, 2\}$, which together define the probability density functions of abnormal returns in the states, the probability of each state occurring φ_s , 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 γ . For known values determining managerial preferences and the state probability, namely $(\gamma, \alpha_0, \alpha_1, \alpha_2)$ and φ_s , plus specific functions defining firm technology, $f_s(x)$ and $g_s(x)$, our analysis in Section 2 provides the basis for solving the model and predicting what effort level shareholders demand, whether the truth telling and sincerity constraints bind or not, and the shape of the optimal compensation schedule.

Alternatively suppose we have data on rational shareholders and managers who have solved the model for a particular specification, but we do not know which one. Given a parameterization, the equations and inequalities derived in Section 2 impose restrictions on to the joint probability distribution of the reported states, the observed returns and managerial compensation. If, for example, we infer from such data that expected utility is equalized across states, but that the constraints are violated, then we conclude that shareholders directly observe the state. If the expected utility is equalized across states, but the constraints are satisfied, then we conclude that whether shareholders directly observe the state or not is

immaterial to the optimal contract. If the data show that expected utility is not equalized across states, but the truth telling and sincerity constraints are satisfied, and at least one of them is binding, then the hybrid model applies.

The union of all possible restrictions generated on the joint probability distribution by the particular parametrization identifies the set of observationally equivalent parametrizations. If a joint probability distribution of the announced states, the observed returns and managerial compensation does not satisfy all the restrictions for any parameterization of the model, then the distribution cannot be generated by a member of the class of models we are considering. In this section we derive the identified set.¹⁷

3.1 Overview

We consider observations for (r_t, x_t) drawn from a (repeated) cross section, or a panel of length T , along with a compensation schedule $w_{rt}(x)$ defined on the space of states and abnormal returns, where the reported state $r_t \in \{1, 2\}$ is an independently distributed random variable with probability (φ_1, φ_2) , and x_t is a random variable with a known conditional density for the reported state. The true state s_t is hidden from shareholders (and econometricians). Only the reported state r_t is observed by shareholders. Under the null hypothesis that the data is generated by our model, managers truthfully reveal the state. Thus in this section on identification, we directly apply the implication of the theory, that $s_t = r_t(s_t)$, by setting $(r_t, x_t) = (s_t, x_t)$. We impose a weak regularity condition on the schedule $w_{rt}(x)$ that for all t and r_t , compensation is bounded, and the bound is approached at extraordinarily high returns:

$$\lim_{x \rightarrow \infty} [w_{rt}(x)] = \sup_{x \in R} [w_{rt}(x)] \equiv \bar{w}_{rt} < \infty \quad (23)$$

For expositional purposes only, this section assumes that the probability φ_1 , the densities $f_1(x)$ and $f_2(x)$, and the compensation schedule $w_{st}(x)$ are known, and that $\alpha_0 \equiv 1$. This implies that φ_1 and $h(x)$ (including \bar{h}) are known too. Setting $\alpha_0 \equiv 1$ simply normalizes the utility level from leaving the firm, meaning that α_j values the nonpecuniary features of engaging in activity $j \in \{1, 2\}$ within the firm relative to the total utility value 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, which is why we now focus on identifying the two mappings $g_s(x)$ plus the constants α_1, α_2 and γ .

To facilitate the discussion we partition the parameter space into γ , the manager's coefficient of absolute risk aversion, and $\theta \equiv (\alpha_1, \alpha_2, g_1(x), g_2(x))$, which characterizes the nonpecuniary benefits to the manager and the costs of shirking to the firm, and denote by (γ^*, θ^*) a generic model within the theoretical framework of Section 2, where $\theta^* \equiv (\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$. Following Chesher (2007) we say the structural parameter

¹⁷In Gayle and Miller (2009) we extend this analysis by investigating identification in a generalized moral hazard model, where the number of states, actions and the information partition of the principal, is finite but arbitrary.

$(\gamma, \theta) \in R^+ \times \Theta$ with true value (γ^*, θ^*) is identified if (γ^*, θ^*) can be written as a functional of the joint distribution of the observed variables, defined from the marginal p and the conditionals $f_s(x)$. If a correspondence, rather than a functional, from the conditional distribution of the observed variables to the parameter space, defines an equivalence class, then structural parameter is set identified. Finally if (r_t, x_t) cannot be rationalized by any $(\gamma, \theta) \in R^+ \times \Theta$ the theoretical framework is rejected.

There are four permutations of the pure and hybrid models to consider, depending the unobserved value of (L_{1t}^o, L_{2t}^o) . After dismissing three permutations, because they are not supported by our data for reasons that are readily apparent, we thoroughly analyze, for the pure and hybrid models, the permutation $(L_{1t}^o, L_{2t}^o) = (1, 1)$, where diligent effort is called forth in both states. We prove three main results on identification. First we prove that if (r_t, x_t) was generated by some (γ, θ) , and γ^* is known, then θ^* is identified. This first result greatly simplifies the analysis of observational equivalence and specification testing, by permitting us to frame the remaining discussion in terms of values the risk aversion parameter γ might take on R^+ , rather than values (γ, θ) might take in the much larger space $R^+ \times \Theta$. We define two sets of risk aversion parameters, Γ_1 for the pure moral hazard case, and Γ_2 for the hybrid case, sets formed from the participation, truth telling and sincerity constraints, by exploiting equalities and inequalities that hold in theory. Second, given a data generating process denoted by $\{r_t, x_t\}$ satisfying the regularity conditions defined above, we prove that if Γ_1 is empty then the data are inconsistent with our pure moral hazard framework, and similarly if Γ_2 is empty then the data are inconsistent with our hybrid framework. This second result provides the basis for specification testing we conduct in Section 4. Third, we prove both sets are sharp and tight. In other words, for $i \in \{1, 2\}$, only the elements belonging to Γ_i are observationally equivalent.

3.2 Some Shirking

We can briefly dispense with the three permutations formed from $(L_{1t}^o, L_{2t}^o) \neq (1, 1)$. If $(L_{1t}^o, L_{2t}^o) = (0, 0)$ or $(L_{1t}^o, L_{2t}^o) = (1, 0)$ then our analysis of the previous section implies that the manager is compensated independently of abnormal returns in the second state. Similarly if $(L_{1t}^o, L_{2t}^o) = (0, 1)$ in the pure moral hazard model, then compensation in the first state does not depend on excess returns. From data on abnormal returns and compensation it is easy to test, state by state, for any given value of exogenous factors, whether the latter depends on the former. In each state taken separately our data strongly rejects the null hypothesis of no effect.

If $(L_{1t}^o, L_{2t}^o) = (0, 1)$ then the first order conditions for the hybrid model imply that compensation depends on excess returns, as the data show. The hybrid model with $(L_{1t}^o, L_{2t}^o) = (0, 1)$ is more parsimonious than the $(L_{1t}^o, L_{2t}^o) = (1, 1)$, because the probability distribution of abnormal returns for diligent work in the second state does not play any role in estimation, yet compensation varies with abnormal returns in both states, as our data indicate. Upon algebraically manipulating the first order conditions to substitute out $g_2(x)$ we obtain:

$$v_{1t}(x)^{-1} = \Upsilon_{0t} + \Upsilon_{1t}h(x)v_{2t}(x)^{-1} - \Upsilon_{2t}h(x) \quad (24)$$

for the $(L_{1t}^o, L_{2t}^o) = (0, 1)$ permutation, where:

$$\begin{aligned}
\Upsilon_{0t} &\equiv \sum_{s=1}^2 \varphi_s (\alpha_s)^{1/(b_t-1)} \\
\Upsilon_{1t} &\equiv \eta_{2t}^{-1} \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} \\
\Upsilon_{2t} &\equiv \eta_{3t} + \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} \left[\eta_{2t}^{-1} \sum_{s=1}^2 \varphi_s (\alpha_s)^{1/(b_t-1)} + (\alpha_2/\alpha_1)^{1/(b_t-1)} \right. \\
&\quad \left. + \eta_{2t}^{-1} \eta_{3t} + \eta_{2t}^{-1} \eta_{4t} \right]
\end{aligned} \tag{25}$$

As we mentioned above, the bond rate b_t is observed, the state probabilities (φ_1, φ_2) are identified from the series on r_t , and $h(x)$ is identified from data on (r_t, x_t) . Hence the vector function $\Upsilon_t \equiv (\Upsilon_{0t}, \Upsilon_{1t}, \Upsilon_{2t})$ is essentially a mapping from the five dimensional parameter vector $(\eta_{2t}, \eta_{3t}, \eta_{4t}, \alpha_1, \alpha_2, \gamma)$, while γ is the only unknown parameter determining $v_{1t}(x)^{-1}$ and $v_{2t}(x)^{-1}$. Estimation entails fitting these six parameters to a linear form that embodies approximately as many restrictions as there are data points, Equation (24) holding exactly for all x . Consequently the $(L_{1t}^o, L_{2t}^o) = (0, 1)$ permutation produces a stochastic singularity in the data. Therefore the only case of interest is the most complicated, $(L_{1t}^o, L_{2t}^o) = (1, 1)$.

3.3 Assuming Risk Preferences are Known

We now prove that if γ^* is known, then $(\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$ is identified from (x_t, r_t, w_t) . This is accomplished by defining a vector function $\theta_t(\gamma)$ and then showing that the true parameter θ^* can be written as $(\gamma^*, \theta_t(\gamma^*))$ for all t .

Proposition 5 *For each (s, t) , let:*

$$\begin{aligned}
v_{st}(x, \gamma) &\equiv \exp[-\gamma w_{st}(x)/b_{t+1}] \\
\bar{v}_{st}(\gamma) &\equiv \exp[-\gamma \bar{w}_{st}/b_{t+1}].
\end{aligned}$$

Then $\theta^ = \theta_t(\gamma^*)$ for all t when $\theta_t(\gamma) \equiv (\alpha_{1t}(\gamma), \alpha_{2t}(\gamma), g_{1t}(x, \gamma), g_{2t}(x, \gamma))$ is recursively defined by the mappings:*

$$\begin{aligned}
\alpha_{2t}(\gamma) &\equiv E[v_{st}(x, \gamma)]^{1-b_t} \\
\eta_{2t}(\gamma) &\equiv \bar{v}_{2t}(\gamma)^{-1} - E_2[v_{2t}(x, \gamma)^{-1}] \\
g_{2t}(x, \gamma) &\equiv \eta_{2t}(\gamma)^{-1} [\bar{v}_{2t}(\gamma)^{-1} - v_{2t}(x, \gamma)^{-1}] \\
\alpha_{1t}(\gamma) &\equiv \alpha_{2t}(\gamma) \left\{ \frac{[\bar{v}_{2t}(\gamma)]^{-1} - E_2[v_{2t}(x, \gamma)]^{-1}}{[\bar{v}_{2t}(\gamma)]^{-1} - E_2[v_{2t}(x, \gamma)^{-1}]} \right\}^{1-b_t} \\
\eta_{4t}(\gamma) &\equiv \frac{E[v_{st}(x, \gamma)]^{-1} E_1[v_{1t}(x, \gamma)] - 1}{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x, \gamma) g_{2t}(x, \gamma) h(x)] - E_1[v_{1t}(x, \gamma) h(x)]} \\
&\quad - \frac{E_1[v_{1t}(x, \gamma) h(x)] \{E_2[v_{2t}(x, \gamma)]^{-1} - E[v_{st}(x, \gamma)]^{-1}\}}{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x, \gamma) g_{2t}(x, \gamma) h(x)] - E_1[v_{1t}(x, \gamma) h(x)]}
\end{aligned}$$

$$\begin{aligned}
\eta_{3t}(\gamma) &\equiv E_2[v_{2t}(x, \gamma)]^{-1} - \eta_{4t}(\gamma) - E[v_{st}(x, \gamma)]^{-1} \\
\eta_{1t}(\gamma) &\equiv [\alpha_{1t}(\gamma)/\alpha_{2t}(\gamma)]^{1/(b_t-1)} \{ \bar{v}_{1t}(\gamma)^{-1} - E[v_{st}(x, \gamma)]^{-1} + \eta_{3t}(\gamma) \bar{h} \} \\
g_{1t}(x, \gamma) &\equiv \eta_{1t}(\gamma)^{-1} \{ \bar{v}_{1t}(\gamma)^{-1} - v_{1t}(x, \gamma)^{-1} + \eta_{3t}(\gamma) [\bar{h} - h(x)] \\
&\quad - \eta_{4t}(\gamma) [\alpha_{1t}(\gamma)/\alpha_{2t}(\gamma)]^{1/(b_t-1)} g_{2t}(x, \gamma) h(x) \}
\end{aligned}$$

A corollary of this proposition is that when $\eta_{3t}(\gamma) = \eta_{4t}(\gamma) = 0$ for all γ and t , the definitions of $\eta_{1t}(\gamma)$ and $g_{1t}(x, \gamma)$ simplify to their second state counterparts. Thus in models of pure moral hazard:

$$\eta_{st}(\gamma) \equiv \bar{v}_{st}(\gamma)^{-1} - E_s[v_{st}(x, \gamma)^{-1}] \quad (26)$$

$$g_{st}(x, \gamma) \equiv \eta_{st}(\gamma)^{-1} [\bar{v}_{st}(\gamma)^{-1} - v_{st}(x, \gamma)^{-1}] \quad (27)$$

for $s \in \{1, 2\}$.

Since our test statistics and estimators are based on sample analogues to the population moments that define these parameters as a function of γ , we briefly describe the intuition to interpret the formulas. The certainty equivalent of a lottery that yields $w_{st}(x)$ for a person with absolute risk aversion parameter γ who then optimally spends it over his lifetime is $E[v_{st}(x, \gamma)]^{b_t-1}$. In the optimal contract for our model the manager is indifferent between accepting the job which provides nonpecuniary benefits of $\alpha_{2t}(\gamma)$ plus that certainty equivalent versus an option with benefits normalized to zero. This explains the formula for $\alpha_{2t}(\gamma)$, the preference parameter for working diligently. The ratio of $\alpha_{1t}(\gamma)$, the shirking parameter, to $\alpha_{2t}(\gamma)$, depend on the difference between $E[v_{2t}(x, \gamma)]^{-1}$ and $E[v_{2t}(x, \gamma)^{-1}]$. By definition $b_t > 1$ and the inverse function $1/v$ is convex decreasing in v , so Jensen's inequality implies $\alpha_{1t}(\gamma) < \alpha_{2t}(\gamma)$. In the special case where a fixed wage is paid $\alpha_{1t}(\gamma) \equiv \alpha_{2t}(\gamma)$, and the more dispersion there is in the random variable $v_{2t} \equiv v_{2t}(x, \gamma^*)$ induced by abnormal returns x and the compensation schedule $w_{2t}(x)$, the more pronounced the inequality. To interpret $g_{1t}(x, \gamma)$ and $g_{2t}(x, \gamma)$, the representations of the shirking to diligence likelihood ratios in the two states, we remark that $\bar{v}_{2t}(\gamma)^{-1} = \exp[\gamma \bar{w}_{st}/b_{t+1}]$ is the exponentially scaled value of the maximal compensation in the second state, which occurs as $x \rightarrow \infty$ and $g_{2t}(x, \gamma) \rightarrow 0$. At values of x where $g_{2t}(x, \gamma) > 0$, the first order condition reveals that compensation is less; thus the compensation gradient from the optimal contract traces out the likelihood ratio in the second state for any given value of γ , up to the shadow value or opportunity costs of marginally relaxing the incentive compatibility constraint, given by $\eta_{2t}(\gamma)$. The likelihood ratio in the first state, $g_{1t}(x, \gamma)$, has a similar representation, modified to reflect optimal adjustments in compensation to ensure truth telling and sincerity occur in the second state, by making payments in the first state sufficiently unattractive.

Proposition 5 does not impose the additional restrictions on the primitives that we mentioned in the previous section which render the sincerity constraint redundant and make the theoretical framework more elegant. Recall $f_1(x)g_1(x) = f_2(x)g_2(x)$ is one of several conditions in Lemma 3 that collectively ensure $\eta_{4t} = 0$. If the extra restriction is imposed on the data then we obtain from formulas for $g_1(x)$ and $g_2(x)$ given in Proposition 5:

$$\frac{f_1(x)}{f_2(x)} \equiv \frac{g_{2t}(x, \gamma)}{g_{1t}(x, \gamma)} = \frac{\eta_{1t}(\gamma)}{\eta_{2t}(\gamma)} \left[\frac{\bar{v}_{2t}(\gamma)^{-1} - v_{2t}(x, \gamma)^{-1}}{\bar{v}_{1t}(\gamma)^{-1} - v_{1t}(x, \gamma)^{-1} + \eta_{3t}(\gamma) h(\bar{x}) - \eta_{3t}(\gamma) h(x)} \right] \quad (28)$$

Noting $f_s(x)$ and $h(x)$ are identified from data on abnormal returns and their states, along with the quantities \bar{w}_{st} and $h(\bar{x})$, this permutation places very strong restrictions on admissible values of γ , because Equation (28) holds for every x in the space of abnormal returns. Even though the data generating process never draws from shirking distribution for our permutation of interest $(L_{1t}, L_{2t}) = (1, 1)$, imposing the regularity condition, that the the state does not affect the shirking distribution, has powerful implications from the perspective of inference that invariably would be rejected.

3.4 Admissible Risk Aversion

Proposition 5 implies the set of identified parameters can be indexed by a Borel set of values for the risk aversion parameter $\gamma \in R^+$ that are observationally equivalent. Next, we analyze the restrictions derived from optimizing behavior in the model that limit the admissible values of γ . Within the class of pure moral models we denote the set of admissible values by Γ_1 . For the class of hybrid models the corresponding set is denoted by Γ_2 . One source for culling values of γ that do not belong to Γ_i are restrictions that arise from the optimality conditions determining the compensation schedule for diligent work. A second source are the differences in firm value from shareholders that induce diligent effort in both states rather than encouraging shirking in at least one state. It is convenient to discuss the pure and hybrid models separately.

Define $\Psi_{1t}(\gamma)$ and $\Psi_{2t}(\gamma)$ by the formula:

$$\Psi_{st}(\gamma) \equiv E_1[v_{11}(x, \gamma)]^{1-b_1} - E_s[v_{st}(x, \gamma)]^{1-b_t} \quad (29)$$

for $s \in \{1, 2\}$. The optimal contract of Section 2 implies that $\Psi_{1t}(\gamma) = \Psi_{2t}(\gamma) = 0$ hold for all t , a condition we call competitive selection. Simply stated this condition ensures the expected utility of the manager is equalized across states in the pure moral hazard model. There is no payment for the luck of being in a favorable state at the beginning of the period.

By symmetry across states in the pure moral hazard model, the identification of the shirking parameter, $\alpha_{1t}(\gamma)$, as a function of γ in Proposition 5 can be established using the first instead of the second state. Making $[\alpha_{1t}(\gamma)/\alpha_{2t}(\gamma)]^{1/(1-b_t)}$ the subject of the equation defining $\alpha_{1t}(\gamma)$ in Proposition 5 we define the expression:

$$\Psi_{3t}(\gamma) \equiv \frac{\bar{v}_{1t}(\gamma)^{-1} - E_1[v_{1t}(x, \gamma)]^{-1}}{\bar{v}_{1t}(\gamma)^{-1} - E_1[v_{1t}(x, \gamma)^{-1}]} - \frac{\bar{v}_{2t}(\gamma)^{-1} - E_2[v_{2t}(x, \gamma)]^{-1}}{\bar{v}_{2t}(\gamma)^{-1} - E_2[v_{2t}(x, \gamma)^{-1}]} \quad (30)$$

and prove by symmetry that $\Psi_{3t}(\gamma^*) = 0$.

Next, consider the ramifications for identification when shareholders induce diligent effort rather than shirking in at least one of the states. By definition the expected value of abnormal returns, conditional on the state s , is zero in the pure moral hazard model, implying $V_{st}E_s[x] = E_s[w_{st}(x)]$. Defining the expression $\Lambda_{st}(\gamma)$ as

$$\Lambda_{st}(\gamma) \equiv E_s[V_{st}xg_{st}(x, \gamma)] - b_{t+1} \log[\alpha_{1t}(\gamma)]/\gamma(b_t - 1) \quad (31)$$

and appealing to Equation (18) and the discussion below it, we infer $\Lambda_{st}(\gamma^*) \geq 0$ since it is less profitable to induce shirking rather than diligence.

To summarize, competitive selection applies to each state taken separately, the taste parameter does not vary across states, and it is suboptimal for shareholders to induce shirking in either state. This yields two inequalities and three equalities in γ . Define the set of $\gamma \in R^+$ as:

$$\bar{\Gamma}_1 \equiv \{\gamma > 0 : \Lambda_{jt}(\gamma) \geq 0 \text{ for } j \in \{1, 2\} \text{ and } \Psi_{kt}(\gamma) = 0 \text{ for } k \in \{1, 2, 3\} \text{ and all } t\}$$

Our analysis implies $\gamma^* \in \bar{\Gamma}_1$ in pure moral hazard models.

The hybrid model also yields a restriction from the same value of $\alpha_{1t}(\gamma)$ appearing in the incentive compatibility conditions for both states. In the hybrid model, this restriction is conveniently stated as $\Psi_{4t}(\gamma^*) = 0$ where:

$$\begin{aligned} \Psi_{4t}(\gamma) \equiv & \eta_{1t}(\gamma) - [\bar{v}_{1t}(\gamma)]^{-1} - \eta_{3t}(\gamma) \bar{h} + E_1 [v_{1t}(x, \gamma)^{-1}] \\ & + \eta_{3t}(\gamma) \varphi_2 / \varphi_1 + \eta_{4t}(\gamma) (\alpha_{1t}(\gamma) / \alpha_{2t}(\gamma))^{1/(b_t-1)} E_1 [g_{2t}(x, \gamma) h(x)] \end{aligned} \quad (32)$$

Appealing to Lemma (2), the competitive selection equations in pure moral hazard models do not hold in hybrid models because $(\eta_{3t} + \eta_{4t})$ is strictly positive, inducing an increase (decrease) between the expected utility received in the second (first) state relative to the expected utility when the employment decision is made. Also since $(\eta_{3t} + \eta_{4t})$ is strictly positive, the Kuhn Tucker conditions require at least one of the truth telling and sincerity constraints in the hybrid moral hazard framework to hold with equality. Define $\Psi_{5t}(\gamma)$ and $\Psi_{6t}(\gamma)$ as :

$$\Psi_{5t}(\gamma) \equiv E_2 [v_{2t}(x, \gamma) - v_{1t}(x, \gamma)] \quad (33)$$

$$\Psi_{6t}(\gamma) \equiv E_2 \left[v_{1t}(x, \gamma) \frac{[\bar{v}_{2t}(\gamma)]^{-1} - [v_{2t}(x, \gamma)]^{-1}}{[\bar{v}_{2t}(\gamma)]^{-1} - E_2 [v_{2t}(x, \gamma)]^{-1}} - v_{2t}(x, \gamma) \right] \quad (34)$$

By Proposition 5, the truth telling constraint 7 implies $\Psi_{5t}(\gamma^*) \geq 0$, while the sincerity constraint 8 implies $\Psi_{6t}(\gamma^*) \geq 0$. In the hybrid model at least one of the constraints must hold with strict equality, further implying $\Psi_{5t}(\gamma^*) \Psi_{6t}(\gamma^*) = 0$. We also require the multipliers associated with truth telling and sincerity, respectively denoted by $\Psi_{7t}(\gamma) \equiv \eta_{3t}(\gamma)$ and $\Psi_{8t}(\gamma) \equiv \eta_{4t}(\gamma)$, to be positive at γ^* . Also both sets of complementary slackness conditions must be satisfied, meaning $\Psi_{5t}(\gamma^*) \Psi_{7t}(\gamma^*) = 0$ and $\Psi_{6t}(\gamma^*) \Psi_{8t}(\gamma^*) = 0$. Finally since $g_1(x)$ is a likelihood ratio in the hybrid model we ensure $g_{1t}(x, \gamma^*) \geq 0$ with unit mass by defining $\Psi_{9t}(\gamma) \equiv E_1 [1 \{g_{1t}(x, \gamma)\} - 1]$ and imposing the restriction that $\Psi_{9t}(\gamma^*) \geq 0$.

Turning now to the effort level induced by shareholders in the hybrid model, we first remark that if shirking is demanded in both states, that is $(L_1, L_2) = (0, 0)$, then compensation is determined as in Equation (18). Since this is suboptimal:

$$\Lambda_{3t}(\gamma) \equiv b_{t+1} \log [\alpha_{1t}(\gamma)] / \gamma (b_t - 1) - E [V_{st} x g_{st}(x, \gamma)] \quad (35)$$

is positive at γ^* . Since $(L_{1t}^o, L_{2t}^o) = (1, 1)$ the expected value to the firm would have been lower if either $(L_1, L_2) = (1, 0)$ or $(L_1, L_2) = (0, 1)$ had been chosen, and this observation yields two extra restrictions on γ^* to be utilized in identification. Given the parameterization

$(\gamma, \theta_t(\gamma))$ for any $\gamma \in R^+$, we denote by $w_{st}^{(L_1, L_2)}(x, \gamma)$ the cost minimizing contracts to induce (L_1, L_2) . Thus $w_{st}(x) \equiv w_{st}^{(1,1)}(x, \gamma^*)$ and from the notation defined in the previous section, $w_{st}^{(L_1, L_2)}(x) \equiv w_{st}^{(L_1, L_2)}(x, \gamma^*)$. Given the parameterization $(\gamma, \theta_t(\gamma))$, the difference in value to shareholders from selecting $(L_1, L_2) = (1, 1)$ versus $(L_1, L_2) = (1, 0)$ is

$$\Lambda_{4t}(\gamma) = -E_2 \left[V_{2t} x g_{2t}(x, \gamma) - w_{2t}^{(1,0)}(x, \gamma) \right] - E_1 \left[V_{1t} x - w_{1t}^{(1,0)}(x, \gamma) \right]$$

At $\gamma = \gamma^*$ this expression is positive since $(L_{1t}^o, L_{2t}^o) = (1, 1)$. In similar fashion we define

$$\Lambda_{5t}(\gamma) = -E_1 \left[V_{1t} x g_{1t}(x, \gamma) - w_{2t}^{(0,1)}(x, \gamma) \right] - E_2 \left[V_{2t} x - w_{1t}^{(0,1)}(x, \gamma) \right] \quad (36)$$

and note that $\Lambda_{5t}(\gamma^*) \geq 0$ because setting $(L_1, L_2) = (1, 1)$ is more profitable than setting $(L_1, L_2) = (0, 1)$.

Summarizing the restrictions directly applied to the hybrid model, the shape of the compensation schedule implies γ^* must satisfy the three inequalities relating to effort selection, the truth telling and sincerity equality which distinguish the hybrid model from pure moral hazard, their two associated complementary slackness conditions which constitute two more equalities, two inequalities defining the multipliers for sincerity and truth telling, an equality relating the same nonpecuniary benefits to both incentive compatibility conditions, and a restriction that the likelihood ratio $g_{1t}(x, \gamma)$ be nonnegative for all x . Defining the set of risk aversion parameters as:

$$\begin{aligned} \bar{\Gamma}_2 \equiv & \{ \gamma > 0 : \Lambda_{kt}(\gamma) \geq 0 \text{ for } k \in \{3, 4, 5\} \text{ and } \Psi_{jt}(\gamma) \geq 0 \text{ for } j \in \{5, 6, 7, 8, 9\} \\ & \text{and } \Psi_{4t}(\gamma) = \Psi_{5t}(\gamma) \Psi_{6t}(\gamma) = \Psi_{5t}(\gamma) \Psi_{7t}(\gamma) = \Psi_{6t}(\gamma) \Psi_{8t}(\gamma) = 0 \text{ and all } t \} \end{aligned}$$

our discussion implies $\gamma^* \in \bar{\Gamma}_2$ in hybrid moral hazard models.

Proposition 6 $\Gamma_i \subseteq \bar{\Gamma}_i$ for $i \in \{1, 2\}$.

Are there any other restrictions on the sets of admissible parameters to further shrink $\bar{\Gamma}_i$ for $i \in \{1, 2\}$ when $\bar{\Gamma}_i$ is not a singleton? The first two propositions only exploit the first order conditions and the Kuhn Tucker complementary slackness conditions of the optimization problem; however the second order conditions are satisfied for all $\gamma > 0$. The theory requires the preference parameters satisfy the inequalities $0 < \alpha_1 < \alpha_2$. Risk aversion also implies the expected compensation exceeds its certainty equivalent for all t , and in the pure moral hazard model for each state $s \in \{1, 2\}$ considered individually. Regarding the distribution of excess returns under shirking, we have only imposed $g_1(x) \geq 0$ in the hybrid model class. Yet the theory requires, for each $s \in \{1, 2\}$ in both Γ_1 and Γ_2 , that $g_s(x) \geq 0$ and also $E[g_s(x)] = 1$, because $g_s(x)$ is a likelihood ratio whose denominator corresponds to the probability density of the expectations operator. Finally all the Kuhn Tucker multipliers are nonnegative, not just those we referred to above. The next proposition shows that none of these restrictions have additional empirical content beyond those already impounded in $\bar{\Gamma}_i$.

Proposition 7 *Supposing $\gamma \in \bar{\Gamma}_i$ for $i = 1$ or $i = 2$, then for all $t \in \{1, 2, \dots\}$ and $s \in \{1, 2\}$*

1. $0 < \alpha_{1t}(\gamma) < \alpha_{2t}(\gamma)$
2. $E_s[w_{st}(x)] > w^{(2)}$ for $\gamma \in \bar{\Gamma}_1$ and $\sum_{s=1}^2 \varphi_s E_s[w_{st}(x)] > w^{(2)}$ for $\gamma \in \bar{\Gamma}_2$
3. $E_s[g_{st}(x, \gamma)] = 1$ and $g_{st}(x, \gamma) \rightarrow 0$ as $x \rightarrow \infty$
4. $g_{st}(x, \gamma) \geq 0$
5. $\eta_{st}(\gamma) \geq 0$

3.5 Observational Equivalence

It remains to show that if $\gamma \in \bar{\Gamma}_i$ for $i \in \{1, 2\}$, then γ is observationally equivalent to γ^* . We entertain four possibilities and prove a more general result for the class of data generating processes defined at the beginning of this section. Letting ϕ denote the empty set:

1. $\Gamma_2 = \phi$ but $\Gamma_1 \neq \phi$. The regular data generating process could arise from a model of pure moral hazard but not from a model with hidden information.
2. $\Gamma_1 = \phi$ but $\Gamma_2 \neq \phi$. The process could arise from a model of moral hazard with hidden information but not from a model of pure moral hazard.
3. $\Gamma_1 \neq \phi$ and $\Gamma_2 \neq \phi$. The process could arise from a model of pure moral hazard or a model with hidden information as well.
4. $\Gamma_1 \cup \Gamma_2 = \phi$. The process is inconsistent with a generalized model of moral hazard.

Our fourth and last proposition in this section establishes the bounds we have constructed are tight. It states that, if the following two conditions hold, any regular compensation schedule is observationally equivalent to a pure or hybrid moral hazard model analyzed in Section 2. First, there exists a positive real number γ satisfying the restrictions summarized by $\Gamma_i = \bar{\Gamma}_i$ for $i = 1$ (in the pure moral hazard case) or $i = 2$ (the hybrid case). Second, $\theta_1(\hat{\gamma}) = \theta_t(\hat{\gamma})$ for all $t \in \{1, \dots, T\}$, which means the other structural parameters implied by $\hat{\gamma}$ do not depend on time. (Violating the second condition would change the structure of the analysis in the previous section.)

Proposition 8 *Let $w_{rt}(x)$ denote any compensation schedule satisfying the regularity conditions (23). Define $\theta_t(\gamma)$ using the formulas in Proposition 5 with respect to the data generation process (r_t, x_t) . Suppose there exists $\hat{\gamma} \in R^+$ satisfying $\theta_1(\hat{\gamma}) = \theta_t(\hat{\gamma}) \in \Theta$ for all $t \in \{1, \dots, T\}$, and denote $\theta_t(\hat{\gamma})$ by $\hat{\theta}$. If $\hat{\gamma} \in \bar{\Gamma}_1$ then (r_t, x_t) is observationally equivalent to a data process generated by a pure moral hazard model, and if $\hat{\gamma} \in \bar{\Gamma}_2$ then (r_t, x_t) is observationally equivalent to a data process generated by a hybrid moral hazard model, both parameterized by $(\hat{\gamma}, \hat{\theta})$.*

4 Empirical Application

The data for our empirical study is based on a panel of N firms, each firm sampled up to T periods; the panel is unbalanced because not all firms are observed for all T periods. Differences between firms observed by the shareholders induce heterogeneity in managerial tasks and hence executive compensation plans.

To capture these differences we denote the characteristics of firm n in period t by the vector z_{nt} , which we treat as exogenous. Reflecting the longitudinal nature of our data, we now let w_{nt} denote the compensation of the CEO of the n^{th} firm in time t , we let x_{nt} denote the excess return of the firm gross of (or before subtracting) CEO compensation, and we denote the report of the n^{th} manager at the beginning of the t^{th} period by r_{nt} . In terms of our model, at time t the bond price is b_t , the characteristics of firm n are z_{nt} , its CEO submits his report r_{nt} , the firm realizes excess returns at the end of the period of x_{nt} , and finally the CEO is paid w_{nt} . We assume the bond price b_t is observed in the data, and in this section explain how consistent estimates of the other variables, firm characteristics, z_{nt} , reports, r_{nt} , excess returns, x_{nt} , and compensation, w_{nt} , were obtained, before describing the most salient features of the data.

4.1 Firm Characteristics and Signals

Let A_{nt} denote total assets reported at the end of the t^{th} period and $Debt_{nt}$ the level of debt reported at the end of the period. Thus $(A_{nt} - Debt_{nt})$ denotes net assets as reported in the annual report of the n^{th} firm up to the end of period t .¹⁸ Managers release information about the state of the firm through accounting statements, and exercise considerable discretion over the values which are reported. They have many ways of directly affect the firm's balance sheets, choosing for example among different valuation methods for credits and liabilities, and using discretionary timing when writing off nonperforming assets. Exercising such liberties provides a mechanism for managers to signal the state of the firm to shareholders. Thus we denote by NA_{nt} the true value of the n^{th} firm's net assets at time t , and let $Dividend_{nt}$ represent total dividends paid out during the period; we assume that $Dividend_{nt}$ is public knowledge, but that in the hybrid model only NA_{nt} is not directly observed by shareholders.

A commonly used accounting measure of the manager's accomplishments and firm's success is the difference between the change in assets and the changes in liabilities plus dividends, called comprehensive income. Normalizing comprehensive income, we define the accounting return π_{nt} for the firm in period t as:

$$\pi_{nt} = (A_{nt} - Debt_{nt} + Dividend_{nt}) / (A_{n,t-1} - Debt_{n,t-1}) \quad (37)$$

We also define the return on net assets as:

$$\pi'_{nt} = (NA_{nt} + Dividend_{nt}) / NA_{n,t-1} \quad (38)$$

¹⁸We measure financial returns annually on a calendar year basis, while the month in which annual reports are released varies by company.

These variables are used to form our measures of s_{nt} and r_{nt} . (Only the latter is observed.) Let $E[\pi_{nt} | z_{nt}]$ denote the expected accounting return of π_{nt} for firm n with characteristics z_{nt} at the beginning of period t before the manager announces total assets A_{nt} . We define the manager's report about the hidden state $r \in \{1, 2\}$ as an indicator variable, telling whether the firm's accounting return is higher or lower than the expected value of accounting returns conditional on the publicly observed state z_{nt} in that period t :

$$r_{nt} \equiv \begin{cases} 1 & \text{if } \pi_{nt} < E[\pi_{nt} | z_{nt}] \\ 2 & \text{otherwise} \end{cases}$$

Similarly define

$$s_{nt} \equiv \begin{cases} 1 & \text{if } \pi'_{nt} < E[\pi'_{nt} | z_{nt}] \\ 2 & \text{otherwise} \end{cases}$$

Note shareholders receive $Dividend_{n,t-1}$ during the course of period $t-1$, and $(A_{n,t-1} - Debt_{n,t-1})$ has been tabled by the end of period $t-1$, leaving π_{nt} to be determined by the manager's current report of $(A_{nt} - Debt_{nt})$. In the model of pure moral hazard $NA_{nt} = (A_{nt} - Debt_{nt})$ and $\pi_{nt} = \pi'_{nt}$. In the hybrid model NA_{nt} is hidden from shareholders, and the CEO can falsely report the first state has occurred by reporting a sufficiently low $(A_{nt} - Debt_{nt})$. This definition is internally consistent with the timing in our model, because the dual effects of a payoff relevant announcement by the manager during the current period are evident in a capital gain or loss through a change in the stock price immediately following, and also in the balance sheet, tabled at the end of the period. In this fashion our model captures the cumulative impact of announcements made throughout the period by their net effect on the balance sheet and shareholder equity value at the end of the period.

For convenience a finite partition was used to differentiate firms. In our study $z_{nt} \in \{1, \dots, Z\}$ are indicators of the sector to which firm n belongs in period t , its total assets at the beginning of the period (or the end of the previous period), total employment, and its debt to equity ratio. The elements of z_{nt} were defined as follows. Let $A_{t-1}(J)$ denote the median total assets for the population representing the probability distribution from which firms in sector J at the beginning of period t are drawn. That is:

$$\Pr[A \leq A_{t-1}(J)] = \Pr[A > A_{t-1}(J)] = 1/2$$

where A is defined as the total assets reported by the manager of a firm randomly sampled from the distribution of firms belonging to the J^{th} sector at the end of period $t-1$. Let $A_{n,t-1} \in \{S, L\}$ indicate whether the reported total assets of firm n lie above or below the median assets for its sector at the end of period $t-1$. Denoting by $J_n \in \{1, 2, 3\}$ the sector to which the n^{th} firm belongs, we set:

$$A_{n,t-1} = \begin{cases} S & \text{if } A_{n,t-1} \leq A_{t-1}(J_n) \\ L & \text{if } A_{n,t-1} > A_{t-1}(J_n) \end{cases}$$

Similarly suppose $W_{t-1}(J)$ denotes the median employment for the population of firms in the J^{th} sector at time $t-1$ and define the indicator variables $W_{n,t-1} \in \{S, L\}$ for whether the number of employees at firm n is above or below the median in its sector at time

$t - 1$, meaning $W_{n,t-1} = S$ if $W_{n,t-1} \leq W_{t-1}(J_n)$ and $W_{n,t-1} = L$ if $W_{n,t-1} > W_{t-1}(J_n)$. Finally let $D_{t-1}(J)$ denote the median debt to equity ratio, and let $D_{n,t-1} \in \{S, L\}$ indicate whether the debt to equity in firm n at period $t - 1$ is greater or less than $D_{t-1}(J)$. Since $z_{nt} \equiv (J_n, A_{n,t-1}, W_{n,t-1}, D_{n,t-1})$ is common knowledge, the n^{th} firm and its manager can condition any contract formed between them at the beginning of period t on z_{nt} without resorting to constraints that induce truth telling.

The medians $A_{t-1}(J)$, $W_{t-1}(J)$ and $D_{t-1}(J)$ are not directly observed, but can be estimated by their sample counterparts $A_{t-1}^{(N)}(J)$, $W_{t-1}^{(N)}(J)$ and $D_{t-1}^{(N)}(J)$. We estimate $A_{n,t-1}$ with $A_{n,t-1}^{(N)} \in \{S, L\}$, which indicates whether the total assets of firm n lie above or below the median assets for the sample firms in its sector at the beginning of period t . Thus $A_{n,t-1}^{(N)} = S$ if and only if $A_{n,t-1} \leq A_{t-1}^{(N)}(J)$. Similarly $W_{n,t-1}^{(N)} \in \{S, L\}$ indicates whether the number of employees (workers) at the firm is above or below the sample median employer, and $D_{n,t-1}^{(N)} \in \{S, L\}$ indicate whether the debt to equity in firm n at the beginning of period t is above or below the median debt to equity. Finally we also estimated the manager's announcement about the hidden state $r \in \{1, 2\}$ by comparing the firm's accounting return with the sample average for all firms in with the same publicly observed state z_{nt} in that period t :

$$r_{nt}^{(N)} \equiv \begin{cases} 1 & \text{if } \pi_{nt} \leq \frac{\sum_{m=1}^N \pi_{mt} I\{z_{mt}=z_{nt}\}}{\sum_{m=1}^N I\{z_{mt}=z_{nt}\}} \\ 2 & \text{otherwise} \end{cases}$$

4.2 Abnormal Returns and Compensation

The definition of compensation used in this study is consistent with our theoretical model, and as a practical issue, follows precedents set in the literature by Antle and Smith (1985,1986), Hall and Liebman (1998), Margiotta and Miller (2000) and Gayle and Miller (2008a, 2008b). In the optimal contract shareholders induce their manager to bear risk on only that part of the return whose probability distribution is affected by his actions. Assuming managers are risk averse, his certainty equivalent for a risk bearing security is less than the expected value of security, so shareholders would diversify amongst themselves every firm security whose returns are independent of the manager's activities, rather than use it to pay the manager. Stock and option grants are treated as directly adding to his wealth, and changes in the value of her holdings of stocks and options only affect her firm based compensation in so far as the changes are attributable to the firm's abnormal returns. Thus managerial compensation is defined as the market value of liquid and illiquid assets the manager receives (including cash and bonus, stock and option grants, pension and retirement benefits), plus the change in the value of the firm's financial securities she holds after netting out market factors, namely the changes that would have occurred if he had held a diversified portfolio instead.

Abnormal returns to the firm are defined as the residual component of returns that cannot be priced by aggregate factors the manager does not control. In an optimal contract, compensation to the manager might depend on this residual in order to provide him with appropriate incentives, but it should not depend on changes in stochastic factors that originate outside the firm, which in any event can be neutralized by adjustments within his wealth

portfolio through the other stocks and bonds he holds. More specifically, let V_{nt} denote the equity value of firm n at time t on the stock market, and let \tilde{x}_{nt} , net abnormal returns, denote the financial return on its stock net of the financial return on the market portfolio in period t . Gross abnormal returns for the n^{th} firm in period t attributable to the manager's actions are defined as net abnormal returns plus compensation as a ratio of firm equity:

$$x_{nt} \equiv \tilde{x}_{nt} + \frac{w_{nt}}{V_{n,t-1}} \quad (39)$$

Neither w_{nt} nor x_{nt} are observed. We assume that true compensation w_{nt} is measured with error, and that measured compensation, denoted \tilde{w}_{nt} , is the sum of true compensation w_{nt} plus an independently distributed disturbance term ε_{nt} , assumed orthogonal to the other variables of interest:

$$\tilde{w}_{nt} = w_{nt} + \varepsilon_{nt} \quad (40)$$

Although $(\tilde{w}_{nt}, \tilde{x}_{nt})$ rather than (w_{nt}, x_{nt}) is observed for each (n, t) , we can, however, construct consistent estimates of (w_{nt}, x_{nt}) from $(\tilde{w}_{nt}, \tilde{x}_{nt})$ given the assumption that all the covariates determining the compensation schedule, denoted $z_{nt} \in Z$, are also observed under a mild regularity condition that states net abnormal returns to shareholders increase with gross abnormal returns, meaning that whole of the increase in the firm value is not appropriated by the manager in the optimal contract.

Lemma 9 *If $V(x_2 - x_1) \neq w(x_2) - w(x_1)$ for all $(x_1, x_2) \in R^2$, then:*

$$w_{nt} = E_t[\tilde{w}_{nt} | \tilde{x}_{nt}, z_{nt}, r_{nt}, V_{n,t-1}] \quad (41)$$

This lemma implies that compensation schedule is the conditional expectation of measured compensation given net abnormal returns and lagged firm size. Pointwise consistent estimates of compensation w_{nt} can be obtained for each observation with Kernel estimators of the cross section taking the form:

$$w_{nt}^{(N)} = \frac{\sum_{m=1, m \neq n}^N \tilde{w}_{mt} I \left\{ z_{mt}^{(N)} = z_{nt}^{(N)}, r_{mt}^{(N)} = r_{nt}^{(N)} \right\} K \left(\frac{\tilde{x}_{mt} - \tilde{x}_{nt}}{\delta_x^{(N)}}, \frac{V_{m,t-1} - V_{n,t-1}}{\delta_v^{(N)}} \right)}{\sum_{n=1, m \neq n}^N I \left\{ z_{mt}^{(N)} = z_{nt}^{(N)}, r_{mt}^{(N)} = r_{nt}^{(N)} \right\} K \left(\frac{\tilde{x}_{mt} - \tilde{x}_{nt}}{\delta_x^{(N)}}, \frac{V_{m,t-1} - V_{n,t-1}}{\delta_v^{(N)}} \right)} \quad (42)$$

where $K(\cdot)$ is a bivariate probability density function with full support and $\delta^{(N)} \equiv (\delta_x^{(N)}, \delta_v^{(N)})$ is the bandwidth satisfying the convergence property $\delta^{(N)} \rightarrow 0$ as $N \rightarrow \infty$. From our estimates of w_{nt} we then constructed a consistent estimator of the gross abnormal return:

$$x_{nt}^{(N)} \equiv \tilde{x}_{nt} + \frac{w_{nt}^{(N)}}{V_{n,t-1}} \quad (43)$$

Our structural analysis inputs vectors of the form $(w_{nt}, x_{nt}^{(N)}, r_{nt}^{(N)}, z_{nt}^{(N)})$ and the subsampling methods we use to obtain test statistics and confidence regions compute the vectors in each subsample. Denote by $f_s(x|z)$ the conditional density of abnormal returns x given

the true state s and the firm's characteristics z . Under the null hypothesis that $s_t = r_t(s_t)$ consistent estimates of $f_s(x|z)$ can be obtained from nonparametric density estimator:

$$f_s^{(N)}(x|z) = \frac{\sum_{t=1}^T \sum_{n=1}^N I \left\{ z_{nt}^{(N)} = z, r_{nt}^{(N)} = s \right\} K_x \left(\frac{x_{nt}^{(N)} - x}{\delta_x^{(N)}} \right)}{\delta_x^{(N)} \sum_{t=1}^T \sum_{n=1}^N I \left\{ z_{nt}^{(N)} = z, r_{nt}^{(N)} = s \right\}} \quad (44)$$

where $K_x(\cdot)$ is a univariate probability density function with full support and the bandwidth $\delta_x^{(N)} \rightarrow 0$ as $N \rightarrow \infty$. Similarly estimates of the compensation schedule for period t are given by the nonparametric regression estimator:

$$w_{st}^{(N)}(x|z) = \frac{\sum_{n=1}^N w_{nt}^{(N)} I \left\{ z_{nt}^{(N)} = z, r_{nt}^{(N)} = s \right\} K_x \left(\frac{x_{nt}^{(N)} - x}{\delta_x^{(N)}} \right)}{\sum_{n=1}^N I \left\{ z_{nt}^{(N)} = z, r_{nt}^{(N)} = s \right\} K_x \left(\frac{x_{nt}^{(N)} - x}{\delta_x^{(N)}} \right)} \quad (45)$$

4.3 Data Summary

Our primary data source is Standard & Poor's ExecuComp database. We extracted compensation data on the current chief executive officer (CEO) of 2,610 firms in the S&P 500, Midcap, and Smallcap indices spanning the years 1992 to 2005. We supplemented these data with firm level data obtained from the S&P COMPUSTAT North America database and monthly stock price data from the Center for Securities Research (CRSP) database. The sample was partitioned into three industrial sectors by GICS code. Sector 1, called primary, includes firms in energy (GICS:1010), materials (1510), industrials (2010,2020,2030), and utilities (5510). Sector 2, consumer goods, comprises firms from consumer discretionary (2510,2520,2530,2540,2550) and consumer staples (3010,3020,3030). Firms in health care (3510,3520), financial services (4010,4020,4030,4040), information technology and telecommunication services (410, 4520, 4030, 4040, 5010) comprise Sector 3, which we call services.

Table 1 summarizes the cross sectional features of our data. Almost twice as many firms in services, as in consumerables, with the primary sector accounting for about half the observations. Average firm size by total assets is highest in the services sector and lowest in the consumer sector. This ordering is reflected by the debt equity ratio, the sector with largest firms by asset also being the most highly leveraged, but reversed when employment is used to measure firm size instead. For this reason we used both total assets and employment as two measures of size, and included the debt equity ratio as a factor that might affect the distribution of abnormal returns, and hence managerial compensation.¹⁹ In this study we assume that firm sector, the firm's total assets, the number of its employees, and its debt equity ratio, is public information.

From Table 1 we note that the average accounting return in the services industry is higher than the other two, but more remarkable is the fact that its standard deviation is much higher. This could be attributable to many factors, but we note that the services sector includes many firms that are intertwined with technological change in a rapidly changing product space, and for that reason alone might rank amongst the hardest firms to value.

¹⁹Findings of several studies, including our own, show this is indeed the case.

Table 2 summarizes the longitudinal features of our data. There are roughly the same number of observations per year, apart from 2005, where we only include data on firms whose financial records for that financial year ended before December.²⁰ In the sample period, financial returns from the stock market to diversified shareholders ranged from a yield to 45 percent in one year to a loss of 14 percent returns in another. Far greater is the variation around the market return by individual firms. As explained above, this latter variation in abnormal returns, rather than variability due to aggregate factors, is critical to explaining managerial compensation. The collective signal managers send about business, average accounting returns, is highly correlated with financial returns, almost without exception rising and falling together. Note though that accounting returns have a considerably higher standard deviation, in part attributable to fixed effects across firms, but also to higher idiosyncratic variability over time.

The term structure of interest rates underlying the bond price series were constructed from data on Treasury bills of varying maturities, and the prices were derived using methods described in Gayle and Miller (2008b). Table 2 shows that over this period, year to year bond price fluctuations are in the order of 5 to 10 percent, but there is no discernible trend in this aggregate variable. Because the optimal contract depends on bond prices, in principle, variation in bond prices can be used to identify the risk aversion parameter from the participation constraint. In practice, the variation over this period is too small to exploit in estimation.

Total assets vary a great deal by firm within and across years, growing by a factor of almost 3 over the period, with year to year standard deviations that are more than twice the mean; thus the cross sectional distribution of firm assets is skewed to the upper tail. The cross sectional distribution of employees is similarly skewed, but in contrast to assets, firm employment on average grows by less than a quarter. More remarkable than changes in annual average debt equity ratio, which ranges between 2.41 and 4.69, is its standard deviation, which varies between 5 and 105.

From Table 2 we see that the mean compensation of managers fluctuates much more than real wages for professional employees, the trough of \$1.7 million for the 12 years occurring only 2 years after the peak of \$4.7 million and just one year before the second highest, \$4.6 million. Variation in CEO compensation between firms within years is greater than the average variation over the 12 years, with a standard deviation of approximately 3 to 10 times the mean, depending on the year, although this feature of the data is partly due to individual variation, reflected in the sectorial differences evident in Table 4 discussed below. To the extent compensation depends on the firm's abnormal return, year to year fluctuations in CEO individual income is of course unpredictable.

Table 3 displays the numbers in each cell of $(z_{nt}^{(N)}, r_{nt}^{(N)})$. For the most part, the (estimated) probability of being in the bad state is higher, implying the median of $r_{nt}^{(N)}$ is less than its mean. However there are exceptions, such as $(A, W, D) = (S, S, L)$ in the primary and consumer sectors.

²⁰Paranthenetically we note that to remove the effects of the accounting month, all current values were deflated back to \$US 2000 from the month and year they accrued.

Table 4 provides a cross sectional summary of the average abnormal returns and CEO compensation conditional on the publicly observable $z_{nt}^{(N)}$ and the accounting report $r_{nt}^{(N)}$ based on the manager’s hidden information. The sample means for returns and compensation are without exception higher when a favorable report indicating the good state is released. Similarly compensation is on average higher when the good state is announced. There is a great deal of dispersion about the sample means. From the numbers of observations in each accounting state $r_{nt}^{(N)}$ provided in Table 3, we infer that many their differences are significant. By way of contrast, there are no systematic differences between sample mean returns that depend on the publicly observed states $z_{nt}^{(N)}$. Compensation tends to be higher in companies that are larger on any of the three dimensions we have measured, and also higher in the service sector.

Figure 1 depicts the estimated probability density functions for abnormal returns, and compensation schedules, in each sector for two of the eight observed states, $(A, W, D) = (S, S, S)$ and (L, L, L) , and both unobserved states. Referring to Table 3 between 1,686 and 3483 observations are used to construct each graph. The probability density functions for the good state exhibit first order stochastic dominance over the bad. This suggests that accounting measures do anticipate financial performance. Hence a manager conditions on these measures when making her effort choice. It immediately follows that these accounting variables are relevant for analyzing empirical models of moral hazard.

Our model does not predict a monotone increasing compensation schedule, nor that compensation is uniformly higher in the good state than the bad, nor that compensation under the good state is tilted to punish poor performance and reward strong results, plausible as these hypotheses might sound. Thus we should not reject the theory because the illustrated compensation schedules in Figure 1, while for the most part upward sloping, are not monotone increasing, and also cross each other more than once. The nature of this data highlight the advantages of a nonparametric approach that directly confronts the theory, effectively eliminating the possibility of spuriously rejecting auxiliary assumptions imposed to accommodate a tightly parametrized formulation of the empirical specification.

5 Testing the Models

Our estimation and testing procedures directly exploits the restrictions in $\bar{\Gamma}_1$ and $\bar{\Gamma}_2$. An alternative approach is the nested full solution method. Given a set of values for the parameters of the model, within an inner algorithm one numerically solves the optimal compensation schedule and nests the solution within a optimization routine that, in an outer algorithm, seeks the best fit with the data, using a maximum likelihood or a methods of moments estimator.²¹ This approach is easy to rationalize if the model in question is identified by a unique parameter value. But identification cannot be imposed as an assumption on the data generating process. Proposition 8 fully characterizes the identified set, and shows that in general the set is not a singleton. It is therefore unclear how the nested full solution

²¹See Ferrall and Shearer (1999), Margiotta and Miller (2000), and Gayle and Miller (2008a, 2008b) for applications of this approach to models of pure moral hazard.

approach to estimation would, in practice, handle observational equivalence between a pure model hazard model with one set of parameters and a hybrid moral hazard model with another. The nested approach to estimation is also ill suited to nonparametric specifications. Our approach avoids these limitations, by explicitly allowing for observational equivalence, and not imposing functional form assumptions on either the probability distributions generating excess returns when managers are diligent, or the unobserved distributions that would be generated if managers shirked.

5.1 The Null Hypothesis

Because Γ_i is defined by a vector function of equalities and inequalities in γ , the framework is amenable to testing whether the data generating process comes from a model of moral hazard or not. To accommodate the heterogeneity observed in our empirical application, we write $\Psi_{jt}(\gamma, z)$ for $\Psi_{jt}(\gamma)$, $\Lambda_{kt}(\gamma, z)$ for $\Lambda_{kt}(z)$, and so forth. Armed with this expanded notation, the sets Γ_i can be expressed as:

$$\Gamma_i = \{\gamma > 0 : Q_i(\gamma) = 0\}$$

for $i \in \{1, 2\}$ where:

$$Q_1(\gamma) \equiv \sum_{t=1}^T \sum_{z=1}^Z \left[\sum_{j=1}^3 \Psi_{jt}(\gamma, z)^2 + \sum_{k=1}^2 \Lambda_{kt}(\gamma, z)^2 \right]$$

$$Q_2(\gamma) \equiv \sum_{t=1}^T \sum_{z=1}^Z \left\{ \sum_{j=5}^9 \min[0, \Psi_{jt}(\gamma, z)]^2 + \sum_{j=6}^7 [\Psi_{5t}(\gamma, z) \Psi_{jt}(\gamma, z)]^2 \right. \\ \left. + \Psi_{4t}(\gamma, z)^2 + [\Psi_{6t}(\gamma, z) \Psi_{8t}(\gamma, z)]^2 + \sum_{k=3}^5 \Lambda_{kt}(\gamma, z)^2 \right\}$$

Appealing to Proposition 8, we reject the null hypothesis of a pure model of moral hazard against the more general alternative of a regular data generating process if and only if Γ_1 is empty. Similarly the hybrid model is rejected if and only if Γ_2 is empty. For $i \in \{1, 2\}$ define:

$$H_0^{(i)} : Q_i(\gamma) = 0 \text{ for some } \gamma > 0$$

$$H_A^{(i)} : Q_i(\gamma) > 0 \text{ for all } \gamma > 0$$

5.2 Constructing a Confidence Region

The empirical analysis is based on observing N firms over T periods, and the asymptotic properties described here are for large N . To test the null hypothesis we define nonparametric

estimators of $\Psi_{jt}(\gamma, z)$ and $\Lambda_{kt}(\gamma, z)$, respectively denoted by $\Psi_{jt}^{(N)}(\gamma, z)$ and $\Lambda_{kt}^{(N)}(\gamma, z)$, to form empirical analogues of $Q_i(\gamma)$, denoted by:

$$\begin{aligned}
Q_1^{(N)}(\gamma) &\equiv \sum_{t=1}^T \sum_{z=1}^Z \left[\sum_{j=1}^3 \Psi_{jt}^{(N)}(\gamma, z)^2 + \sum_{k=1}^2 \Lambda_{kt}^{(N)}(\gamma, z)^2 \right] \\
Q_2^{(N)}(\gamma) &\equiv \sum_{t=1}^T \sum_{z=1}^Z \left\{ \sum_{j=5}^9 \min \left[0, \Psi_{jt}^{(N)}(\gamma, z) \right]^2 + \sum_{j=6}^7 [\Psi_{5t}^{(N)}(\gamma, z) \Psi_{jt}^{(N)}(\gamma, z)]^2 \right. \\
&\quad \left. + \Psi_{4t}^{(N)}(\gamma, z)^2 + [\Psi_{6t}^{(N)}(\gamma, z) \Psi_{8t}^{(N)}(\gamma, z)]^2 + \sum_{k=3}^5 \Lambda_{kt}^{(N)}(\gamma, z)^2 \right\} \quad (46)
\end{aligned}$$

and a confidence region for Γ_i , defined as:

$$\Gamma_i^{(N)} \equiv \{\gamma > 0 : A_N Q_i^{(N)}(\gamma) \leq c_\delta\}$$

where A_N is the asymptotic rate of convergence of $Q_i^{(N)}(\gamma)$, and c_δ is the δ critical value of the test statistic. We reject the pure moral hazard model at level δ if $\Gamma_1^{(N)}$ is empty, and interpret $\Gamma_2^{(N)}$ in a similar manner.

The estimated functions $\Psi_{jt}^{(N)}(\gamma, z)$ and $\Lambda_{kt}^{(N)}(\gamma, z)$ are formed from estimates of their components. In the previous section we described our estimates of the compensation scheme, $w_{st}^{(N)}(x, z)$, the probability densities, $f_s^{(N)}(x, z)$, and the probabilities, $\varphi_s^{(N)}(z)$. From these estimated functions, we directly form the estimated weighted ratio $h^{(N)}(x, z)$, as well as $\Psi_{jt}^{(N)}(\gamma, z)$ for $j \in \{1, 2, 5\}$ using the definitions of $\Psi_{jt}(\gamma, z)$ given in the previous section.

To estimate $\Psi_{3t}^{(N)}(\gamma, z)$ we require an estimate of $\bar{w}_{st}(\gamma, z) \equiv \exp[-\gamma \bar{w}_{st}(z)/b_{t+1}]$. We use the fact that although \bar{w}_{st} is unknown, $w_{st}(x, z)$ is a locally non-decreasing function in x in the limit as $x \rightarrow \infty$. Following Brunk (1958), for each state $s \in \{1, 2\}$, covariate value $z \in \{1, \dots, Z\}$, and period $t \in \{1, \dots, T\}$, we rank the observations on returns in decreasing order by $x_{st}^{(1)}(z)$, $x_{st}^{(2)}(z)$, \dots and so on, denoting by $w_{st}^{(1)}(z)$, $w_{st}^{(2)}(z)$, \dots the corresponding (estimated) compensations, and estimate $\bar{w}_{st}(z)$ with:

$$\bar{w}_{st}^{(N)}(z) \equiv \max_q \sum_{r=1}^q \frac{w_{st}^{(r)}(z)}{q} \quad (47)$$

Finally to estimate $\Psi_{jt}^{(N)}(\gamma, z)$ for $j \in \{4, 7, 8, 9\}$ and $\Lambda_{kt}^{(N)}(\gamma, z)$ for $k \in \{1, \dots, 5\}$ we also require estimates of $g_s(x, z)$, which we denote by $g_s^{(N)}(\gamma, x, z)$. Note from Proposition 5 that $g_2^{(N)}(\gamma, x, z)$ can be directly found from $\bar{w}_{st}^{(N)}(z)$ but that $g_1^{(N)}(\gamma, x, z)$ also requires an estimate of $\bar{h}(z)$. L'Hospital's rule yields:

$$\bar{h}(z) = \frac{\varphi_2(z)}{\varphi_1(z)} \left\{ \lim_{x \rightarrow \infty} \left[\frac{f_2(x, z)}{f_1(x, z)} \right] \right\} = \frac{\varphi_2(z)}{\varphi_1(z)} \left\{ \lim_{x \rightarrow \infty} \left[\frac{1 - F_2(x, z)}{1 - F_1(x, z)} \right] \right\}$$

Ranking excess returns realized in the first state achieved at the end of any period $t \in \{1, \dots, T\}$, conditional on $z \in \{1, \dots, Z\}$, we obtain the decreasing sequence $x^{(1)}(z)$,

$x^{(2)}(z), \dots$ Again, following Brunk (1958), we estimated $\bar{h}(z)$ with:

$$\bar{h}^{(N)}(z) \equiv \max_q \left[\frac{\sum_{t=1}^T \sum_{n=1}^N 1 \left\{ x_{nt}^{(N)} \geq x^{(q)}(z) \right\} 1 \left\{ z_{nt}^{(N)} = z, s_{nt}^{(N)} = 2 \right\}}{q} \right] \quad (48)$$

Under standard regularity conditions $\Psi_{jt}^{(N)}(\gamma, z)$ and $\Lambda_{kt}^{(N)}(\gamma, z)$ converge in probability to $\Psi_{jt}(\gamma, z)$ and $\Lambda_{kt}(\gamma, z)$. Although $w_{st}^{(N)}(x, z)$ and $f_s^{(N)}(x, z)$ are estimated nonparametrically in the first stage, and converge pointwise at a slower rate than $N^{1/2}$, we appeal to results in Newey and MacFadden (1994) to establish:

$$N^{1/2} \left[\Psi_{jt}^{(N)}(\gamma, z) - \Psi_{jt}(\gamma, z) \right] \implies N(0, \Omega_j(\gamma, z)) \quad (49)$$

for a given $\gamma > 0$ and $j \in \{1, 2, 5\}$. However $N^{1/2}$ convergence does not necessarily extend to $\Psi_{jt}^{(N)}(\gamma, z)$ for $j \in \{3, 4, 6, \dots, 9\}$ or $\Lambda_{kt}^{(N)}(\gamma, z)$ for $k \in \{1, \dots, 5\}$, because of the components $\bar{w}_{st}^{(N)}(z)$ and $\bar{h}^{(N)}(z)$. In particular the regularity condition about the upper bound \bar{x}_{st} plays a role.

Suppose there exists a finite \bar{x}_{st} such that $F_s(\bar{x}_{st}, z) < 1$ and if $x > \bar{x}_{st}$, then $g_s(x) = 0$. In that case the derivative of $w_{st}(x)$ at \bar{x}_{st} is zero, following Parsons (1978) the norming constant is $N^{1/2}$, and hence (49) holds for all $j \in \{1, \dots, 9\}$. An analogous result applies to $\Lambda_{kt}^{(N)}(\gamma, z)$ too. However, if we relax the assumption about the existence of a finite \bar{x}_{st} , and assume, less restrictively, that $\lim_{x \rightarrow \infty} g_s(x) = 0$ then as Wright (1981) shows, the norming constant is $N^{1/3}$. In that case we replace $N^{1/2}$ with $N^{1/3}$ in (49).

Although the assumption about \bar{x}_{st} does not affect the estimation of the model or the identification results, it does affect the rate of convergence of the estimates and the asymptotic covariance matrix. Thus in our model $A_N = N^a$, where $a = 1$ if there exists a finite \bar{x}_{st} such that $F_s(\bar{x}_{st}, z) < 1$ with $g_s(x) = 0$ for all $x > \bar{x}_{st}$, but where $a = 2/3$ under the weaker assumption that $\lim_{x \rightarrow \infty} g_s(x) = 0$. If a finite \bar{x}_{st} does not exist and $A_N = N^{2/3}$, then asymptotic covariance matrix is driven by the tail observations. If a finite \bar{x}_{st} does exist and $A_N = N$, then all the observations help determine the asymptotic properties.

Assuming the following condition (10) holds, then by Lemma 3.1 of Chernozhukov, Tamer and Hong (2007), c_δ is the δ -quantile of the distribution of C_i .

Condition 10 For:

$$C_i^{(N)} = \sup_{\gamma \in \Gamma} \left[N^a Q_i^{(N)}(\gamma) \right]$$

$P\{C_i^{(N)} \leq c\} \rightarrow P\{C_i \leq c\}$ for each $c \in [0, \infty)$, where the probability distribution function for C_i is nondegenerate and continuous on $[0, \infty)$.

Since c_α is unknown, the test cannot be implemented as stated, but a modified subsampling procedure outlined in Chernozhukov, Tamer and Hong (2007), and described here for expositional convenience, can be used to obtain a consistent estimator of this critical value

and thus conduct the test. Since several of the components to the test statistic are ill defined because $v_{st}(x, 0) = 1$ for all x , the modification bounds the set of γ we consider away from zero.

Algorithm 11 (Subsampling) Consider all subsets of the data with size $N_b < N$, where $N_b \rightarrow \infty$ but $N_b/N \rightarrow 0$, and denote the number of subsets by B_N . Define c_{0i} and $\Gamma_{0i}^{(N)}$ as:

$$\begin{aligned} c_{0i} &= \inf_{\gamma > 0} \left[N^a Q_i^{(N)}(\gamma) \right] + \kappa_N \\ \Gamma_{0i}^{(N)} &= \{ \gamma \geq \gamma_N : N^a Q_i^{(N)}(\gamma) \leq c_{0i} \} \end{aligned}$$

where $\kappa_N \propto \ln N$ and γ_N converges to zero at a rate faster than N^a . For each subset $j \in \{1, \dots, B_N\}$ of size N_b define:

$$C_i^{(j, N_b)} \equiv \sup_{\gamma \in \Gamma_{0i}^{(N)}} \left[N_b^a Q_i^{(j, N_b)}(\gamma) \right]$$

Denoting the δ -quantile of the sample $\{C_i^{(1, N_b)}, \dots, C_i^{(B_N, N_b)}\}$ by $\widehat{c}_{\delta i}$, let:

$$\widehat{\Gamma}_{0i}^{(N)} = \{ \gamma > \gamma_N : N^a Q_i^{(N)}(\gamma) \leq \widehat{c}_{\delta i} \}$$

We reject the null hypothesis of private information if $\widehat{\Gamma}_{0i}^{(N)}$ is empty.

5.3 Test Results

We conducted this test separately in each sector for both the pure and the hybrid moral hazard models, simulating subsamples of $N_b = 3000$ stratified by the 16 states so that that the subsampling procedure generated the states in the proportion they were observed in the data. While the test statistics apply to a universal subsample, following empirical practice we simulated 100 draws. We tested the pure moral hazard model for every sector, only to discover the set $\widehat{\Gamma}_{01}^{(N)}$ is empty, essentially because the state dependent compensation schedules, illustrated in Figure 1 for (S, S, S) and (L, L, L) , do not satisfy the competitive selection constraint for any value of the risk aversion parameter. We therefore rejected the pure moral hazard model.

Table 5 depicts the results for the hybrid model. We cannot reject that model at the 5 percent confidence level in any sector. In both the primary and consumer goods sectors the confidence regions for the identified set of risk aversion parameters consists of two intervals, whereas in the services sector there is only one. The fact that the bands are relatively wide, especially in the primary and consumer goods sectors, should be interpreted as evidence that for a wide range of risk aversion parameters, there is little evidence against the null that managers have private information and that shareholders recognize this by the contracts they set. Moreover since there is a common region of overlap across the three sectors, namely $\gamma \in [0.0037, 0.0042]$, there is no evidence that managers with different attitudes towards risk are sorting into different sectors.

6 Estimation

Our analysis of identification and the ensuing tests showed there is a set of values Γ for the risk aversion parameter such that its elements $\gamma \in \Gamma$ are observationally equivalent. Consequently one approach to structural estimation is to delineate the set of $\theta \in \Theta$ implied by the risk aversion parameters in the identified set, thus bounding them. Rather than pursue that approach, we minimized the sum of squared criterion function $Q_2^{(N)}(\gamma)$ in γ to obtain a consistent estimator of one of the parameters in the identified set, denoted by $\hat{\gamma}$ and formed the estimators $\hat{\theta} \equiv \theta(\hat{\gamma})$. Consistent estimators of the standard errors were computed for $\hat{\gamma}$ by minimizing the criterion function $Q_2^{(j, N_b)}(\gamma)$ in γ for each subsample $j \in \{1, \dots, B_N\}$ and computing the standard deviation of $\hat{\gamma}_j$.

This section presents our estimates of the manager's preference parameters γ , α_2/α_1 and α_2 , along with their estimated standard errors, and reports on several measures of the role of hidden information and moral hazard. We estimated the losses shareholders would incur from letting the manager tend his own interests instead of maximizing the expected value of the firm before taking account of his compensation, denoted by Δ_1 , and how much managers would gain from tending their own interests instead of their firm's, denoted by Δ_2 . Their difference, $\Delta_1 - \Delta_2$, represents the net gains from internalizing conflicting goals of managers and shareholders. We also estimated how much shareholders would pay to rid the firm of moral hazard problem altogether, denoted by Δ_3 and how much shareholders would pay to make the private information public rather than induce revelation through the manager's compensation schedule, Δ_4 .

6.1 Structural Parameter Estimates

Table 6 reports our estimates of γ , α_2/α_1 and α_2 , along with their estimated standard errors. The value obtained 0.0038 for $\hat{\gamma}$ is precisely estimated, with standard error 0.0005, and is comparable to the levels to risk aversion found in previous work on managerial compensation by Margiotta and Miller (2000) and Gayle and Miller (2008a), (2008b), who applied a fully parametric estimator to data on U.S. managers from other industries over different time periods. For example our estimate of γ implies that a manager would pay \$185,600 to avoid an equal chance of losing versus gaining \$1,000,000. By way of contrast, intuitively plausible consumption based measures of this curvature parameter are notoriously difficult to obtain from applying Euler equation methods to aggregate data on consumption, or panel data on labor supply and food consumption. More generally, concentrating on a relatively homogeneous group of workers who face a high degree of uncertainty about something they deeply care about, such as annual compensation, may be key to developing our understanding about attitudes to risk.

Our estimates of α_2 and α_2/α_1 are imprecise, but the null hypotheses that $\alpha_2 = \alpha_1$ and that $\alpha_2 = 1$ for all the sectors is rejected against the one side alternatives that $\alpha_2 > \alpha_1$ and that $\alpha_2 > 1$. All the numerical estimates in the lower panel indicate $\alpha_2 > \alpha_1$, which means that the incentives of the manager are not naturally aligned with the shareholder owners. Similarly almost all the α_2 parameter estimates exceed one, which means that the manager

would not accept the position and work diligently unless he was paid compensation with a positive certainty equivalent wage. The inequality $\alpha_2 > 1$ is not implied by theory, and cannot be tested without simultaneously estimating (or knowing) the manager’s attitude towards risk. However the null hypothesis that $\alpha_2 = 1$ implies that total compensation to managers is purely a risk premium with certainty equivalent of zero, and hence that the primary attraction for holding a managerial position comes from the exercise of executive power rather than financial reward.

Both sets of estimates of α_2 and α_2/α_1 are lower, or closer to one, than those found in Gayle and Miller (2008a, 2008b) for pure moral hazard on comparable data sets. Our previous work did not, however, account for systematic variation in managerial compensation attributable to accounting information, but treated it instead as unexplained variation about true compensation. Fitting a pure moral hazard model to data which has been generated by a hybrid model, and omitting the variables that arise from private information, gives rise to two biases in statistical inference and estimation. First, the risk premium for fluctuations in compensation due to the omitted variables registers as a nonpecuniary cost of accepting a position in management, creating an upward bias in α_2 . Briefly, omitting the effects of accounting information on compensation excessively smooths fluctuations in the estimated compensation schedule, but leaves the mean intact, leading us to wrongly infer management has other undesirable characteristics. Second, because accounting returns are positively correlated with financial returns, the effect of ignoring private information is to attribute too much of the fluctuation in compensation to moral hazard. This makes shirking appear more attractive than it really is, and thus generates upwards bias in α_2/α_1 . Ignoring the role of hidden information leads us to overstate the value to the manager from not pursuing shareholder objectives.

6.2 The Importance of Hidden Information and Hidden Actions

Our measures of the role private information and moral hazard play in determining managerial compensation and the performance of the firm, are determined by the structural parameters. Gross shareholder loss from shirking, Δ_1 , is a function of the likelihood ratios $g_s(x)$. The manager’s gain from shirking, Δ_2 , depends on the ratio of utility parameters α_2/α_1 and the risk aversion parameter γ . The social cost of generalized moral hazard, Δ_3 , depends on the parameters α_2 and γ . Finally Δ_4 , the cost of hidden information, is determined by computing $w_{st}^{(M)}(x)$, optimal compensation in a pure moral hazard model.

Estimates of Δ_1 through Δ_3 are depicted in the top three panels of Table 7. They were computed from $\theta(\hat{\gamma})$ and numerically integrating over x where appropriate. The first measure, denoted Δ_1 , is the expected gross output loss to the firm switching from the distribution of abnormal returns for diligent work to the distribution for shirking, that is the difference between the expected output to the plant from the manager pursuing the firm’s goals versus his or her own, before netting out expected managerial compensation given that the state are reported truthfully. In symbols:

$$\Delta_1 \equiv E \{x [1 - g_s(x)]\} = -E [xg_s(x)]$$

where the expectation following the second equality is over (x, s) and exploits the fact that abnormal returns have mean zero, implying $E[x] = 0$.

The estimated percentage losses per year range from 0.2 to 7.6, depending on the sector and size of the firm, highly significant in 15 out of the 24 firm categories. For most categories of size and the debt equity ratio, Sector 2 firms experience higher gross losses than firms in the other two sectors. Elsewhere, in Gayle and Miller (2008b), we found that the total gross loss from shirking increases with firm size, and the result is replicated in this study. Applying our measures of market value of the firms to our estimates of Δ_1 , we see that the loss to the firm varies from \$12,000,000 to \$492,000,000 in the primary sector, \$93,000,000 to \$546,000,000 in the consumer goods sector and \$93,000,000 to \$478,000,000 in the services sector per year. However the effect of asset and employee size on gross losses from shirking per unit of equity depends on the measure of size used and the sector. For example in services, our estimates of Δ_1 are decreasing in employment but increasing in assets. In the other two sectors there is no monotone pattern using either measure, except in the consumer sector, where gross losses per unit of equity decline with increases in assets after controlling for employment and the debt equity ratio. Controlling for both measures of firm size, the gross loss is greater for those firms with a lower debt equity ratio than the median firm in two sectors, while in Sector 3 there is no significant pattern.

Partially offsetting the benefits to the firm of having the manager follow a policy of value maximization are the smaller costs of the forgone opportunity to the manager from pursuing his own goals, denoted by Δ_2 . This second measure of generalized moral hazard can be expressed as the difference between $w_t^{(2)}$, the manager's reservation certainty equivalent wage to work under perfect monitoring, given in Equation (12), and $w_t^{(1)}$, the manager's reservation certainty equivalent wage to shirk given by Equation (18). Differencing the two, and averaging over the sample time frame (during which the bond price hardly changed) we obtain:

$$\Delta_2 \equiv \frac{1}{T} \sum_{t=1}^T \left(w_t^{(2)} - w_t^{(1)} \right) = \frac{\log(\alpha_2/\alpha_1)}{\gamma T} \sum_{t=1}^T \frac{b_{t+1}}{(b_t - 1)} \quad (50)$$

Our estimates of this compensating differential are positive and highly significant in 17 firm categories, ranging from \$180,000 and \$4,640,000 per year, depending on the characteristics of the firm. Given size and debt equity, the ranking of these benefits across sectors is almost identical to the ranking of the gross losses; for example in the same five categories the consumer sector has the highest estimated Δ_1 and also the highest Δ_2 . Similarly the qualitative effects of increasing assets, given firm employment and financial leverage, are almost identical for Δ_2 as for Δ_1 ; thus increasing assets lead to increases in Δ_1 and Δ_2 in the same three categories in the primary sector, and a fall in the other one. Controlling for assets and the debt equity ratio, the net benefits from shirking to the manager increase with the number of employees at firms in the consumer sector, but no clear pattern emerges in the other two sectors. Thus our estimates show that the costs of incentivizing managers are tiny compared to the gross benefits to shareholders, confirming previous estimates found for models of pure moral hazard that a policy of paying chief executive officers a fixed wage would generate large social losses.

The other two measures show how much the firm pays to induce diligence and truthful

revelation, in other words its willingness to pay for eliminating the moral hazard problem. Under a perfect monitoring scheme shareholders would pay the manager a fixed wage of $w_t^{(2)}$. Hence the expected value of a perfect monitor to shareholders, denoted Δ_3 , is the time averaged difference between expected compensation under the current optimal scheme and $w_t^{(2)}$, or:

$$\Delta_3 \equiv \frac{1}{T} \sum_{t=1}^T \left\{ E[w_{st}(x)] - w_t^{(2)} \right\} = \frac{1}{T} \sum_{t=1}^T E[w_{st}(x)] - \frac{\log(\alpha_2)}{\gamma T} \sum_{t=1}^T \frac{b_{t+1}}{(b_t - 1)}$$

Our estimates of Δ_3 are positive and highly significant in 19 firm categories, and insignificant in the remaining ones. Conditioning on firm size, the costs of generalized moral hazard are, with just one exception (S, L, L), higher in the services than in the primary or consumer sectors. In contrast to our earlier study of pure moral hazard for a much more specialized group of industries investigated for a much longer period, firm size does not appear to play such a dominant role in explaining differences in welfare costs across the different categories. Somewhat surprisingly only in the primary sector, which coincidentally contains the three industries we examined previously, Gayle and Miller (2008b), is the original hypothesis of Berle and Means (1932), that increased employment is associated with an increased welfare cost, confirmed in all four categories of assets and debt equity ratios. More generally our estimates show that the qualitative effect of changing a firm's assets or debt equity ratio on the welfare cost depends on its characteristics.

The fourth measure is the willingness of the shareholders to pay to rid the firm of the hidden information, which is the difference in expected compensation under the existing arrangements a pure moral hazard situation:

$$\Delta_4 \equiv \frac{1}{T} \sum_{t=1}^T E \left[w_{st}(x) - w_{st}^{(M)}(x) \right]$$

where $w_{st}^{(M)}(x)$ is defined below Equation (15) and denotes the optimal compensation schedule for a pure moral hazard problem in state $s \in \{1, 2\}$. Figure 2 depicts the compensation schedule for (S, S, S) in the primary sector. For almost any given abnormal return, managers are rewarded for announcing (beforehand) that the firm is in the second state relative to the moral hazard case. Apart from the upwards shift there is little to distinguish them from each other, which reflects the fact that both $v_{st}(x)^{-1}$ and $v_{st}^{(M)}(x)^{-1} \equiv \exp \left[\gamma w_{st}^{(M)}(x) / b_{t+1} \right]$ are linear transformations in $g_2(x)$. In the bad state the contract for the hybrid model is tilted relative to the moral hazard case, inducing more uncertainty and hence lower expected utility than in the pure moral hazard contract (where expected utility is equalized across states). The tilting, found by differencing the first order conditions for the hybrid and pure moral hazard models, is due to the terms:

$$\left(\eta_{1t} - \eta_{1t}^{(M)} \right) g_2(x) - \eta_{3t} h(x) - \eta_{4t} (\alpha_1 / \alpha_2)^{1/(b_t - 1)} g_2(x) h(x)$$

While the likelihood ratio for abnormal returns between states given diligent work, $h(x)$, is for the most part increasing in x , the likelihood ratio for abnormal returns between shirking

and diligence in the second state $g_2(x)$, apparently declines more steeply, tilting the schedule to be more dependent on abnormal returns.

We emphasize the upwards tilting of the contract in the less promising state is not a theoretical implication or an internal inconsistency of our model. If the sincerity constraint does not bind, and $h(x)$ is increasing, then the schedule in the bad state would be flatter, and lower, than its moral hazard counterpart. Note also that the schedules for the pure moral hazard model are not monotone increasing in abnormal returns x . By definition $v_{st}^{(M)}(x)^{-1}$ is monotone increasing in $w_{st}^{(M)}(x)$. It follows from the first order condition that neither $g_1(x)$ or $g_2(x)$ are monotone decreasing in x , thus violating a standard monotone likelihood regularity condition useful to impose in parametric estimation, but not relevant for our analysis of identification, testing and nonparametric estimation.

Our estimates of Δ_4 are positive and highly significant in just under half the firm categories, ranging up to \$4,330,000, and more dispersed than those for Δ_3 , total welfare costs. Controlling for size and financial leverage, the costs of private information are lowest in the primary sector, achieving significance in only two firm categories, and having the lowest estimated cost in four. Similar to Δ_3 , increasing employment in the primary sector is associated with higher private information costs, but apart from that empirical regularity, there is no clear pattern relating firm size or financial leverage to the cost of private information. Appealing to the formula defining $w_{st}^{(M)}(x)$, Equation (16), the difference between Δ_3 and Δ_4 is the risk premium for paying $w_{st}^{(M)}(x)$ rather than its certainty equivalent $w_t^{(2)}$:

$$\begin{aligned} \Delta_3 - \Delta_4 &= \frac{1}{T} \sum_{t=1}^T E \left[w_{st}^{(M)}(x) \right] - w_t^{(2)} \\ &= \frac{1}{T} \sum_{t=1}^T \frac{b_{t+1}}{\gamma} E \left\{ \log \left[1 + \eta_{st}^{(M)} (\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_{st}^{(M)} g_s(x) \right] \right\} \end{aligned}$$

Alternatively stated, $\Delta_3 - \Delta_4$ is the welfare cost of moral hazard in the absence of hidden information. It is evident from the table that pure moral hazard costs exhibit less dispersion than the total welfare cost in cross section. Intuitively the contribution of private information to the cost of generalized moral hazard varies markedly by sector, firm and size and financial leverage, from no economic or statistical significance, such as (L, S, S) in the primary and consumer sectors, to almost the total cost, for example (S, L, S) in services.

7 Conclusion

A controversial topic, executive compensation is the perennial target of populist governments. One reason managers are assailed more than rock stars and professional sportsmen, whose incomes are also high, volatile and temporary, is that the compensation of a manager is partly determined by how well he claims his own firm is performing. This stylized fact is hard to reconcile with models of pure moral hazard, which predict that the expected utility of managers is equalized across states. We derive the equilibrium restrictions from models of optimal contracting to predict the shape of the compensation schedule when there are only hidden actions (pure moral hazard), and when there is hidden information as well (hybrid).

Our framework does not make parametric assumptions about the conditional distributions of abnormal returns, or the functional form of the contract. Yet our identification analysis establishes sharp and tight bounds on the identified set, in this way exhausting the empirical content of the pure and hybrid models of moral hazard we investigate. We implement tests and estimators derived from these bounds.

Our empirical investigation is based on a large panel data set measuring compensation of chief executive officers, financial and accounting returns, as well as size and sector background characteristics of the publicly trade firms they manage. In the hybrid, model we interpret data on accounting income as information reported by the CEO that cannot be fully corroborated by shareholders, in this way capturing the considerable discretion accountants and auditors wield when describing the state of their firm. Pure moral hazard models assume that accounting income can be (retrospectively) verified by shareholders.

The pure moral hazard model is rejected. In a hybrid model, with both hidden actions and hidden information, the expected utility of managers is not equalized across states. However pure moral hazard models are not nested within hybrid models. In a hybrid model, the optimal contract equates the expected utility of telling the truth and working diligently with the expected utility of lying and working diligently, and/or the expected utility of lying and shirking. The hybrid model is not rejected.

We find that the benefits of optimally deterring managers from acting against shareholder interests, and also the premium paid to executives for accepting risky compensation, are comparable to previous estimates obtained by estimating parametric models of pure moral hazard. However misspecified structural estimators of pure moral hazard models that ignore the effects of accounting information falsely attribute the positive effect of hidden information on the mean level of compensation (a risk premium in hybrid models), to nonpecuniary burdens of management. Misspecified pure models of moral hazard also overstate the value of deviating from shareholder interests, by attributing all the incentive alignment problem to unmonitored actions, thus neglecting the cost to shareholders of inducing the manager to reveal his hidden information. Similar to previous structural estimates of pure moral hazard models, our estimates show the social costs of failing to motivate managers are very large, while the social costs of motivating managers to act in the interests of shareholders are relatively modest. Last, our empirical results show that the degree of hidden information varies with firm background variables, implying that the value of accounting information to shareholders depends on firm size and sector.

8 Proofs

Proof of Lemma 1. Let λ_r be the date t price of a contingent claim made on a consumption unit at date r , implying the bond price is defined as:

$$b_t \equiv E_t \left[\sum_{r=t}^{\infty} \lambda_s \right]$$

and let q_t denote the date t price of a security that pays off the random quantity:

$$q_t \equiv E_t \left[\sum_{r=t}^{\infty} \lambda_s (\log \lambda_r - s \log \beta) \right]$$

From Equation (15) on page 680 of Margiotta and Miller (2000), the value to a manager with current wealth endowment e_{nt} , from announcing state $r_t(s)$ in period t when the true state is s , choosing effort level l_{st2} in anticipation of compensation $w_{r_t(s)t}(x)$ at the beginning of period $t + 1$ when he retires one period later, is:

$$-b_t \alpha_2^{1/b_t} \alpha_0^{1-1/b_t} \left\{ E \left[\exp \left(-\frac{\gamma w_{r_t(s)t}(x)}{b_{t+1}} \right) \right] \right\}^{1-1/b_t} \exp \left(-\frac{q_t + \gamma e_{nt}}{b_{t+1}} \right)$$

the corresponding value from choosing effort level l_{st1} is:

$$-b_t \alpha_1^{1/b_t} \alpha_0^{1-1/b_t} \left\{ E_t \left[\exp \left(-\frac{\gamma w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x)] \right] \right\}^{1-1/b_t} \exp \left(-\frac{q_t + \gamma e_{nt}}{b_{t+1}} \right)$$

whereas from their Equation (8) on page 678, the value from retiring immediately is:

$$-b_t \alpha_0 \exp \left(-\frac{q_t + \gamma e_{nt}}{b_{t+1}} \right)$$

Dividing each expression through by the retirement utility it immediately follows that the manager chooses $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$ to minimize the negative of expected utility:

$$\begin{aligned} & l_{t0} + \left(\frac{\alpha_1}{\alpha_0} l_{st1} + \frac{\alpha_2}{\alpha_0} l_{st2} \right)^{1/b_t} \left\{ E \left[\exp \left(-\frac{\gamma w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x) l_{st1} + l_{st2}] \right] \right\}^{1-1/b_t} \\ &= l_{t0} + \left\{ \left(\frac{\alpha_1}{\alpha_0} l_{st1} + \frac{\alpha_2}{\alpha_0} l_{st2} \right)^{1/(b_t-1)} E_t \left[\exp \left(-\frac{\gamma w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x) l_{st1} + l_{st2}] \right] \right\}^{(b_t-1)/b_t} \end{aligned}$$

Since $l_{t0} \in \{0, 1\}$ and $b_t > 1$ the solution to this optimization problem also solves:

$$l_{t0} + \left(\frac{\alpha_1}{\alpha_0} l_{st1} + \frac{\alpha_2}{\alpha_0} l_{st2} \right)^{1/(b_t-1)} E_t \left[\exp \left(-\frac{\gamma w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x) l_{st1} + l_{st2}] \right]$$

Multiplying through by the factor $[(\alpha_1 l_{st1} + \alpha_2 l_{st2}) / \alpha_0]^{1/(b_t-1)}$ and summing over the two states $s \in \{1, 2\}$ yields the minimand in Lemma 1. ■

Proof of Lemma 2. Multiplying each first order equation in the text by $\varphi_s v_{st}(x) f_s(x)$, then summing and integrating over x yields:

$$1 = \eta_{0t} \left[\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s v_{st}(x) f_s(x) dx \right] \equiv \eta_{0t} E[v_{st}(x)]$$

where we make use of the complementary slackness conditions. Substituting for $\eta_{0t} = E[v_{st}(x)]^{-1}$ gives the first numbered item in the lemma.

Multiplying the first order conditions for the second state by $v_{2t}(x)$, after solving for η_{0t} we obtain:

$$1 = E[v_{st}(x)]^{-1} v_{2t}(x) + \eta_{3t} v_{2t}(x) + \eta_{2t} v_{2t}(x) [(\alpha_2 / \alpha_1)^{1/(b_t-1)} - g_2(x)] + \eta_{4t} v_{2t}(x)$$

Taking the expectation with respect to x conditional on the second state occurring, and noting the incentive compatibility constraint is satisfied with equality in both states, yields:

$$\begin{aligned} 1 &= E[v_{st}(x)]^{-1} E_2[v_2(x)] + \eta_{3t} E_2[v_{2t}(x)] + \eta_{4t} E_2[v_{2t}(x)] \\ &= E_2[v_{2t}(x)] (E[v_{st}(x)]^{-1} + \eta_{3t} + \eta_{4t}) \end{aligned}$$

Dividing through by $E_2[v_{2t}(x)]$ proves the second numbered item in the lemma. ■

Proof of Lemma 3. Lemma 2, the incentive compatibility constraint for the second state, and the assumption of the lemma, successively imply:

$$\begin{aligned} \int v_{2t}(x) f_2(x) dx &\leq \int v_{1t}(x) f_1(x) dx \\ &\leq \int \left(\frac{\alpha_1}{\alpha_2}\right)^{1/(b_t-1)} v_{1t}(x) f_1(x) g_1(x) dx \\ &= \int \left(\frac{\alpha_1}{\alpha_2}\right)^{1/(b_t-1)} v_{1t}(x) f_2(x) g_2(x) dx \end{aligned}$$

■

Proof of Lemma 4. In our model the proof of Proposition 5 in Margiotta and Miller (2000) can be simply adapted to show that Theorem 3 of Fudenberg, Holmstrom and Milgrom (1990) applies, thus demonstrating that the long term optimal contract can be sequentially implemented. An induction completes the proof, by establishing that the sequential contract implementing the optimal long term contract for a manager who will retire in $\bar{\tau}$ periods replicates the one period optimal contract. In the optimal short term contract, the participation constraint is satisfied with strict equality, which implies that at the beginning of period $\bar{\tau} - 1$ the expected lifetime utility of the manager is determined by setting $t = \bar{\tau} - 1$ in the equation:

$$-b_t \alpha_0 \exp\left(-\frac{a_t + \gamma e_t}{b_t}\right) \quad (51)$$

Suppose that at the beginning of all periods $t \in \{\tau + 1, \tau + 2, \dots, \bar{\tau} - 1\}$, the expected lifetime utility of the manager is given by Equation (51). We first show the expected lifetime utility of the manager at τ is also given by Equation (51). From Lemma 1 the problem shareholders solve at τ is identical to the short term optimization problem solved in the text. In the solution to each cost minimization subproblem for the four (L_{1t}, L_{2t}) choices, the manager's participation constraint is met with equality. Consequently the manager achieves an expected lifetime utility of the manager at τ is given by Equation (51) as claimed. Therefore the problem of participating at time τ and possibly continuing with the firm for more than one period reduces to the problem of participating at time τ one period at most, solved in Lemma 1. The induction step now follows. ■

Proof of Proposition 5. Writing $\theta^* \equiv (\alpha_1, \alpha_2, g_1(x), g_2(x))$, where $v_{st}(x) \equiv \exp[-\gamma^* w_{st}(x)/b_{t+1}]$ and $\bar{v}_{st} \equiv \exp[-\gamma^* \bar{w}_{st}/b_{t+1}]$, we prove $\theta^* = \theta_t(\gamma^*)$ by successively treating each component of θ^* .

1. First we show $\alpha_2 = \alpha_{2t}(\gamma^*)$. Since the participation constraint is met with equality in the optimal contract:

$$\alpha_2 = E[v_{st}(x, \gamma^*)]^{1-b_t} = \alpha_{2t}(\gamma^*)$$

2. Proving $\eta_{2t} = \eta_{2t}(\gamma^*)$ comes from substituting the solution for η_{0t} into the first order condition for the second state, which yields:

$$v_{2t}(x)^{-1} = E[v_{st}(x)]^{-1} + \eta_{2t}[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)] + \eta_{3t} + \eta_{4t}$$

Taking expectations we obtain:

$$E_2[v_{2t}(x)^{-1}] = E[v_{st}(x)]^{-1} + \eta_{2t}[(\alpha_2/\alpha_1)^{1/(b_t-1)} - 1] + \eta_{3t} + \eta_{4t}$$

Also:

$$\bar{v}_{2t}^{-1} = E[v_{st}(x)]^{-1} + \eta_{2t}(\alpha_2/\alpha_1)^{1/(b_t-1)} + \eta_{3t} + \eta_{4t}$$

Differencing the second two equations:

$$\eta_{2t} = \bar{v}_{2t}^{-1} - E_2[v_{2t}(x)^{-1}] = \eta_{2t}(\gamma^*)$$

3. Proving $g_2(x) = g_{2t}(x, \gamma^*)$ comes from differencing:

$$v_{2t}(x)^{-1} = E[v_{st}(x)]^{-1} + \eta_{2t}[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)] + \eta_{3t} + \eta_{4t}$$

from:

$$\bar{v}_{2t}^{-1} = E[v_{st}(x)]^{-1} + \eta_{2t}(\alpha_2/\alpha_1)^{1/(b_t-1)} + \eta_{3t} + \eta_{4t}$$

to give:

$$\bar{v}_{2t}^{-1} - v_{2t}(x)^{-1} = \eta_{2t}g_2(x)$$

Upon rearrangement, we appeal to the result in Item 2, that $\eta_{2t} = \eta_{2t}(\gamma^*)$ to obtain:

$$g_2(x) = \eta_{2t}^{-1} [\bar{v}_{2t}^{-1} - v_{2t}(x)^{-1}] = g_{2t}(x, \gamma^*)$$

4. To show $\alpha_1 = \alpha_{1t}(\gamma^*)$ we substitute the solution for η_{2t} above into the first order condition for the second state evaluated at the limit $x \rightarrow \infty$ to obtain:

$$\bar{v}_{2t}^{-1} = E[v_{st}(x)]^{-1} + \{\bar{v}_{2t}^{-1} - E_2[v_{2t}(x)^{-1}]\} (\alpha_2/\alpha_1)^{1/(b_t-1)} + \eta_{3t} + \eta_{4t}$$

or, upon appealing to Lemma 2:

$$\begin{aligned} (\alpha_2/\alpha_1)^{1/(b_t-1)} &= \frac{\bar{v}_{2t}^{-1} - E[v_{st}(x)]^{-1} - \eta_{3t} - \eta_{4t}}{\bar{v}_{2t}^{-1} - E_2[v_{2t}(x)^{-1}]} \\ &= \frac{\bar{v}_{2t}^{-1} - E[v_{2t}(x)]^{-1}}{\bar{v}_{2t}^{-1} - E_2[v_{2t}(x)^{-1}]} \end{aligned}$$

Making α_1 the subject of the equation:

$$\alpha_1 = \alpha_2 \left[\frac{\bar{v}_{2t}^{-1} - E[v_{2t}(x)]^{-1}}{\bar{v}_{2t}^{-1} - E_2[v_{2t}(x)^{-1}]} \right]^{1-b_t} = \alpha_{1t}(\gamma^*)$$

5. To prove $\eta_{4t} = \eta_{4t}(\gamma^*)$ we first multiply the first order conditions for the first state by $v_{1t}(x)$, after solving for η_{0t} to obtain:

$$\begin{aligned} & 1 - \eta_{1t} v_{1t}(x) [(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x)] \\ &= \{E[v_{st}(x)]\}^{-1} v_{1t}(x) - \eta_{3t} v_{1t}(x) h(x) - \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t}(x) g_2(x) h(x) \end{aligned}$$

Conditioning on the first state and taking expectations with respect to x yields:

$$1 = \{E[v_{st}(x)]\}^{-1} E_1[v_{1t}(x)] - \eta_3 E_1[v_{1t}(x) h(x)] - \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x) g_2(x) h(x)]$$

since the incentive compatibility condition drops out. Substituting out the solution for:

$$\eta_{3t} = \{E_2[v_{2t}(x)]\}^{-1} - E[v_{st}(x)]^{-1} - \eta_{4t}$$

we obtained from lemma 2 reduces this expression to:

$$\begin{aligned} 1 &= E[v_{st}(x)]^{-1} E_1[v_{1t}(x)] - \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x) g_2(x) h(x)] \\ &\quad - \{E_2[v_{2t}(x)]^{-1} - E[v_{st}(x)]^{-1} - \eta_{4t}\} E_1[v_{1t}(x) h(x)] \end{aligned}$$

Upon collecting terms:

$$\begin{aligned} & \eta_{4t} \{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x) g_2(x) h(x)] - E_1[v_{1t}(x) h(x)]\} \\ &= \{E[v_{st}(x)]\}^{-1} E_1[v_{1t}(x)] - E_1[v_{1t}(x) h(x)] \{E_2[v_{2t}(x)]^{-1} - E[v_{st}(x)]^{-1}\} - 1 \end{aligned}$$

so solving for η_4 we now have:

$$\begin{aligned} \eta_{4t} &= \frac{E[v_{st}(x)]^{-1} E_1[v_{1t}(x)] - E_1[v_{1t}(x) h(x)] \{E_2[v_{2t}(x)]^{-1} - E[v_{st}(x)]^{-1}\} - 1}{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x) g_2(x) h(x)] - E_1[v_{1t}(x) h(x)]} \\ &= \eta_{4t}(\gamma^*) \end{aligned}$$

6. Proving $\eta_{3t} = \eta_{3t}(\gamma^*)$ follows directly from the lemma above, which implies:

$$\eta_{3t} \equiv E_2[v_{2t}(x)]^{-1} - \eta_{4t}(\gamma^*) - E[v_{st}(x)]^{-1}$$

7. To prove $\eta_{1t} = \eta_{1t}(\gamma^*)$, rewrite the first order condition for the first state as:

$$\eta_{1t} [(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x)] = v_{1t}(x)^{-1} - E[v_{st}(x)]^{-1} + \eta_{3t} h(x) + \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x) h(x)$$

At the limit $x \rightarrow \infty$ we have:

$$\eta_{1t} (\alpha_2/\alpha_1)^{1/(b_t-1)} = \bar{v}_{1t}^{-1} - E[v_{st}(x)]^{-1} + \eta_{3t} \bar{h} \quad (52)$$

Making η_{1t} the subject of the equation now demonstrates $\eta_{1t} = \eta_{1t}(\gamma^*)$.

8. Differencing the first order condition for the first state and its limit as $x \rightarrow \infty$ gives:

$$\eta_{1t} g_1(x) = \bar{v}_{1t}^{-1} - v_{1t}(x)^{-1} + \eta_{3t} [\bar{h} - h(x)] - \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x) h(x)$$

Dividing both sides by η_{1t} we thus establish $g_1(x) = g_1(x, \gamma^*)$ for all t .

■

Proof of Proposition 6. We prove $\Psi_{1t}(\gamma) = \Psi_{2t}(\gamma) = 0$, derive the expression for $\Psi_{4t}(\gamma)$ given by (32) and show the sincerity constraint implies $\Psi_{6t}(\gamma) \geq 0$. Setting $\eta_{3t} = \eta_{4t} = 0$, and appealing to symmetry implies from the statement of Proposition 1 that:

$$\alpha_1 = \alpha_2 \left\{ \frac{\bar{v}_{1t}^{-1} - E[v_{1t}(x)]^{-1}}{\bar{v}_{1t}^{-1} - E_2[v_1(x)]^{-1}} \right\}^{1-b_t} = \alpha_2 \left\{ \frac{\bar{v}_{2t}^{-1} - E[v_{2t}(x)]^{-1}}{\bar{v}_{2t}^{-1} - E_2[v_2(x)]^{-1}} \right\}^{1-b_t}$$

which proves $\Psi_{1t}(\gamma) = 0$. Alternatively the equality follows as a special case for the proof of $\Psi_{2t}(\gamma) = 0$.

The expected value of the first order condition in the first state can be expressed as:

$$\eta_{1t} [(\alpha_2/\alpha_1)^{1/(b_t-1)} - 1] = E_1[v_{1t}(x)]^{-1} - E[v_{st}(x)]^{-1} + \eta_{3t} \frac{\varphi_2}{\varphi_1} + \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[g_2(x)h(x)]$$

Differencing this expression from (52) yields:

$$\eta_{1t} = \bar{v}_{1t}^{-1} + \eta_{3t} \bar{h} - E_1[v_{1t}(x)]^{-1} - \eta_{3t} \frac{\varphi_2}{\varphi_1} - \eta_{4t} (\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[g_2(x)h(x)]$$

Rearranging this expression then yields the equality satisfied by $\Psi_{4t}(\gamma)$.

The sincerity constraint is:

$$E_2 \left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t}(x) g_2(x) - v_{2t}(x) \right] \geq 0$$

Substituting $(\gamma, \theta(\gamma))$ for $(\gamma^*, \theta(\gamma^*))$ in this inequality we obtain:

$$\begin{aligned} 0 &\leq E_2 \left[[\alpha_{1t}(\gamma)/\alpha_{2t}(\gamma)]^{1/(b_t-1)} v_{1t}(x, \gamma) g_2(x, \gamma) - v_{2t}(x, \gamma) \right] \\ &= E_2 \left[\left\{ \frac{1 - \bar{v}_{st}(\gamma) E_2[v_{2t}(x, \gamma)]^{-1}}{1 - \bar{v}_{st}(\gamma) E_2[v_{2t}(x, \gamma)]^{-1}} \right\} \left\{ \frac{1 - \bar{v}_{2t}(\gamma)/v_{2t}(x, \gamma)}{1 - \bar{v}_{2t}(\gamma) E_2[v_{2t}(x, \gamma)]^{-1}} \right\} v_{1t}(x, \gamma) - v_{2t}(x, \gamma) \right] \\ &= E_2 \left[\left\{ \frac{1 - \bar{v}_{2t}(\gamma)/v_{2t}(x, \gamma)}{1 - \bar{v}_{st}(\gamma) E_2[v_{2t}(x, \gamma)]^{-1}} \right\} v_{1t}(x, \gamma) - v_{2t}(x, \gamma) \right] \\ &\equiv \Psi_{6t}(\gamma) \end{aligned}$$

■

Proof of Proposition 7. We prove each numbered item in order.

1. The fact that $\alpha_{2t}(\gamma) > 0$ immediately follows from its definition, and hence $\alpha_{1t}(\gamma) > 0$ from its definition too. By Jensen's inequality:

$$\{E_2[v_{2t}(x, \gamma)]\}^{-1} < E_2[v_{2t}(x, \gamma)]^{-1}$$

so:

$$1 - \bar{v}_{st}(\gamma) \{E_2[v_{2t}(x, \gamma)]\}^{-1} > 1 - \bar{v}_{st}(\gamma) E_2[v_{2t}(x, \gamma)]^{-1}$$

and consequently the quotient:

$$\frac{\alpha_{1t}(\gamma)}{\alpha_{2t}(\gamma)} = \left\{ \frac{1 - \bar{v}_{st}(\gamma) E_2 [v_{2t}(x, \gamma)]^{-1}}{1 - \bar{v}_{st}(\gamma) E_2 [v_{2t}(x, \gamma)^{-1}]} \right\}^{1-b_t} < 1$$

since $b_t > 1$. Thus $\alpha_{1t}(\gamma) < \alpha_{2t}(\gamma)$.

2. For $\gamma \in \bar{\Gamma}_2$ the participation constraint implies:

$$\begin{aligned} b_{t+1} \log [\alpha_{2t}(\gamma)] &= b_{t+1} \log \left\{ \sum_{s=1}^2 \varphi_s E_s [\exp [-\gamma w_{st}(x) / b_{t+1}]] \right\}^{1-b_t} \\ &= (1 - b_t) b_{t+1} \log \left\{ \sum_{s=1}^2 \varphi_s E_s [\exp [-\gamma w_{st}(x) / b_{t+1}]] \right\} \\ &< (1 - b_t) b_{t+1} \left\{ \sum_{s=1}^2 \varphi_s E_s [-\gamma w_{st}(x) / b_{t+1}] \right\} \\ &= -\gamma (1 - b_t) \left\{ \sum_{s=1}^2 \varphi_s E_s [w_{st}(x)] \right\} \\ &= \gamma (b_t - 1) \left\{ \sum_{s=1}^2 \varphi_s E_s [w_{st}(x)] \right\} \end{aligned}$$

The proof for $\gamma \in \bar{\Gamma}_1$ is found by replacing $\sum_{s=1}^2 \varphi_s E_s [\exp [-\gamma w_{st}(x) / b_{t+1}]]$ with $E_s [\exp [-\gamma w_{st}(x) / b_{t+1}]]$ for $s \in \{1, 2\}$.

3. It follows directly from the definition of $g_{st}(x, \gamma)$ that $g_{st}(x, \gamma) \rightarrow 0$ as $x \rightarrow \infty$ for $s \in \{1, 2\}$, and that $E_s [g_{st}(x, \gamma)] = 1$.

4. Since $g_2(x) \rightarrow 0$ as $x \rightarrow \infty$ and $g_2(x) > 0$, it follows from the first order condition associated with the second state that $v_{2t}(x) \geq \bar{v}_2$ for all $x \in R$, so from the definition of $v_{2t}(x)$ we infer:

$$\bar{w}_{2t} \equiv \lim_{x \rightarrow \infty} [w_{2t}(x)] = \sup_{x \in R} [w_{2t}(x)]$$

Hence $v_{2t}(x, \gamma) \geq \bar{v}_{2t}(\gamma)$ for all $\gamma > 0$ and $x \in R$, thus proving from its definition that $g_{2t}(x, \gamma) \geq 0$. In the pure moral hazard model, $\eta_{3t} = \eta_{4t} = 0$, and hence the same logic shows $g_{1t}(x, \gamma) \geq 0$ if $\gamma \in \bar{\Gamma}_1$. Finally the inequality $\Psi_{9t}(\gamma) \geq 0$ guarantees $g_{1t}(x, \gamma) \geq 0$ for $\gamma \in \bar{\Gamma}_2$.

5. In the proof of Item 2 of Proposition 1 we established $\eta_{2t} \equiv \bar{v}_{2t}^{-1} - E_2 [v_{2t}(x)^{-1}]$, and also proved in Item 2 above that $\bar{v}_{2t} \leq v_{2t}(x)$. Hence $\eta_{2t} \geq 0$. The same arguments apply to $\eta_{1t}(\gamma)$ when $\gamma \in \bar{\Gamma}_1$ thus establishing $\eta_{1t}(\gamma) \geq 0$ in that case.

■

Proof of Proposition 8. For $j \in \{1, 2\}$ in the pure moral hazard model, and for $j \in \{1, \dots, 4\}$ in the hybrid model, define $\eta_{jt}(\gamma)$ using the formulas in Proposition 5 with respect to a regular data generation process (s_t, x_t) and the $w_{st}(x)$ schedule, and similarly let $\eta_{0t}(\gamma) \equiv E \{ \exp [-\gamma w_{st}(x) / b_{t+1}] \}^{-1}$. We show $w_{st}(x)$ satisfies the first order conditions

of the Kuhn Tucker formulation of an optimal contracting problem for a model with parameterization $(\hat{\gamma}, \hat{\theta})$, and that $\hat{\eta}_{jt} \equiv \eta_{jt}(\hat{\gamma})$ are Kuhn Tucker multipliers associated with this solution. Since the objective function for the underlying maximization problem is strictly concave, and the constraints are linear, the first order and complementary slackness conditions in the Kuhn Tucker formulation uniquely determine the solution to the optimal contracting problem, thus proving the proposition. It is convenient to treat the pure and hybrid moral hazard models separately.

Considering first models of pure moral hazard, we show that the participation and incentive compatibility constraints are met with equality, that the first order condition is satisfied, and that $\hat{\eta}_{jt} > 0$ solve the equations defining the Kuhn Tucker multipliers.

1. Since $\Psi_{1t}(\hat{\gamma}) = \Psi_{2t}(\hat{\gamma}) = 0$, the competitive selection constraints ensure:

$$E_1 [v_{1t}(x, \hat{\gamma})]^{1-b_t} = E_2 [v_{2t}(x, \hat{\gamma})]^{1-b_t}$$

for all $t \in \{1, \dots, T\}$, so from the definition of $\hat{\alpha}_2$ we have:

$$\hat{\alpha}_2 \equiv E [v_{st}(x, \hat{\gamma})]^{1-b_t} = E_1 [v_{1t}(x, \hat{\gamma})]^{1-b_t} = E_2 [v_{2t}(x, \hat{\gamma})]^{1-b_t}$$

implying the participation constraint is met with equality in each state, as required by the solution to the optimization problem.

2. From the definitions of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{g}_s(x)$:

$$\hat{g}_s(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} = \hat{\eta}_{st}^{-1} [\bar{v}_{st}(\hat{\gamma})^{-1} - v_{st}(x, \hat{\gamma})^{-1}] - \left\{ \frac{\bar{v}_{st}(\hat{\gamma})^{-1} - E_s [v_{st}(x, \hat{\gamma})]^{-1}}{\bar{v}_{st}(\hat{\gamma})^{-1} - E_s [v_{st}(x, \hat{\gamma})]^{-1}} \right\}$$

Multiplying both sides by:

$$\hat{\eta}_{st} \equiv \bar{v}_{st}(\hat{\gamma})^{-1} - E_s [v_{st}(x, \hat{\gamma})^{-1}]$$

we obtain:

$$\begin{aligned} \hat{\eta}_{st} \left[\hat{g}_s(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] &= [\bar{v}_{st}(\hat{\gamma})^{-1} - v_{st}(x, \hat{\gamma})^{-1}] - [\bar{v}_{st}(\hat{\gamma})^{-1} - E_s [v_{st}(x, \hat{\gamma})]^{-1}] \quad (51) \\ &= E_s [v_{st}(x, \hat{\gamma})]^{-1} - v_{st}(x, \hat{\gamma})^{-1} \end{aligned}$$

Substituting $\eta_{0t}(\hat{\gamma})$ for $E_s [v_{st}(x, \hat{\gamma})]^{-1}$ yields the first order condition in for $v_{st}(x, \hat{\gamma})$ state $s \in \{1, 2\}$ of the pure moral hazard problem for the parameterization $(\hat{\gamma}, \hat{\theta})$.

3. From its definition $\hat{\eta}_{st} > 0$. Also the equation above implies:

$$\left[\hat{g}_s(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] = \hat{\eta}_{st}^{-1} [E_s [v_{st}(x, \hat{\gamma})]^{-1} - v_{st}(x, \hat{\gamma})^{-1}]$$

Multiplying through by $v_{st}(x, \hat{\gamma})$ and taking the expectation with respect to x conditional on the s^{th} state yields:

$$\begin{aligned} E_2 \left\{ \left[\hat{g}_s(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] v_{st}(x, \hat{\gamma}) \right\} &= \hat{\eta}_{st}^{-1} E_s \left\{ \left[E_s[v_{st}(x, \hat{\gamma})]^{-1} - v_{st}(x, \hat{\gamma})^{-1} \right] v_{st}(x, \hat{\gamma}) \right\} \\ &= \hat{\eta}_{st}^{-1} E_s \left\{ v_{st}(x, \hat{\gamma}) / E_s[v_{st}(x, \hat{\gamma})] - 1 \right\} \\ &= 0 \end{aligned}$$

Therefore the incentive compatibility condition is satisfied with equality in each state $s \in \{1, 2\}$.

4. By construction $\eta_{0t}(\hat{\gamma})$ is the Lagrange multiplier for the participation constraint. From the Equation (54):

$$E_2 \left\{ v_{st}(x, \hat{\gamma}) \left[\hat{g}_s(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] \right\} = 0$$

Using the first order condition as expressed by Equation (53) to substitute out $v_{st}(x, \hat{\gamma})$ yields $\hat{\eta}_{st}$ as a solution to the equation:

$$E_2 \left\{ \frac{\hat{g}_s(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)}}{\eta \hat{g}_s(x) - \eta (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} - E_s[v_{st}(x, \hat{\gamma})]^{-1}} \right\} = 0$$

in η for $s \in \{1, 2\}$, which define the Kuhn Tucker multipliers associated with incentive compatibility. This completes the proof for the pure moral hazard case.

We now consider the hybrid case, and suppose $\hat{\gamma} \in \bar{\Gamma}_2$. We show that the participation and incentive compatibility constraints are met with equality, that the first order conditions hold, that the truth telling and sincerity constraints are satisfied, and that $\hat{\eta}$ solves the equations defining the Kuhn Tucker multipliers.

1. The definition of $\hat{\alpha}_2 \equiv \{E[v_{st}(x, \hat{\gamma})]\}^{1-b_t}$ directly implies the participation constraint is met with equality. The truth telling and sincerity constraints are directly imposed from $\hat{\gamma}$ belonging to $\bar{\Gamma}_2$ through the equality $\Psi_{7t}(\hat{\gamma}) \Psi_{8t}(\hat{\gamma}) = 0$ and the two inequalities $\Psi_{7t}(\hat{\gamma}) \geq 0$, $\Psi_{8t}(\hat{\gamma}) \geq 0$. This only leaves the three tasks of establishing $(\hat{\gamma}, \hat{\theta})$ satisfies the first order conditions, that the incentive compatibility conditions are met with equality, and that $\hat{\eta}_{1t}$ through $\hat{\eta}_{4t}$ together solve the equations defining the Kuhn Tucker multipliers.
2. Noting the definitions of $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{g}_2(x)$ and $\hat{\eta}_{2t}$ are identical to their counterparts in the pure moral hazard model we appeal to Item 2 in the pure moral hazard case to establish:

$$\hat{\eta}_{2t} \left[\hat{g}_2(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] = E_2[v_{2t}(x, \hat{\gamma})]^{-1} - v_{st}(x, \hat{\gamma})^{-1}$$

From the definition of $\widehat{\eta}_{3t}$ we have:

$$\{E[v_{st}(x, \widehat{\gamma})]\}^{-1} + \widehat{\eta}_{3t} + \widehat{\eta}_{4t} = E_2[v_{2t}(x, \widehat{\gamma})]^{-1}$$

Subtracting the first equation from the second and substituting $\widehat{\eta}_{0t}$ for $E[v_{st}(x, \widehat{\gamma})]^{-1}$ we obtain the first order condition for the second state in the hybrid model:

$$\widehat{\eta}_{0t} + \widehat{\eta}_{3t} + \widehat{\eta}_{2t} \left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_2(x) \right] + \widehat{\eta}_{4t} = v_{2t}(x, \widehat{\gamma})^{-1}$$

Turning to the first state, the definition of $\widehat{g}_1(x)$ implies:

$$\begin{aligned} \widehat{\eta}_{1t}\widehat{g}_1(x) &= \bar{v}_{1t}(\widehat{\gamma})^{-1} - v_{1t}(x, \widehat{\gamma})^{-1} + \widehat{\eta}_{3t}[\bar{h} - h(x)] \\ &\quad - \widehat{\eta}_{4t}(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}\widehat{g}_2(x)h(x) \end{aligned}$$

From the definition of $\widehat{\eta}_{1t}$:

$$\eta_{3t}\bar{h} = \widehat{\eta}_{1t}(\alpha_2/\alpha_1)^{1/(b_t-1)} - E[v_{st}(x, \widehat{\gamma})]^{-1} - \bar{v}_{1t}(\widehat{\gamma})^{-1}$$

Substituting out $\eta_{3t}\bar{h}$ in the expression above for $\widehat{\eta}_{1t}\widehat{g}_1(x)$, and again noting $\widehat{\eta}_{0t} \equiv E[v_{st}(x, \widehat{\gamma})]^{-1}$ now yields the first order condition in the first state upon rearrangement:

$$v_{1t}(x, \widehat{\gamma})^{-1} = \widehat{\eta}_{0t} + \widehat{\eta}_{1t} \left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_1(x) \right] - \widehat{\eta}_{3t}h(x) - \widehat{\eta}_{4t}(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}\widehat{g}_2(x)h(x)$$

3. Next we show that the incentive compatibility constraints are satisfied with equality. In the second state, we again appeal to the fact that the definitions of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{g}_2(x)$ and $\widehat{\eta}_{2t}$ are identical to their counterparts in the pure moral hazard model, which implies from Item 2 in the moral hazard case that:

$$\widehat{\eta}_{2t} \left[\widehat{g}_2(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \right] = E_2[v_{2t}(x, \widehat{\gamma})]^{-1} - v_{2t}(x, \widehat{\gamma})^{-1}$$

Multiplying by $v_{2t}(x, \widehat{\gamma})$ and taking expectations conditional on the second state then proves the incentive compatibility constraint is satisfied with equality in the second state:

$$\begin{aligned} &E_2 \left\{ \widehat{\eta}_{2t}v_{2t}(x, \widehat{\gamma}) \left[\widehat{g}_2(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \right] \right\} \\ &= E_2 \left\{ \widehat{\eta}_{2t}v_{2t}(x, \widehat{\gamma}) \left[E_2[v_{2t}(x, \widehat{\gamma})]^{-1} - v_{2t}(x, \widehat{\gamma})^{-1} \right] \right\} \\ &= 0 \end{aligned}$$

Multiplying the expression we derived for the first order condition in Item 2 above by $v_{1t}(x, \widehat{\gamma})$, using the identity $\widehat{\eta}_{0t} \equiv E[v_{st}(x, \widehat{\gamma})]^{-1}$, and taking the expectation conditional on the first state yields implies:

$$\begin{aligned} &\widehat{\eta}_{1t}E_1 \left\{ v_{1t}(x, \widehat{\gamma}) \left[\widehat{g}_1(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \right] \right\} \\ &= E_1[v_{1t}(x, \widehat{\gamma})] E[v_{st}(x, \widehat{\gamma})]^{-1} - \widehat{\eta}_{3t}E_1[v_{1t}(x, \widehat{\gamma})h(x)] \\ &\quad - \widehat{\eta}_{4t}(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)} E_1[v_{1t}(x, \widehat{\gamma})\widehat{g}_2(x)h(x)] - 1 \end{aligned}$$

We now prove the right side of this equation is zero by successively substituting the definitions of $\widehat{\eta}_{3t}$ and $\widehat{\eta}_{4t}$ into it:

$$\begin{aligned}
& E_1 [v_{1t}(x, \widehat{\gamma})] E [v_{st}(x, \widehat{\gamma})]^{-1} - \widehat{\eta}_{3t} E_1 [v_{1t}(x, \widehat{\gamma})h(x)] \\
& - \widehat{\eta}_{4t} (\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)} E_1 [v_{1t}(x, \widehat{\gamma})\widehat{g}_2(x)h(x)] - 1 \\
= & E_1 [v_{1t}(x, \widehat{\gamma})] E [v_{st}(x, \widehat{\gamma})]^{-1} - \{E_2 [v_{2t}(x, \widehat{\gamma})]^{-1} - E [v_{st}(x, \widehat{\gamma})]^{-1} - \widehat{\eta}_{4t}\} E_1 [v_{1t}(x, \widehat{\gamma})h(x)] \\
& - \widehat{\eta}_{4t} (\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)} E_1 [v_{1t}(x, \widehat{\gamma})\widehat{g}_2(x)h(x)] - 1 \\
= & E_1 [v_{1t}(x, \widehat{\gamma})] E [v_{st}(x, \widehat{\gamma})]^{-1} - \{E_2 [v_{2t}(x, \widehat{\gamma})]^{-1} - E [v_{st}(x, \widehat{\gamma})]^{-1}\} E_1 [v_{1t}(x, \widehat{\gamma})h(x)] - 1 \\
& + \widehat{\eta}_{4t} \left\{ E_1 [v_{1t}(x, \widehat{\gamma})h(x)] - (\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)} E_1 [v_{1t}(x, \widehat{\gamma})\widehat{g}_2(x)h(x)] \right\} \\
= & 0
\end{aligned}$$

Therefore:

$$\widehat{\eta}_{1t} E_1 \left\{ v_{1t}(x, \widehat{\gamma}) \left[\widehat{g}_1(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \right] \right\} = 0$$

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

4. Turning now to the four equations defining Lagrangian multipliers, it follows from the definition of the mappings $\widehat{v}_{st}^{-1}(x, \widehat{\gamma})$, $\widehat{g}_s(x)$ and $\widehat{\alpha}_s$ for $s \in \{1, 2\}$ and the defined elements $\widehat{\eta}_{1t}$ through $\widehat{\eta}_{4t}$ that:

$$\begin{aligned}
\widehat{v}_{1t}(x, \widehat{\gamma})^{-1} &= E [v_{st}(x, \widehat{\gamma})]^{-1} + \widehat{\eta}_{1t} \left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_1(x) \right] \\
&\quad - \widehat{\eta}_{3t}h(x) - \widehat{\eta}_{4t} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \widehat{g}_2(x)h(x)
\end{aligned}$$

and

$$\widehat{v}_{2t}^{-1}(x, \widehat{\gamma}) = E [v_{st}(x, \widehat{\gamma})]^{-1} + \widehat{\eta}_{2t} [(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_2(x)] + \widehat{\eta}_{3t} + \widehat{\eta}_{4t}$$

Substituting for $\widehat{v}_{1t}(x, \widehat{\gamma})^{-1}$ and $\widehat{v}_{2t}^{-1}(x, \widehat{\gamma})$ in the truth telling constraint:

$$\begin{aligned}
0 &= E_2 [\widehat{v}_{1t}(x, \widehat{\gamma}) - \widehat{v}_{1t}(x, \widehat{\gamma})] \\
&= E_2 \left\{ \left[\begin{array}{c} E [v_{st}(x, \widehat{\gamma})]^{-1} + \widehat{\eta}_{1t} \left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_1(x) \right] \\ - \widehat{\eta}_{3t}h(x) - \widehat{\eta}_{4t} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \widehat{g}_2(x)h(x) \end{array} \right]^{-1} \right\} \\
&\quad - E_2 \left\{ \left[E [v_{st}(x, \widehat{\gamma})]^{-1} + \widehat{\eta}_{2t} [(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_2(x)] + \widehat{\eta}_{3t} + \widehat{\eta}_{4t} \right]^{-1} \right\}
\end{aligned}$$

Similarly, since the incentive compatibility and first order conditions are satisfied with equality in each state by $\widehat{\gamma}$, $\widehat{\theta}$ and $\widehat{\eta}_{0t} \equiv E [v_{st}(x, \widehat{\gamma})]^{-1}$:

$$\begin{aligned}
0 &= E_1 \left\{ v_{1t}(x, \widehat{\gamma}) \left[\widehat{g}_1(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \right] \right\} \\
&= E_1 \left[\frac{\widehat{g}_1(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}}{\left(\begin{array}{c} E [v_{st}(x, \widehat{\gamma})]^{-1} + \widehat{\eta}_{1t} \left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_1(x) \right] \\ - \widehat{\eta}_{3t}h(x) - \widehat{\eta}_{4t} (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \widehat{g}_2(x)h(x) \end{array} \right)} \right]
\end{aligned}$$

and:

$$\begin{aligned}
0 &= E_2 \left\{ \widehat{v}_{2t}^{-1}(x, \widehat{\gamma}) \left[\widehat{g}_2(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} \right] \right\} \\
&= E_2 \left[\frac{\widehat{g}_2(x) - (\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}}{\{E[v_{st}(x, \widehat{\gamma})]\}^{-1} + \widehat{\eta}_{2t}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_2(x)] + \widehat{\eta}_{3t} + \widehat{\eta}_{4t}} \right]
\end{aligned}$$

From its definition $\widehat{\eta}_{4t} = 0$ when $\Psi_{6t}(\widehat{\gamma}) > 0$, and when $\Psi_{6t}(\widehat{\gamma}) = 0$:

$$\begin{aligned}
0 &= E_2[v_{2t}(x, \widehat{\gamma})] - (\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)} E_2[v_{1t}(x, \widehat{\gamma})\widehat{g}_2(x)] \\
&= E_2 \left\{ \frac{1}{1 + \widehat{\eta}_{2t}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_2(x)] + \widehat{\eta}_{3t} + \widehat{\eta}_{4t}} \right\} \\
&\quad - E_2 \left\{ \frac{(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}\widehat{g}_2(x)}{\left(\begin{array}{c} \{E[v_{st}(x, \widehat{\gamma})]\}^{-1} + \widehat{\eta}_{1t}[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)} - \widehat{g}_1(x)] \\ -\widehat{\eta}_{3t}h(x) - \widehat{\eta}_{4t}(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}\widehat{g}_2(x)h(x) \end{array} \right)} \right\}
\end{aligned}$$

This equations demonstrate that $\widehat{\eta}_{1t}$ through $\widehat{\eta}_{4t}$ solve the equations defining the Lagrange multipliers for the parameterization defined by $(\widehat{\gamma}, \widehat{\theta})$ in the hybrid moral hazard model, thus completing the proof of the proposition.

■

Proof of Lemma 9. For notational convenience, and without loss of generality, we suppress the dependence of compensation w_{nt} on $(z_{nt}, s_{nt}, b_t, b_{t+1})$. Let \tilde{x} denote the net excess returns, x gross excess returns, $w(x)$ the compensation schedule as a mapping from gross excess returns, and let V denote the value of the firm at the beginning of the period. By our definition of net and gross excess returns:

$$\tilde{x} = x - w(x) / V \tag{55}$$

Suppose there exists for some (\tilde{x}_0, V_0) two distinct values of net excess returns, denoted $x_1 \in R$ and $x_2 \in R$, satisfying Equation (55). Then:

$$\tilde{x}_0 = x_i - w(x_i) / V_0$$

for $i \in \{1, 2\}$ which implies:

$$V_0(x_2 - x_1) = w(x_2) - w(x_1)$$

But this possibility is ruled out as a possibility in the premise of the Lemma. Therefore a unique solution to the relation defined by Equation (55) exists for each pair (\tilde{x}, V) , and we can denote the solution mapping by $x \equiv X(\tilde{x}, V)$. Substituting $X(\tilde{x}, V)$ for x in $w(x)$ we define $\Lambda(\tilde{x}, V) \equiv w[X(\tilde{x}, V)]$. The lemma now follows because the measurement error on compensation is assumed to independent of (\tilde{x}, V) , so $E[\tilde{w}|\tilde{x}, V] = \Lambda(\tilde{x}, V)$. ■

References

- [1] Andrews, D., S. Berry and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location", Department of Economics, Yale University.
- [2] Antle, R. and Smith, A. (1985): "Measuring Executive Compensation: Methods and an Application", *Journal of Accounting Research*, **23**, 296-325.
- [3] Antle, R. and Smith, A. (1986): "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives", *Journal of Accounting Research*, **24**, 1-39.
- [4] Bajari, P., L. Bankard and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition", *Econometrica*, **75**, 1331-1371.
- [5] Baker, G., M. Gibbs and B. Holmstrom (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data", *Quarterly Journal of Economics*, **109**, 881-919.
- [6] Baker, G., M. Gibbs and B. Holmstrom (1994b): "The Wage Policy of a Firm", *Quarterly Journal of Economics*, **109**, 921-955.
- [7] Berle, A. A., and G. C. Means (1932): *The Modern Corporation and Private Property*. Harcourt, Brace, & World Inc., New York.
- [8] Bebchuk, L. and J. Fried (2003): "Executive Compensation as an Agency Problem", *Journal of Economic Perspectives*, **17**, 71-92.
- [9] Bertrand, M. and S. Mullainathan (2001): "Are CEOs Rewarded for Luck? The Ones without Principals Are", *Quarterly Journal of Economics*, **116**, 901-932.
- [10] Bolton, P. and M. Dewatripont (2005): *Contract Theory*, MIT Press, Cambridge.
- [11] Brunk, H. D. (1958): "On the Estimation of Parameters Restricted by Inequalities", *Annals of Mathematical Statistics*, **29**, 437-454.
- [12] Chernozhukov, V., H. Hong and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243-1284.
- [13] Chesher, A. (2007): "Instrumental Values", *Journal of Econometrics*, **139**, 15-34.
- [14] Chiappori, P.A. and B. Salanie (2000): "Testing Contract Theory: A Survey of Some Recent Work", in *Advances in Economics and Econometrics*, **1**, M. Dewatripont, L.Hansen and S. Turnovsky eds. , Cambridge University Press.
- [15] Ciliberto, F. and E. Tamer (2007): "Market Structure and Multiple Equilibria in Airline Markets", Department of Economics, Northwestern University.
- [16] Dubois P. and T. Vukina (2005): "Optimal Incentives under Moral Hazard and Heterogeneous Agents: Evidence from Production Contracts Data", Working paper

- [17] Dufflo E., R. Hanna and S. Ryan (2007): "Monitoring Works: Getting Teachers to Come to School", Working paper, MIT.
- [18] Dye, R. A. and S. S. Sridhar (2005): "Moral Hazard Severity and Contract Design", *RAND Journal of Economics*, **36**, 78-92.
- [19] Ferrall C. and B. Shearer (1999): "Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records", *Review of Economic Studies*, **66**, 309-338.
- [20] Finnerty, J.E. (1976): "Insiders and Market Efficiency", *Journal of Business*, **31**, 1141-1148.
- [21] Fudenberg, D., B. Holmstrom and P. Milgrom (1990): "Short-Term Contracts and Long-Term Agency Relationships", *Journal of Economic Theory*, **50**, 1-31.
- [22] Gayle, G.-L. and R. A. Miller (2008a): "Has Moral Hazard become a More important Factor in Managerial Compensation?", Forthcoming *American Economic Review*.
- [23] Gayle, G.-L. and R. A. Miller, (2008b): "Insider Information and Performance Pay", Forthcoming, *CESifo Economic Studies*.
- [24] Gayle, G.-L. and R. A. Miller (2009): "The Empirical Content of Generalized Models of Moral Hazard", Tepper School of Business, Carnegie Mellon University.
- [25] Gayle, G.-L., L. Golan and R. A. Miller (2008b): "Are There Glass Ceilings for Female Executives", Tepper School of Business, Carnegie Mellon University.
- [26] Gayle, G.-L., L. Golan and R. A. Miller (2008b): "Promotion, Turnover and Compensation in the Executive Market", Tepper School of Business, Carnegie Mellon University.
- [27] Gayle, G.-L., C. Li and R. A. Miller (2009): "The Unintended Consequences of the Sarbanes-Oxley Act on CEO Compensation", Tepper School of Business, Carnegie Mellon University.
- [28] Green, J. R. and J.-J. Laffont (1986): "Partially Verifiable Information and Mechanism Design", *Review of Economic Studies*, **53**, 447-456.
- [29] Grossman, S. and O. Hart (1983): "An Analysis of the Principal-Agent Problem", *Econometrica*, **51**, 7-46.
- [30] Hall, B. J. and J. B. Liebman, (1998): "Are CEOs Really Paid Like Bureaucrats?", *Quarterly Journal of Economics*, **113**, 653-680.
- [31] Haubrich, J. (1994): "Risk Aversion, Performance Pay, and the Principal Agent Problem," *Journal of Political Economy*, **102**, 258-276.
- [32] Ho, K. (2008): "Insurer-Provider Networks in the Medical Care Market" , Forthcoming, *American Economic Review*.

- [33] Ho, K., J. Ho, and J. Mortimer (2008): "The Effects of Full-Line Forcing Contracts" Department of Economics, Columbia University.
- [34] Huang Y., I. Perrigne, and Q. Vuong (2007): "Nonlinear Pricing in Yellow Pages", Department of Economics, Pennsylvania State University.
- [35] Jaffe, J.F. (1974): " Special information and Insider Trading", *Journal of Business*, **47**, 410-428.
- [36] Laffont, J.-J. and D. Martimort (2002): *The Theory of Incentives*, Princeton University Press, Princeton.
- [37] Levine, A. (2008): "Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry", Department of Economics, Stanford University.
- [38] Lorie, J.H. and V. Niederhoffer (1968): " Predictive and Statistical Properties of Insider Trading", *Journal of Law and Economics*, **11**, 35-51.
- [39] Malcomson, J. M. and F. Spinnewyn (1988): " The Multiperiod Principal-Agency Problem", *Review of Economic Studies*, **55**, 391-407.
- [40] Margiotta, M. and R. A. Miller (2000): "Managerial Compensation and The Cost of Moral Hazard", *International Economic Review*, **41**, 669-719.
- [41] Mirrlees, J. A. (1975): "The Theory of Moral Hazard and Unobservable Behavior, Part 1", mimeo, Oxford.
- [42] Murphy, K. (1999): "Executive Compensation", in *Handbook of Labor Economics*, 3, O. Ashenfelter and D. Card, eds., Amsterdam: North-Holland, 2485-2563.
- [43] Myerson, R. (1982): "Optimal Coordination Mechanisms in Generalized Principal Agent Problem", *Journal of Mathematical Economics*, **10**, 67–81.
- [44] Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in *Handbook of Econometrics*, eds. R. F. Engle and D. L. McFadden. Amsterdam: North-Holland. 2111–2245.
- [45] Pakes, A., J. Porter, K. Ho and J. Ishii (2006): " Moment Inequalities and Their Application", CEMMAP Working Paper.
- [46] Parsons, V. L. (1978): "The Limiting Distribution of the Isotonic Estimator at a Point", Unpublished Manuscript.
- [47] Perrigne, I. and Q. Vuong (2007): "Nonparametric Identification of Incentive Regulation Models", CAPCP Working Paper.
- [48] Prendergast, C. (1999): "The Provision of Incentives in Firms", *Journal of Economic Literature*, **37**, 7-63.

- [49] Rey, P. and B. Salanie (1990): "Long-term, Short-term and Renegotiation: On the Value of Commitment in Contracting", *Econometrica*, **58**, 597-619.
- [50] Romano, J.P. and A.M. Shaikh (2006): "Inference for Identifiable Parameters in Partially Identified Econometric Models", working paper.
- [51] Seyhun, H. N. (1986): "Insiders' Profits, Cost of Trading, and Market Efficiency", *Journal of Financial Economics*, **16**, 189-212.
- [52] Seyhun, H. N (1992a): "The Effectiveness of Insider Trading Sanctions", *Journal of Law and Economics*, **35**, 149-182.
- [53] Seyhun, H. N (1992b): " Why Does Aggregate Insider Trading Predict Future Stock Returns?", *Quarterly Journal of Economics*, **107**, 1303-1331.
- [54] Shearer, B. (2004): " Piece Rates, Fixed Wages and Incentives: Evidence From a Field Experiment", *Review of Economic Studies*, **71**, 513-534.
- [55] Wright, F.T. (1981): " The Asymptotic Behavior of Monotone Regression Estimates", *Annals of Statistics*, **9**, 443-448.

Table 1
 Cross Section Summary of Firm Characteristics by Sector
 (Assets in millions of 2000 \$US, Employees in thousands,
 Compensation in thousands of 2000 \$US)

Variable	Primary	Consumer	Services
Observations	8,980	6,762	11,144
Assets	6,322 (27773)	5,277 (22124)	17,776 (67133)
Employees	15.8 (40.8)	32.23 (78.75)	11.9 (26.59)
Debt/Equity Ratio	2.07 (40.9)	1.94 (26.21)	4.56 (50.63)
r_{nt}	1.15 (4.54)	1.13 (1.68)	1.28 (7.26)
Market Value	6,480 (25160)	7,811 (21975)	11,664 (35002)

Table 2
Time Series Summary of Firm Characteristics and Compensation
(Assets in millions of 2000 \$US, Employees in thousands, Compensation in thousands of \$US)

Year	Bond	Assets	Employees	Debt/Equity	x_t	r_{nt}	Compensation	Observation
1993	15.9	8896 (26269)	18.02 (46.15)	2.83 (7.24)	1.19 (0.45)	1.18 (0.51)	1854 (12412)	1574
1994	13.72	7770 (25284)	16.18 (43.41)	2.87 (5.04)	0.97 (0.29)	1.07 (2.52)	2714 (10909)	1876
1995	14.00	8187 (28650)	16.43 (44.41)	3.45 (33.4)	1.26 (0.47)	1.18 (0.64)	1781 (13252)	1867
1996	13.79	8357 (29029)	17.31 (45.92)	2.41 (17.2)	1.16 (0.38)	1.17 (0.87)	3257 (14824)	1926
1997	13.67	8770 (31797)	17.94 (47.96)	2.76 (41.4)	1.30 (0.48)	1.22 (3.06)	4691 (17791)	1997
1998	15.00	9486 (40145)	17.67 (45.91)	3.91 (71.3)	1.05 (0.53)	1.20 (1.11)	2726 (18530)	2012
1999	13.97	10303 (43087)	18.34 (45.75)	2.84 (11.57)	1.14 (0.76)	1.31 (8.27)	1652 (21631)	1970
2000	13.18	10484 (45936)	19.59 (54.08)	2.64 (8.31)	1.14 (0.68)	1.18 (1.5)	4624 (21641)	1865
2001	14.16	12015 (52064)	20.10 (56.50)	2.69 (14.9)	1.08 (0.54)	1.17 (1.86)	3314 (18842)	1851
2002	14.32	12115 (57166)	19.47 (54.51)	4.69 (105)	0.86 (0.42)	0.996 (2.43)	3165 (16077)	1877
2003	14.87	13869 (66331)	19.15 (52.85)	2.51 (35.2)	1.45 (0.64)	1.53 (16.1)	3151 (18830)	1814
2004	14.17	14429 (70812)	21.05 (64.83)	2.77 (9.39)	1.16 (0.37)	1.11 (1.38)	4069 (17195)	1687
2005	13.89	20925 (89832)	22.19 (52.34)	2.63 (12.27)	1.07 (0.36)	1.16 (1.63)	4397 (19992)	751

Table 3
Estimate of $\varphi(j)$ Probabilities

S_{1t}	N	Primary		N	Consumer		N	Services		Total	
		Good	Bad		Good	Bad		Good	Bad		
(S,S,S)	1	2598	0.0917	0.1975	2023	0.1227	0.1764	3483	0.1249	0.1877	8103
(S,L,S)	2	319	0.0141	0.0214	268	0.0121	0.0275	210	0.0040	0.0149	797
(S,L,L)	3	469	0.0257	0.0266	418	0.0229	0.0389	1210	0.0337	0.0749	2097
(S,S,L)	4	1326	0.0763	0.0713	961	0.0725	0.0696	952	0.0434	0.0421	3239
(L,S,S)	5	541	0.0272	0.0331	498	0.0308	0.0427	760	0.0248	0.0434	1799
(L,L,S)	6	1105	0.0635	0.0595	734	0.0593	0.0493	927	0.0164	0.0668	2766
(L,L,L)	7	2398	0.1118	0.1552	1686	0.0879	0.1614	3056	0.0865	0.1878	7140
(L,S,L)	8	224	0.0127	0.0123	175	0.0145	0.0114	546	0.0262	0.0227	945
Total		8980	0.423	0.577	6762	0.423	0.577	11,144	0.360	0.640	26886

Table 4
 Cross Section Summary of Returns and Compensation by State and Sector
 (Compensation in thousands of 2000 \$US)

S_{1t}	Variable	Primary		Consumer		Services	
		Good	Bad	Good	Bad	Good	Bad
Return							
(S,S,S)	1	0.11 (0.56)	-0.097 (0.43)	0.095 (0.57)	-0.15 (0.47)	0.28 (0.94)	-0.05 (0.70)
(S,L,S)	2	-0.03 (0.40)	-0.14 (0.34)	-0.007 (0.31)	-0.07 (0.35)	0.02 (0.43)	-0.10 (0.46)
(S,L,L)	3	-0.03 (0.36)	-0.11 (0.37)	0.07 (0.38)	-0.12 (0.36)	0.10 (0.41)	-0.05 (0.34)
(S,S,L)	4	0.07 (0.52)	-0.11 (0.44)	0.05 (0.62)	-0.11 (0.55)	0.20 (0.75)	-0.11 (0.82)
(L,S,S)	5	-0.005 (0.34)	-0.10 (0.39)	0.006 (0.41)	-0.12 (0.40)	0.08 (0.71)	-0.09 (0.52)
(L,L,S)	6	-0.07 (0.29)	-0.11 (0.33)	-0.06 (0.32)	-0.12 (0.34)	0.15 (0.61)	-0.05 (0.47)
(L,L,L)	7	-0.03 (0.27)	-0.13 (0.30)	-0.005 (0.41)	-0.16 (0.38)	0.02 (0.32)	-0.06 (0.37)
(L,S,L)	8	0.02 (0.30)	-0.13 (0.40)	0.07 (0.47)	-0.24 (0.49)	0.13 (0.85)	-0.12 (0.59)
Compensation							
(S,S,S)	1	3889 (14651)	670 (10779)	3397 (19178)	-1501 (15235)	6063 (20034)	1701 (17316)
(S,L,S)	2	4384 (9381)	2339 (14243)	4922 (30677)	-486 (23882)	8015 (24615)	-1183 (25740)
(S,L,L)	3	3742 (11903)	521 (15710)	9194 (19898)	821 (11820)	7096 (14740)	2274 (14363)
(S,S,L)	4	2522 (9855)	721 (8851)	3977 (14844)	908 (11504)	4154 (16068)	-150 (14255)
(L,S,S)	5	3079 (20381)	-850 (15773)	4235 (20107)	-510 (16940)	3386 (18844)	1629 (19287)
(L,L,S)	6	4154 (13375)	2422 (16220)	4727 (20989)	-429 (21784)	8035 (24244)	5496 (26472)
(L,L,L)	7	5781 (12807)	2200 (12208)	6897 (19288)	2775 (19118)	9846 (24075)	5595 (19936)
(L,S,L)	8	4396 (14831)	-3729 (18890)	4742 (19288)	-2442 (14448)	5647 (20347)	1718 (17612)

Table 5 : Subsampling Results

	Primary	Consumer	Services
N	7796	5600	8536
N_b	3000	3000	3000
5% critical value	15	2	10
5% Confidence Interval ρ	$[2.1e-05, 0.0026] \cup [0.0037, 0.0042]$	$[2.1e-05, 0.0013] \cup [0.0019, 0.0057]$	$[0.0027, 0.0053]$
Number replication	100	100	100

Table 6
Structural Estimates of Manager's Preferences

Parameter			Full Sample			Estimates	S.E.
γ			0.0038			0.0005	
	S_{1t}	Primary Estimates	S.E..	Consumer Estimates	S.E..	Services Estimates	S.E..
α_2/α_0	(S,S,S)	1.0015	6.7E-04	0.9971	1.7E-03	1.0011	6.5E-04
	(S,L,S)	1.0051	2.7E-03	0.9986	5.1E-03	0.9997	5.1E-03
	(S,L,L)	1.0027	1.5E-03	1.0087	5.9E-03	1.0081	2.9E-03
	(S,S,L)	1.0023	9.9E-04	1.0007	9.2E-04	1.0014	9.9E-04
	(L,S,S)	1.0001	2.7E-03	1.0015	2.2E-03	1.0016	1.9E-03
	(L,L,S)	1.0078	3.4E-03	1.0029	2.2E-03	1.0121	4.7E-03
	(L,L,L)	1.0086	3.6E-03	1.0097	5.8E-03	1.0164	5.8E-03
	(L,S,L)	0.9909	1.8E-03	0.9997	3.1E-03	1.0065	2.8E-03
α_2/α_1	(S,S,S)	1.0011	1.2E-03	1.0057	3.8E-03	1.0026	9.7E-04
	(S,L,S)	1.0031	1.3E-03	1.0124	9.4E-03	1.0026	2.9E-02
	(S,L,L)	1.0001	4.0E-04	1.0027	3.3E-03	1.0021	6.9E-04
	(S,S,L)	1.0005	2.4E-04	1.0027	1.8E-03	1.0017	1.3E-03
	(L,S,S)	1.0093	3.9E-03	1.0039	6.2E-03	1.0048	2.4E-03
	(L,L,S)	1.0011	1.2E-03	1.0054	3.9E-03	1.0021	1.1E-03
	(L,L,L)	1.0015	4.8E-03	1.0032	2.8E-03	1.0041	1.3E-03
	(L,S,L)	1.0032	1.4E-03	1.0008	1.2E-03	1.0031	1.2E-03

Table 7
Structural Estimates of the Costs and Benefits
of Hidden Information and Actions

(Δ_1 is measured in percentage per year

Δ_2 , Δ_3 and Δ_4 are measured in 100,000 of 2000 \$US)

	S_{1t}	Primary		Consumer		Services	
		Estimates	S.E..	Estimates	S.E..	Estimates	S.E..
Δ_1	(S,S,S)	0.7	0.4	3.5	0.7**	2.2	0.5**
	(S,L,S)	3.9	1.4**	7.0	2.7**	0.8	6.2
	(S,L,L)	1.3	0.9**	2.2	1.7	2.0	2.5
	(S,S,L)	0.2	.07**	3.1	0.6**	2.4	1.7
	(L,S,S)	7.6	1.4**	2.4	2.6	4.1	2.1**
	(L,L,S)	1.6	1.6	4.7	1.3**	0.9	1.9
	(L,L,L)	1.6	0.6**	1.8	0.8**	3.8	0.7**
	(L,S,L)	4.2	1.5**	1.2	1.3	3.8	1.1**
Δ_2	(S,S,S)	3.8	1.9**	21.4	4.6**	9.9	2.6**
	(S,L,S)	11.4	4.5**	46.4	18.5**	9.6	13.4
	(S,L,L)	2.8	1.5	10.1	8.9	7.8	3.7**
	(S,S,L)	1.8	0.7**	10.01	2.4**	6.5	5.4
	(L,S,S)	35.2	8.9**	14.7	13.0	18.3	7.5**
	(L,L,S)	4.2	3.4	20.1	6.3**	7.8	10.4
	(L,L,L)	5.7	1.7**	12.1	3.7**	15.4	4.3**
	(L,S,L)	12.0	5.6**	3.1	4.3**	11.2	4.6**
Δ_3	(S,S,S)	7.9	1.1**	17.1	2.1**	23.8	2.1**
	(S,L,S)	23.9	4.6**	5.7	13.5	43.9	18.2**
	(S,L,L)	19.0	4.5**	7.3	4.7	12.7	6.1**
	(S,S,L)	3.2	1.3**	12.3	1.9**	13.3	3.0**
	(L,S,S)	-1.0	5.3	10.0	5.0**	21.8	5.7**
	(L,L,S)	8.1	3.8**	15.9	5.4**	16.7	9.0
	(L,L,L)	9.8	2.6**	8.5	4.2**	11.9	5.4**
	(L,S,L)	4.7	4.3	14.1	5.2**	19.0	4.4**
Δ_4	(S,S,S)	1.12	1.4	15.0	1.1**	7.7	1.8**
	(S,L,S)	4.2	0.9**	4.3	9.8	43.3	4.4**
	(S,L,L)	2.8	1.4**	2.7	1.2**	2.4	2.8
	(S,S,L)	1.0	1.8	8.0	1.5**	11.8	5.6**
	(L,S,S)	-0.2	2.2	-0.1	1.0	1.4	1.13
	(L,L,S)	3.4	5.4	3.6	0.8**	3.3	1.7**
	(L,L,L)	3.4	4.0	4.1	9.2	2.48	2.8
	(L,S,L)	0.05	0.7	0.1	0.8	-0.08	0.9

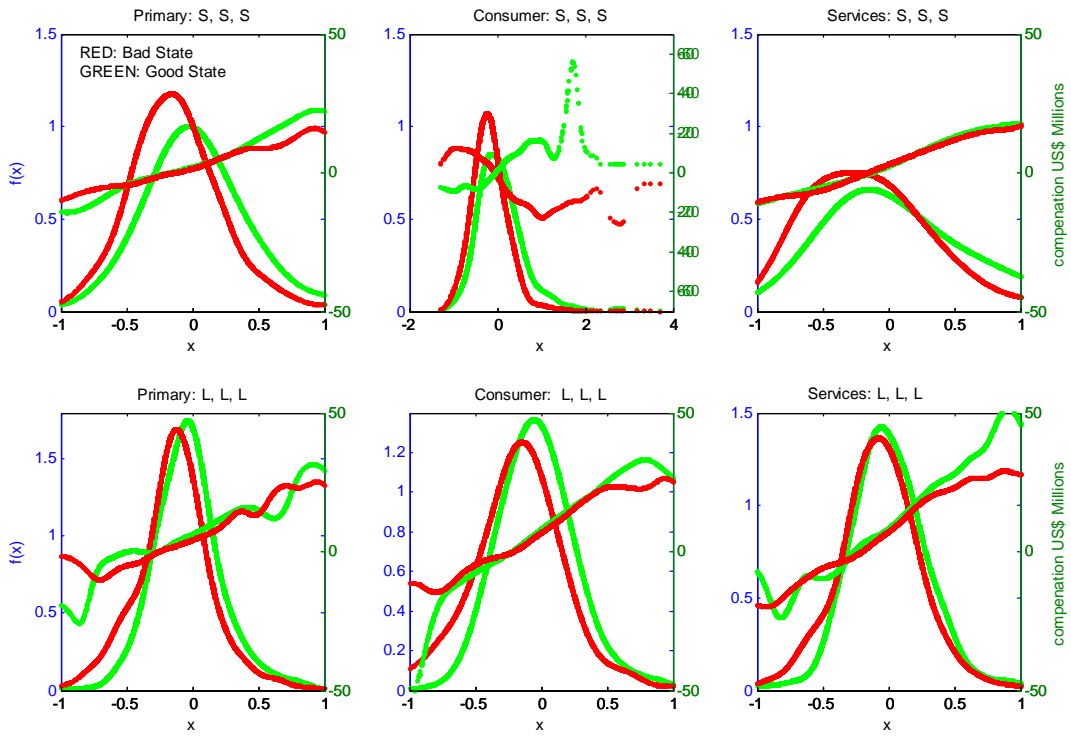


Figure 1: Compensation Schedules and Financial Return Densities

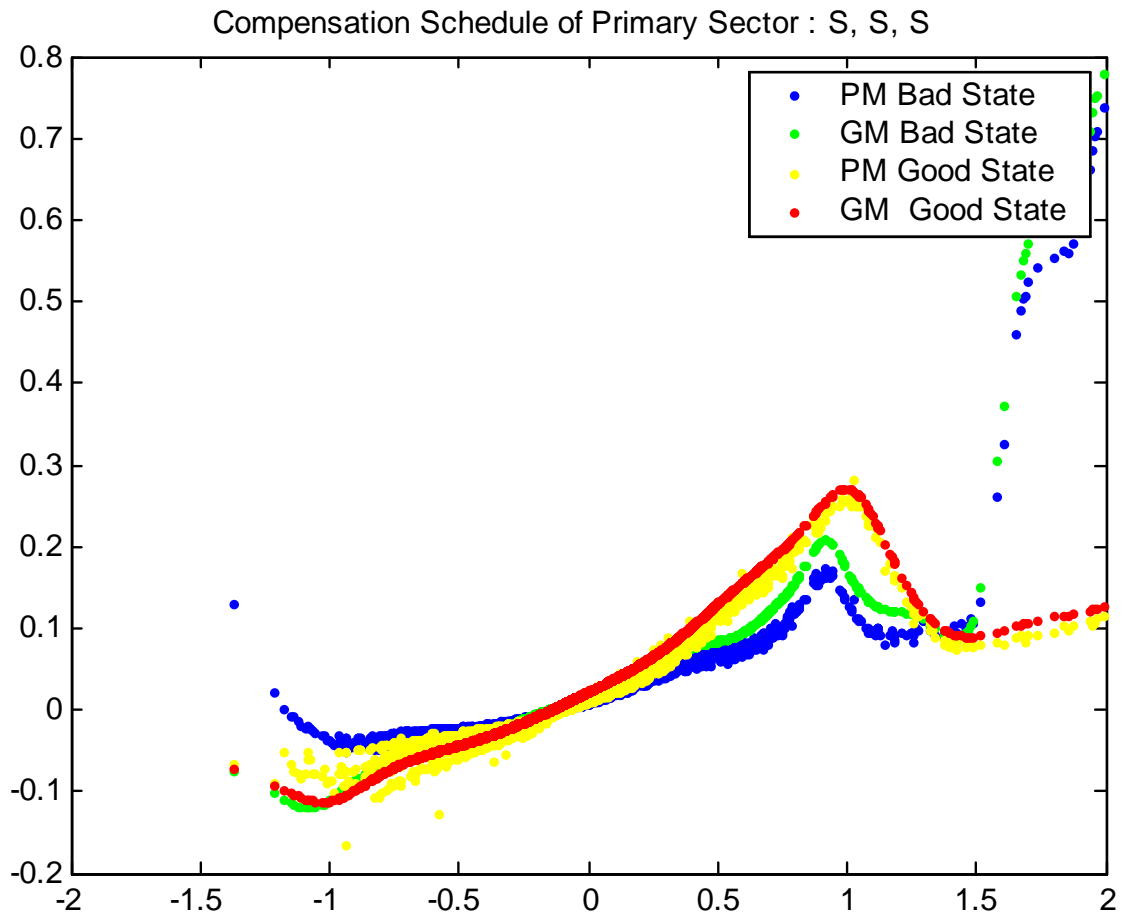


Figure 2: Predicted Pure versus Actual Generalized Compensation Schedules