IDENTIFYING AND TESTING GENERALIZED MORAL HAZARD MODELS OF MANAGERIAL COMPENSATION

George-Levi Gayle
and Robert A. Miller*

In the first part of this paper we exploit restrictions from the theory of optimal contracting to fully characterize set identification in a class of generalized moral hazard models, and demonstrate how the effects of hidden information are differentiated from the effects of pure moral hazard. Then we apply nonparametric methods to test the model, and quantify the various factors identified in the first stage, using a large longitudinal data set on chief executive officers. Our empirical study provides the first structural estimates of the importance of hidden information relative to pure moral hazard in executive compensation packages.

Keywords: Private information, Hidden actions, Optimal contracting, Set identification, Semiparametric estimation, Moment inequalities, Executive Compensation, Accounting Information, Financial returns, Welfare costs.

1. INTRODUCTION

This paper fully characterizes the empirical content of an important set of principal agent models with asymmetric information and moral hazard. From the equations and inequalities defining the optimal contract in a generalized moral hazard model, we derive the subset of models that generate the observationally equivalent data processes characterizing the joint distribution of private information, hidden actions, the return to a risk neutral principal and the risk averse agent's compensation. Thus the first part of this paper exploits restrictions from the theory of optimal contracting to establish whether the data can be rationalized by any generalized moral hazard model within the set we consider, define the observationally equivalent subset, and demonstrate how the data generating process differentiate private information from hidden actions. Then we apply our identification results to a large panel containing data on the compensation of chief executive officers and the firms they manage. 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 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 incentive regulation, and Huang, Perrigne and Vuong (2007), who nonparametrically identify and estimate

*We thank the participants of Econometric Society 2007 Summer Meetings, SITE Theory-Based Micro-Econometric Modelling Workshop 2008, Society of Economic Dynamics 2008 and seminar participants at Carnegie Mellon University, University of Pittsburgh, Georgetown University, Queens University and the University of Toranto for comments and suggestions.

a nonlinear pricing model of advertising. Nevertheless our work differs from theirs in many respects. The differences range from the theoretical structures considered (such as discrete versus continuous choices by the agent), to the results obtained (for example element versus set identification), and the nature of empirical problems to which the identification analysis applies (executive compensation versus regulation and nonlinear pricing).

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. The genesis for the class of generalized moral hazard models we study is in Myerson (1982), our least cost approach to optimal contracting extends the pure moral hazard model developed in Grossman and Hart (1980), and the decentralization of the optimal long term contract follows from results similar to those derived in Fudenberg, Holmstrom and Milgrom (1990).

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. This is accomplished by deriving sharp and tight bounds on the parameter space within a structural framework derived from the theory of optimal mechanism design, without imposing strong parametric assumptions on the conditional distributions of abnormal returns, or the functional form of the contract.

Section 4 motivates our empirical study by briefly reviewing the published empirical evidence showing how much managers exploit private information about their firms for personal financial gain. Then we describe the sources of data on chief executive officers of publicly traded companies, explain how the data were compiled, and summarize the 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) and Standard and Poors' Compusat data bases, as well as background characteristics on the sector and size of the firms.

We develop test statistics for both the pure and hybrid models in Section 5 and present our results for the CEO compensation data. The methods we develop to achieve set identification, and test the specification, draw from Chernozhukov, Hong and Tamer (2007) and Romano and Shaikh (2006), while our semi-parametric estimators exploit moment inequalities, in this latter respect similar to recent parametric estimation of structural models in industrial organization by Andrews, Berry and Jia (2004), Ciliberto and Tamer (2007), Ho (2008), Ho, Ho and Mortimer (2008), Levine (2007), and Pakes, Porter, Ho, and Ishii (2006). Our tests reject the pure moral hazard model, but we cannot reject the hybrid model, where there are hidden actions and private information.

Section 6 reports our estimates of the various components to the costs of hidden information and moral hazard. Our empirical study is one of several that take a structural approach to investigating the role of labor market incentives, including Ferrall and Shearer (1999) and Shearer (2004) on payment schedules to plant trees, Dubois and Vukina (2005) on livestock contracts with farmers, Dufflo, Hann and Ryan (2007) for motivating teachers. Even more closely related is the structural estimation conducted by Margiotta and Miller (2000), who estimate a parametric structural model of executive compensation when there is pure moral hazard, Gayle and Miller (2008b) who apply the identification results in this paper to estimate how the costs of motivating managers have changed over time, and Golan, Gayle and Miller (2008), who apply our identification and estimation techniques to examine career incentives and turnover within executive markets. In this paper we find that the benefits of contracting to deter managers from deviating from shareholder interests, and also the risk premium paid to executives for taking uncertain pay, are comparable to previous estimates obtained by estimating parametric models of pure moral hazard, and that the degree of private information varies considerably across sectors and over firm size.

2. THE MODEL

In our model the manager of a firm is subject to moral hazard, but also has private information about the firms future returns at the beginning of each period. 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 she will have higher expected utility from employment with the firm rather than another one, an incentive compatibility constraint, that induces her to maximize the value of the firm rather than using the resources of the firm to pursue some other objective, and two other conditions that induce the manager to truthfully reveal her private information. After paying the manager for her work in the previous period, 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 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 firms assets are realized at the end of the period. It depends on how well the firm was managed during the period, the private information available to the manager, as well as other unanticipated factors. The objective of the manager is to sequentially maximize her expected lifetime utility, and the goal of the firm is expected value maximization.

More specifically, at the beginning of period t the manager is paid compensation denoted w_t for her work in period $t-1$ according to the schedule the shareholders had previously committed, and her managerial contracts is up for renewal. She makes her consumption choice, a positive real number denoted by c_t, and the board proposes a new contract. At that time 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 she chooses to be engaged inside the firm.

If $l_{t0} = 0$, the prospects of the firm are then fully revealed to the manager but partially hidden to the shareholders. We assume throughout that managers privately observe $s_t \in \{1, 2\}$ in period t, information that affects the distribution of the firm's abnormal returns. The board announces how managerial compensation will be determined as a function of $s'_t \in \{1, 2\}$, what she tells them about the firm's prospects and its subsequent performance, as measured by abnormal returns x_{t+1} revealed at the beginning period $t+1$. The manager truthfully declares or lies about the firm's prospects by announcing $s'_t \in S$, effectively selecting one from many schedules $w\left(s'_t, x_{t+1}\right)$ indexed by her announcement s'_t.

She then makes her unobserved labor effort choice, denoted by $l_{tj} \in \{0, 1\}$ for $j \in \{1, 2\}$ in each period t. There are two possibilities, to work diligently for the firm by pursuing the shareholders objectives of value maximization, and indicated by setting $l_{t2} = 1$, or to be employed by the firm but shirk, following different objectives than maximizing the firm's value, and here denoted by $l_{t1} = 1$. Let $l_t \equiv (l_{t0}, l_{t1}, l_{t2})$. Since leaving the firm, working diligently and shirking are mutually exclusive activities then $\sum_{j=0}^{3} l_{tj} = 1$.

At the beginning of the period $t + 1$ abnormal returns x_{t+1} for the firm are drawn from a probability distribution which depends on the true state s_t and the manager's action l_t. We denote the probability distribution function for abnormal returns in period t when the manager works diligently and the state is s by $F_s\left(x_{t+1}\right)$, and assume it is differentiable with density $f_s\left(x_{t+1}\right)$. Similarly, let $f_s\left(x_{t+1}\right) g_s\left(x_{t+1}\right)$ denote the probability density function for abnormal returns in period t when the manager shirks. Since $f_s(x)\, g_s(x)$ is a density, $g_s(x)$ must be a positive mapping with $E_s\left[g_s(x)\right] = 1$,where the expectation is taken with respect to $f_s(x)$. Compensation to the manager is denoted by $w_{t+1} \equiv w\left(s'_t, x_{t+1}\right)$. We also 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 we assume that $\lim_{x\to\infty}\left[g_s(x)\right] = 0$ for each $s \in \{1, 2\}$.

We assume there are a complete set of markets for all publicly disclosed events and denote by w_{t+1}, the manager's compensation in period t, in units of current consumption. The manager's wealth is endogenously determined by her consumption and compensation. By assuming markets exist for consumption contingent on any public event, we effectively attribute all deviations from the law of one price to the particular market imperfections under consideration.

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:

$$-\sum_{t=0}^{T}\sum_{j=0}^{J}\beta^{t}\alpha_{j}l_{tj}\exp\left(-\rho c_{t}\right) \tag{2.1}$$

where β is the constant subjective discount factor, ρ is the constant absolute level of risk aversion, and α_j is a utility parameters with consumption equivalent $-\rho^{-1}\log\left(\alpha_j\right)$ 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.

2.1. *Feasible Short Term Contracts*

At the end of the next 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 $w\left(s_t', x_{t+1}\right)$, the compensation contract she anticipates receiving from the firm, l_t, her labor supply choices, and s_t', her announcement about the firm's prospects which may depend on the state of the firm s_t, which she observes after making her employment but not her effort decision. Appealing to Myerson (1982), the revelation principle implies that we can without loss of generality restrict the set of feasible contracts to those that respect the participation, incentive compatibility and truth telling constraints we define. The participation constraint states that the manager is indifferent between working one period and then leaving, versus not working for the firm at all. We show this is a necessary and sufficient condition for the worker to prefer managing the firm for a period, regardless of the choices she makes in the future. The incentive compatibility constraint restricts short term contracts to those payment schedules in which the manager prefers to work diligently rather than shirk. The truth telling condition requires shareholders to write contracts that induce the manager to select a compensation schedule that reveals the firm's prospects. Finally the contract must also guard against the possibility of the manager lying about the state and also shirking, which we name the sincerity constraint.

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. To obtain it, let b_t denote the price of a bond that pays of a unit of consumption from period t through to period T, relative to the price of a unit of consumption in period t. For expositional convenience, Lemma 1 states this indirect utility function in terms of the utility she would receive from immediate retirement.

Lemma 2.1 *If the manager anticipating a contract of* $w\left(s_t', x_{t+1}\right)$ *retires in period* t *or period* $t+1$ *by setting:* $\left(1-l_{t0}\right)\left(1-l_{t+1,0}\right)=0$*, she optimally chooses* $\left(l_t, s_t'\right)$ *to* *minimize*

$$(\alpha_0/\alpha_j)^{1/(b_t-1)}\, l_{t0} + E_t\left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right)\left[g_s\left(x_{t+1}\right) l_{t1} + l_{t2}\right]\right] \tag{2.2}$$

Suppressing the bond price for expositional convenience, let $v_{st}(x)$ measure how utility is scaled up by compensation if abnormal returns x are realized at the end of the current period t:

$$v_{s,t+1}(x) \equiv \exp\left(-\frac{\rho w(s,x)}{b_{t+1}}\right) \tag{2.3}$$

To induce an honest, diligent manager to participate, her expected utility from employment must exceed the utility she would obtain from retirement. Setting $(l_{t2}, s_t') = (1, s_t)$ in (2.2) and substituting in $v_s(x_{t+1})$, the participation constraint is thus:

$$\left[\sum_{s=1}^{2}\int_{\underline{x}}^{\infty}\varphi_s v_{st}\left(x_{t+1}\right) f_s\left(x_{t+1}\right) dx\right] \equiv E\left[v_{s,t+1}(x)\right] \leq (\alpha_0/\alpha_j)^{1/(b_t-1)} \tag{2.4}$$

Given her decision to stay with the firm one more period, and to truthfully reveal the state, the incentive compatibility constraint induces the manager to prefer working diligently to shirking. Substituting the definition of $v_s(x)$ into (2.2) 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:

$$0 \leq \int_{\underline{x}}^{\infty}\left(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}\right) v_s(x) f_s(x)\, dx \equiv E_s\left[\left(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}\right) v_s(x)\right] \tag{2.5}$$

Information hidden from shareholders further restricts the set of contracts that can be implemented. We assume throughout that legal considerations induce the manager not to overstate the firm's prospects but that incentives must be provided to persuade the manager from understating them. 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 condition:

$$0 \leq \int\left[v_1(x) - v_2(x)\right] f_2(x)\, dx \equiv E_2\left[v_1(x) - v_2(x)\right] \tag{2.6}$$

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

$$
\begin{aligned}
0 &\leq \int \left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_1(x) g_2(x) - v_2(x)\right] f_2(x)\, dx \\
&\equiv E_2\left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_1(x) g_2(x) - v_2(x)\right] \qquad (2.7)
\end{aligned}
$$

where $(\alpha_1/\alpha_2)^{1/(b_t-1)} v_1(x)$ is proportional to the utility obtained from shirking and announcing the first state, and $f_2(x) g_2(x)$ is the probability density function associated with shirking when the second state occurs.

2.2. *The Optimal Contract*

Having defined the constraints, we are led to formulating the cost minimization problem shareholders solve for each possible effort by state. Recalling from its definition that $\log v_s(x)$ is proportional to $-w_s(x)$, deriving $w(s_t, x_{t+1})$ to minimize expected compensation of inducing diligent work in both states subject to the three constraints is equivalent to choosing $v_s(x_{t+1})$ to maximize:

$$
\sum_{s=1}^{2} \int_{\underline{x}}^{\infty} \varphi_s \log v_s(x_{t+1}) f_s(x_{t+1})\, dx_{t+1} \equiv E[\log v_s(x)] \qquad (2.8)
$$

subject to the same four constraints. To achieve diligent work, shareholders maximize:

$$
\begin{aligned}
&\sum_{s=1}^{2} \varphi_s \int_{\underline{x}}^{\infty} \left\{\log v_s(x) + \eta_0\left[(\alpha_0/\alpha_2)^{1/(b_t-1)} - v_{st}\right]\right\} f_s(x)\, dx \qquad (2.9)\\
&+\sum_{s=1}^{2} \varphi_s \eta_s \int_{\underline{x}}^{\infty} v_s(x)\left[(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}\right] f_s(x)\, dx \\
&+\varphi_2\eta_3 \int_{\underline{x}}^{\infty} [v_1(x) - v_2(x)] f_2(x)\, dx \\
&+\varphi_2\eta_4 \int \left[(\alpha_1/\alpha_2)^{1/(b_t-1)} v_1(x) g_2(x) - v_2(x)\right] f_2(x)\, dx
\end{aligned}
$$

with respect to $v_s(x_{t+1})$, where η_0 through η_4 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 $v_s(x_{t+1})$ for each x_{t+1}. 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:

$$
\begin{aligned}
v_{1t}(x)^{-1} &= \eta_0 + \eta_1\left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x)\right] - \eta_3 h(x) \\
&\quad -\eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x) h(x) \qquad (2.10)\\
v_{2t}(x)^{-1} &= \eta_0 + \eta_2\left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)\right] + \eta_3 + \eta_4
\end{aligned}
$$

where:

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

is the weighted likelihood ratio of abnormal returns for the two states given diligent work. The following lemma is helpful for interpreting the first order conditions.

LEMMA 2.2 *The Lagrange multipliers satisfy:*

1. $\eta_0 = (\alpha_2/\alpha_0)^{1/(b_t-1)}$
2. $\eta_3 + \eta_4 = \{E_2 [v_2 (x)]\}^{-1} - \{E [v_{st} (x)]\}^{-1}$

By inspection, the first equality in Lemma 2.2 demonstrates that (α_2/α_0) only enters through the definition of η_0, the shadow value of participation, and conversely the value of η_0 does not depend on which other constraints are binding. From the second equality we infer that if $\eta_3 = \eta_4 = 0$, then:

$$E_2 [v_2 (x)] = E [v_{st} (x)] = E_1 [v_1 (x)]$$

In words, if neither the truth telling nor the sincerity constraints bind, the pure moral hazard case, then expected utility is equalized across states. Otherwise $(\eta_3 + \eta_4)$ 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:

$$E_2 [v_2 (x)] < E [v_{st} (x)] < E_1 [v_1 (x)]$$

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.

The cost minimizing solution 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 2.3 below this ensures the sincerity constraint does not bind and $\eta_4 = 0$. 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_1, η_2, and η_3. They are:

$$\begin{aligned} & \int_{\underline{x}}^{\infty} \frac{1}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_3 h(x) + \eta_1 \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x)\right]} f_2(x)\, dx \\ (2.11) \qquad = & \int_{\underline{x}}^{\infty} \frac{1}{(\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_3 + \eta_2 \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)\right]} f_2(x)\, dx \end{aligned}$$

$$
\begin{aligned}
0 &= \int_{\underline{x}}^{\infty} \frac{g_1(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_3 h(x) + \eta_1 (\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_1 g_1(x)} f_1(x)\, dx \\
&= \int_{\underline{x}}^{\infty} \frac{g_2(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_3 + \eta_2 (\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_2 g_2(x)} f_2(x)\, dx
\end{aligned} \tag{2.12}
$$

Lemma 2.3 *If $f_1(x) g_1(x) \equiv f_2(x) g_2(x)$, then $\eta_4 = 0$.*

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 maximization problem and the associated first order conditions are defined by setting $\eta_3 = \eta_4 = 0$ in (2.9) and Lemma 2.2. Similarly two of the Kuhn Tucker equations drop out, leaving the solution to η_s uniquely defined by:

$$
\int_{\underline{x}}^{\infty} \frac{\left[(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}\right] f_s(x)}{(\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_s[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_s(x)]} dx = 0 \tag{2.13}
$$

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

$$
\int_{\underline{x}}^{\infty} \left\{ \log v_s(x) + \eta_{0s} \left[(\alpha_0/\alpha_2)^{1/(b_t-1)} - v_{st} \right] + \eta_s v_s(x) \left[(g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)} \right] \right\} f_s(x)\, dx \tag{2.14}
$$

In this case, however, solved in Margiotta and Miller (2000), 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 2.2 it is straightforward to show that $\eta_{01} = \eta_{02}$.

The cost minimizing way of achieving the three other combinations of effort can be derived using the same procedure as we have just explained for the case of diligent work in both states. The wage for employing a manager to shirk in both states is independent of the firm's abnormal return, and yields a inverse utility equivalent of:

$$
v_{1t}(x)^{-1} = (\alpha_1/\alpha_0)^{1/(b_t-1)} \tag{2.15}
$$

that just offsets the value of leaving the firm and consequently does not depend on the state. To induce effort in the first state only the firm pays the equivalent of:

$$
\begin{aligned}
v_{1t}(x)^{-1} &= \textstyle\sum_{s=1}^{2} (1-\varphi_s)(\alpha_s/\alpha_0)^{1/(b_t-1)} + \eta_1 \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x) \right] \\
&\quad - \eta_3 h(x) - \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x) h(x) \\
v_{2t}(x)^{-1} &= \textstyle\sum_{s=1}^{2} (1-\varphi_s)(\alpha_s/\alpha_0)^{1/(b_t-1)} + \eta_3 + \eta_4
\end{aligned} \tag{2.16}
$$

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 $(\eta_3+\eta_4)$ is positive in the hybrid model but zero when there is no private information. To induce effort in the second state only:

$$
\begin{aligned}
v_{1t}(x)^{-1} &= \textstyle\sum_{s=1}^{2} \varphi_s \left(\alpha_s/\alpha_0\right)^{1/(b_t-1)} - \eta_3 h(x) - \eta_4 \left(\alpha_1/\alpha_2\right)^{1/(b_t-1)} g_2\left(x\right) h\left(x\right) \\
v_{2t}(x)^{-1} &= \textstyle\sum_{s=1}^{2} \varphi_s \left(\alpha_s/\alpha_0\right)^{1/(b_t-1)} + \eta_2 \left[\left(\alpha_2/\alpha_1\right)^{1/(b_t-1)} - g_2\left(x\right)\right] + \eta_3 + \eta_4
\end{aligned}
\tag{2.17}
$$

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.

Anticipating his reaction to any proposed compensation package, shareholders demand a pair of effort choices from the manager, that maximize the expected value of the firm, by selecting one of the four cost minimizing contracts. Our analysis demonstrates that the cost minimizing schedule selected for each state depends on the effort level the shareholders wish to induce in both states. Accordingly let $L_s \in \{0,1\}$ indicate a directive to work diligently in state $s \in \{1,2\}$, where $L_s = 0$ means the shareholders anticipate the manager to shirk in state s, and write $w_{st}\left(x, L_1, L_2\right)$ for the contract associated with a directive of $\left(L_1, L_2\right)$, computed using the cost minimization programs described above. Shareholders demand $\left(L_1, L_2\right)$ of managers to maximize:

$$
\textstyle\sum_{s=1}^{2} \int_{\underline{x}}^{\infty} \varphi_s \left\{L_s V x + \left(1-L_s\right) V x g_s\left(x\right) - w_{st}\left(x, L_1, L_2\right)\right\} f_s\left(x\right) dx \tag{2.18}
$$

where V is the value of the firm at the beginning of the period. Optimally choosing $\left(L_1, L_2\right)$ thus completes 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.[1] Lemma 2.4 verifies the assumptions of Fudenberg, Holmstom and Milgrom (1990) are met, thus establishing that the long term optimal contact decentralizes to a sequence of short term contracts satisfying the first order conditions as stated. The four main assumptions are that the firm is assumed to have no better access to financial markets than its manager, the signal about the state the manager receives only applies to the abnormal returns next period, similarly the effort the manager selects has no long term repercussions that are unforeseen by the end of the period, and 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.

LEMMA 2.4 *The optimal long term contract can be implemented by a T period replication of the optimal short term contract.*

[1] Malcomson and Spinnewyn (1988), Fudenberg, Homstrom and Milgom (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 2.4 in the Appendix draws extensively upon their results.

3. IDENTIFICATION

The model is characterized by $f_s(x)$ and $g_s(x)$ for each state $s \in S$, which together define the probability density functions of abnormal returns in the states, the probability distribution for the states, $(\alpha_0, \alpha_1, \alpha_2)$, the preference parameters for leaving the firm, versus shirking and working within the firm, and the risk aversion parameter ρ. For expositional purposes this section assumes that the probability distribution for $s \in S$, that $f_s(x)$ are known for each s, and that $\alpha_0 \equiv 1$. Although the states are partially hidden from shareholders ex-ante, the nature of the optimal contract reveals the states ex-post, explaining why we assume s_t is observed. 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. This is why we now focus on identifying the two mappings $g_s(x)$ plus the constants α_1, α_2 and ρ. There are eight permutations of the pure and hybrid models to consider, depending the unobserved value of (L_1^o, L_2^o). We thoroughly analyze the two permutations that apply to our empirical application, $(L_1^o, L_2^o) = (1, 1)$ for the pure and hybrid models, in which diligent effort is called forth in both states, but also briefly discuss the other permutations at the end of this section.

To facilitate the discussion we partition the parameter space into ρ, 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 (ρ^*, θ^*) the generalized model of moral hazard generating the data process (x_t, s_t, w_t), where $\theta^* \equiv (\alpha_1^*, \alpha_1^*, g_1^*(x), g_2^*(x))$. Following Chesher (2007) we say the structural parameter $(\rho, \theta) \in R^+ \times \Theta$ with true value (ρ^*, θ^*) is identified if (ρ^*, θ^*) can be written as a functional of the conditional distribution of the observed variables. 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 the conditional distribution of the observed variables cannot be rationalized by any $(\rho, \theta) \in R^+ \times \Theta$ the class of models is rejected.

First we prove that if ρ^* is known, then θ^* is identified from the compensation schedule alone. Then, treating the pure and hybrid models separately, we write down conditions that arise from the shape of the optimal contract which define the set to which ρ^* belongs to. We prove this set is sharp and tight: Every element inside the set is observationally equivalent, and every element outside the set is not. Finally our characterization provides a basis for testing the specification of the model. Within a broad class of regular data generating processes for states, abnormal returns and compensation defined below, we show that if the set is empty, then the data are inconsistent with our theoretical framework.

3.1. *Assuming Risk Preferences are Known*

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

PROPOSITION 3.1 *For each $(s, t,)$, let:*

$$\begin{aligned} v_{st}(x,\rho) &\equiv \exp\left[-\rho w_s(x)/b_{t+1}\right] \\ \overline{v}_{st}(\rho) &\equiv \exp\left[-\rho \overline{w}_s/b_{t+1}\right]. \end{aligned}$$

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

$$\begin{aligned} \alpha_{2t}(\rho) &\equiv \left\{E\left[v_{st}(x,\rho)\right]\right\}^{1-b_t} \\ \eta_{2t}(\rho) &\equiv \overline{v}_{2t}(\rho)^{-1} - E_2\left[v_{2t}(x,\rho)^{-1}\right] \\ g_{2t}(x,\rho) &\equiv \eta_{2t}(\rho)^{-1}\left[\overline{v}_{2t}(\rho)^{-1} - v_{2t}(x,\rho)^{-1}\right] \\ \alpha_{1t}(\rho) &\equiv \alpha_{2t}(\rho)\left\{\frac{\left[\overline{v}_{2t}(\rho)\right]^{-1} - \left\{E_2\left[v_{2t}(x,\rho)\right]\right\}^{-1}}{\left[\overline{v}_{2t}(\rho)\right]^{-1} - E_2\left[v_{2t}(x,\rho)^{-1}\right]}\right\}^{1-b_t} \\ \eta_{4t}(\rho) &= \frac{\left\{E\left[v_{st}(x,\rho)\right]\right\}^{-1} E_1\left[v_{1t}(x,\rho)\right] - 1}{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1\left[v_{1t}(x,\rho)\, g_{2t}(x,\rho)\, h(x)\right] - E_1\left[v_{1t}(x,\rho)\, h(x)\right]} \\ &\quad - \frac{E_1\left[v_{1t}(x,\rho)\, h(x)\right]\left[\left\{E_2\left[v_{2t}(x,\rho)\right]\right\}^{-1} - \left\{E\left[v_{st}(x,\rho)\right]\right\}^{-1}\right]}{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1\left[v_{1t}(x,\rho)\, g_{2t}(x,\rho)\, h(x)\right] - E_1\left[v_{1t}(x,\rho)\, h(x)\right]} \\ \eta_{3t}(\rho) &\equiv \left\{E_2\left[v_{2t}(x,\rho)\right]\right\}^{-1} - \eta_{4t}(\rho) - \left\{E\left[v_{st}(x,\rho)\right]\right\}^{-1} \\ \eta_{1t}(\rho) &= \left[\alpha_{1t}(\rho)/\alpha_{2t}(\rho)\right]^{1/(b_t-1)}\left[\overline{v}_{1t}(\rho)^{-1} - \left\{E\left[v_{st}(x,\rho)\right]\right\}^{-1} + \eta_{3t}(\rho)\,\overline{h}\right] \\ g_{1t}(x,\rho) &\equiv \eta_{1t}(\rho)^{-1}\left\{\overline{v}_{1t}(\rho)^{-1} - v_{1t}(x,\rho)^{-1} + \eta_{3t}(\rho)\left[\overline{h} - h(x)\right]\right. \\ &\quad \left. -\eta_{4t}(\rho)\left[\alpha_{1t}(\rho)/\alpha_{2t}(\rho)\right]^{1/(b_t-1)} g_{2t}(x,\rho)\, h(x)\right\} \end{aligned}$$

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

$$\eta_{st}(\rho) \equiv \overline{v}_{st}(\rho)^{-1} - E_s\left[v_{st}(x,\rho)^{-1}\right] \tag{3.1}$$

$$g_{st}(x,\rho) \equiv \eta_{st}(\rho)^{-1}\left[\overline{v}_{st}(\rho)^{-1} - v_{st}(x,\rho)^{-1}\right] \tag{3.2}$$

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 ρ, we briefly describe the intuition to interpret the formulas. The certainty equivalent of a lottery that yields $w(x)$ for a person with absolute risk aversion parameter ρ who then optimally spends it over his lifetime is $\left\{E\left[v_{st}(x,\rho)\right]\right\}^{b_t-1}$. In the optimal contract for our model the manager is indifferent between accepting the job which provides nonpecuniary benefits of $\alpha_{2t}(\rho)$ plus that certainty equivalent versus an option with benefits normalized to zero. This explains the formula for $\alpha_{2t}(\rho)$, the preference parameter for working diligently. The ratio of $\alpha_{1t}(\rho)$, the shirking parameter, to $\alpha_{2t}(\rho)$, depend on the difference between $\{E_2[v_{2t}(x,\rho)]\}^{-1}$ and $E_2\left[v_{2t}(x,\rho)^{-1}\right]$. By definition $b_t > 1$ and the inverse function $1/v$ is convex decreasing in v, so Jensen's inequality implies $\alpha_{1t}(\rho) < \alpha_{2t}(\rho)$. In the special case where a fixed wage is paid $\alpha_{1t}(\rho) \equiv \alpha_{2t}(\rho)$, and the more dispersion there is in the random variable $v_{2t} \equiv v_{2t}(x,\rho^*)$ induced by abnormal returns x and the compensation schedule $w(x)$, the more pronounced the inequality. To interpret $g_{2t}(x,\rho)$ and $g_{2t}(x,\rho)$, the representations of the shirking to diligence likelihood ratios in the two states, we remark that $\overline{v}_{2t}(\rho)^{-1} = \exp[\rho\overline{w}_s/b_{t+1}]$ is the exponentially scaled value of the maximal compensation in the second state, which occurs as $x \to \infty$ and $g_{2t}(x,\rho) \to 0$. At values of x where $g_{2t}(x,\rho) > 0$, 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 ρ, up to the shadow value or opportunity costs of marginally relaxing the incentive compatibility constraint, given by $\eta_{2t}(\rho)$. The likelihood ratio in the first state, $g_{1t}(x,\rho)$, 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.

3.2. *Admissible Risk Aversion*

Proposition 3.1 implies the set of identified parameters can be indexed by a Borel set of values for the risk aversion parameter $\rho \in R^+$ that are observationally equivalent. Next, we analyze the restrictions derived from optimizing behavior in the model that limit the admissible values of ρ. 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 of restrictions arises from the optimality conditions determining the compensation schedule for diligent work are a second source for culling values of ρ that do not belong to Γ_i. The other source is due to differences in firm value from shareholders inducing 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.

For models of pure moral hazard, the optimal contract of Section 2 implies that competitive selection constraints of the form $\Psi_{1t}(\rho) = \Psi_{2t}(\rho) = 0$ hold for all t, where $\Psi_{1t}(\rho)$ and $\Psi_{2t}(\rho)$ are respectively defined by:

$$\Psi_{st}(\rho) \equiv \{E_1[v_{11}(x,\rho)]\}^{1-b_1} - \{E_s[v_{st}(x,\rho)]\}^{1-b_t} \tag{3.3}$$

for $s \in \{1,2\}$. By symmetry, the identification of the shirking parameter, $\alpha_{1t}(\rho)$, as a function of ρ in Proposition 3.1 can be established using the first instead of the second state. This directly leads to a further restriction on ρ that:

$$\Psi_{3t}(\rho) \equiv \frac{\overline{v}_{1t}(\rho)^{-1} - \{E_1[v_{1t}(x,\rho)]\}^{-1}}{\overline{v}_{1t}(\rho)^{-1} - E_1\left[v_{1t}(x,\rho)^{-1}\right]} - \frac{\overline{v}_{2t}(\rho)^{-1} - \{E_2[v_{2t}(x,\rho)]\}^{-1}}{\overline{v}_{2t}(\rho)^{-1} - E_2\left[v_{2t}(x,\rho)^{-1}\right]} = 0 \tag{3.4}$$

Next, consider the ramifications for identification when shareholders induce diligent effort rather than shirking in at least one of the states. Given the parameterization $(\rho, \theta_t(\rho))$ for any $\rho \in R^+$, we denote by $w_{st}(x, \rho, L_1, L_2)$ the cost minimizing contracts to induce (L_1, L_2). Thus the compensation data comprise $w_{st}(x) \equiv w_{st}(x, \rho^*, 1, 1)$. It is easy to show that if the manager is induced to shirk in state s, then the optimal compensation is:

$$w_{1t}(x,\rho,L_1,0) = w_{2t}(x,\rho,0,L_2) = b_{t+1}\log[\alpha_1(\rho)]/\rho(b_t-1) \tag{3.5}$$

where $\alpha_1(\rho)$ is defined in Proposition 3.1. Moreover optimal compensation for diligence does not depend on the effort level induced in the other state. Thus $w_{1t}(x,\rho,1,0) = w_{st}(x,\rho,1,1)$ for $s=1$, and similarly $w_{2t}(x,\rho,0,1) = w_{st}(x,\rho,1,1)$ for $s=2$. By definition the expected value of abnormal returns, conditional on the state s, is zero in the pure moral hazard model, meaning $V_{st}E_s[x] = E_s[w_{st}(x)]$. Since it is less profitable to induce shirking rather than diligence, we conclude $\Lambda_{st}(\rho^*) \geq 0$ where:

$$\Lambda_{st}(\rho) = E_s[V_{st}xg_{st}(x,\rho)] - b_{t+1}\log[\alpha_{1t}(\rho)]/\rho(b_t-1) \tag{3.6}$$

To recapitulate, competitive selection applies to each state taken separately, the taste parameter does not vary across states, and it is nonoptimal for shareholders to induce shirking in either state. This yields five inequalities in ρ. Defining the set:

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

our discussion implies $\rho^* \in \overline{\Gamma}_1$ in pure moral hazard models.

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

$$\begin{aligned}\Psi_{4t}(\rho) \equiv\ & \eta_{1t}(\rho) - [\overline{v}_{1t}(\rho)]^{-1} - \eta_{3t}(\rho)\overline{h} + E_1\left[v_{1t}(x,\rho)^{-1}\right] \\ & +\eta_{3t}(\rho)\varphi_2/\varphi_1 + \eta_{4t}(\rho)(\alpha_{1t}(\rho)/\alpha_{2t}(\rho))^{1/(b_t-1)}E_1[g_{2t}(x,\rho)h(x)]\end{aligned} \tag{3.7}$$

In place of the competitive selection equations in pure moral hazard models, the truth telling and sincerity constraints in the hybrid moral hazard framework yields an equation in ρ. Defining $\Psi_{5t}(\rho)$ and $\Psi_{6t}(\rho)$ as :

$$\Psi_{5t}(\rho) \equiv E_2[v_{2t}(x,\rho) - v_{1t}(x,\rho)] \tag{3.8}$$

$$\Psi_{6t}(\rho) \equiv E_2\left[v_{1t}(x,\rho)\frac{[\overline{v}_{2t}(\rho)]^{-1} - [v_{2t}(x,\rho)]^{-1}}{[\overline{v}_{2t}(\rho)]^{-1} - E_2[v_{2t}(x,\rho)]^{-1}} - v_{2t}(x,\rho)\right] \tag{3.9}$$

it follows that $\Psi_{5t}(\rho)\Psi_{6t}(\rho) = 0$ for all t and $\rho \in \Gamma_2$. That is the truth telling and/or the sincerity constraint must hold in the hybrid model. In any event $\Psi_{5t}(\rho) \geq 0$ and $\Psi_{6t}(\rho) \geq 0$. We also require the multipliers associated with truth telling and sincerity, respectively denoted by $\Psi_{7t}(\rho) \equiv \eta_{3t}(\rho)$ and $\Psi_{8t}(\rho) \equiv \eta_{4t}(\rho)$, to be positive. Also both sets of complementary slackness conditions must be satisfied, meaning $\Psi_{5t}(\rho)\Psi_{7t}(\rho) = 0$ and $\Psi_{6t}(\rho)\Psi_{8t}(\rho) = 0$.

Finally since $g_1(x)$ is a likelihood ratio in the hybrid model we impose the restriction that:

$$\Psi_{9t}(\rho) \equiv E_1\left[1\left\{g_{1t}(x,\rho)\right\} - 1\right] \geq 0 \tag{3.10}$$

to ensure $g_{1t}(x,\rho) \geq 0$ with unit mass.

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 (2.15). Since this is not optimal:

$$\Lambda_{3t}(\rho) = -E\left[V_{st}xg_{st}(x,\rho)\right] + b_{t+1}\log\left[\alpha_{1t}(\rho)\right]/\rho\left(b_t - 1\right) \tag{3.11}$$

is positive at ρ^*. The two remaining inequalities also require us to solve the cost minimizing compensation schedule when the nonoptimal choices $(L_1, L_2) = (0,1)$ or $(L_1, L_2) = (1,0)$ are chosen for the parameterization $(\rho, \theta_t(\rho))$. Noting the optimization problem determining the compensation plan as a function of (ρ, L_1, L_2), is globally concave and satisfies the Kuhn Tucker conditions, it follows that the solution is determined by the remaining three multipliers $\eta_1(L_1, L_2, \rho)$ through $\eta_4(L_1, L_2, \rho)$ and the first order conditions:

$$\begin{aligned} w_{1t}(x,\rho,L_1,L_2) &= -\frac{b_{t+1}}{\rho}\log\left\{\left[\alpha_{2t}(\rho)\right]^{1/(b_t-1)} + \eta_1(L_1,L_2,\rho)\left[\frac{\alpha_{2t}(\rho)}{\alpha_{1t}(\rho)}\right]^{1/(b_t-1)}\right. \\ &\quad -\eta_1(L_1,L_2,\rho)\,g_{1t}(x,\rho) - \eta_3(L_1,L_2,\rho)\,h(x) \\ &\quad \left. -\eta_4(L_1,L_2,\rho)\left[\frac{\alpha_{1t}(\rho)}{\alpha_{2t}(\rho)}\right]^{1/(b_t-1)} g_{2t}(x,\rho)\,h(x)\right\} \end{aligned} \tag{3.12}$$

$$\begin{aligned} w_{2t}(x,\rho,L_1,L_2) &= -\frac{b_{t+1}}{\rho}\log\left\{\left[\alpha_{2t}(\rho)\right]^{1/(b_t-1)} + \eta_2(L_1,L_2,\rho)\left[\frac{\alpha_{2t}(\rho)}{\alpha_{1t}(\rho)}\right]^{1/(b_t-1)}\right. \\ &\quad \left. -\eta_2(L_1,L_2,\rho)\,g_{2t}(x,\rho) + \eta_{32}(L_1,L_2,\rho) + \eta_{42}(L_1,L_2,\rho)\right\} \end{aligned} \tag{3.13}$$

Substituting Equations (3.12) and (3.13) into the incentive compatibility, truth telling and sincerity constraints defined for the parameterization $(\rho, \theta_t(\rho))$, we solve the four equations in four unknowns for a given value of ρ and $(L_1, L_2) = (0,1)$ or $(L_1, L_2) = (1,0)$. The Kuhn Tucker theorem ensures the positive solution to $\eta_1(L_1, L_2, \rho)$ through $\eta_4(L_1, L_2, \rho)$ is unique. Substituting the respective solutions into the compensation equations defined above we obtain the two restrictions $\Lambda_{4t}(\rho^*) \geq 0$ and $\Lambda_{5t}(\rho^*) \geq 0$:

$$\Lambda_{4t}(\rho) = -E_2\left[V_{2t}xg_{2t}(x,\rho) - w_{2t}(x,\rho,1,0)\right] - E_1\left[V_{1t}x - w_{1t}(x,\rho,1,0)\right] \tag{3.14}$$

$$\Lambda_{5t}(\rho) = -E_1\left[V_{1t}xg_{1t}(x,\rho) - w_{2t}(x,\rho,0,1)\right] - E_2\left[V_{2t}x - w_{1t}(x,\rho,0,1)\right] \tag{3.15}$$

Summarizing the restrictions directly applied to the hybrid model, the shape of the compensation schedule implies ρ^* 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 $g_{1t}(x,\rho)$ be nonnegative for all x. Defining the set of risk aversion parameters as:

$$\begin{aligned}\overline{\Gamma}_2 \equiv \{\rho > 0 : \ & \Lambda_{kt}(\rho) \geq 0 \text{ for } k \in \{3,4,5\} \text{ and } \Psi_{jt}(\rho) \geq 0 \text{ for } j \in \{1,2,5,6,7,8,9\} \\ & \text{and } \Psi_{4t}(\rho) = \Psi_{5t}(\rho)\Psi_{6t}(\rho) = \Psi_{5t}(\rho)\Psi_{7t}(\rho) = \Psi_{6t}(\rho)\Psi_{8t}(\rho) = 0 \text{ and all } t\}\end{aligned}$$

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

PROPOSITION 3.2 $\Gamma_i \subseteq \overline{\Gamma}_i$ *for* $i \in \{1,2\}$.

Are there any other restrictions on the sets of admissible parameters to further shrink $\overline{\Gamma}_i$ for $i \in \{1,2\}$ when $\overline{\Gamma}_i$ 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 $\rho > 0$. 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 $\overline{\Gamma}_i$.

PROPOSITION 3.3 *Supposing* $\rho \in \overline{\Gamma}_i$ *for* $i = 1$ *or* $i = 2$, *then for all* $t \in \{1,2,\ldots\}$ *and* $s \in \{1,2\}$

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

3.3. *Regular Data Generating Processes*

It remains to show that if $\rho \in \overline{\Gamma}_i$ for $i \in \{1,2\}$, then ρ is observationally equivalent to ρ^*. We prove a more general result by considering the class of regular data generating processes, generically denoted by $\{p(s\,|s_{t-1}, x_{t-1}), f_s(x), w_s(x)\}_{s=1}^{S}$ for $(s,x) = (s_t, x_t)$, where s_t is sequentially generated by $p(s\,|s_{t-1}, x_{t-1})$, a Markov probability transition, x_t is an independently distributed random variable with conditional density $f_s(x)$, and realized compensation can be expressed as the known mapping $w_t = w_s(x)$ where $w_s^*(x)$ is defined on the space of states and abnormal returns. For expositional convenience we maintain the assumptions that the probability distribution for $s \in S$, that $f_s(x)$ are known for each s. We also impose the regularity condition that:

$$\lim_{x \to \infty} [w_s(x)] = \sup_{x \in R} [w_s(x)] \equiv \overline{w}_s \tag{3.16}$$

but discard the premise that the data was necessarily generated by a generalized model of moral hazard, and ignore intertemporal variation from the bond price b_t by assuming the data is strictly cross sectional taken at a single point in time t. In this broader context we entertain four possibilities, tested in the next 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 final proposition states that any regular generating process is observationally equivalent to a generalized moral hazard model if there is a positive real number γ that obeys the restrictions described above. Consequently $\Gamma_i = \overline{\Gamma}_i$ for $i \in \{1, 2\}$ and the bounds we have constructed are tight.

Proposition 3.4 *A regular data generating process is observationally equivalent to a pure moral hazard model indexed by $\gamma > 0$ if $\gamma \in \overline{\Gamma}_1$, and is observationally equivalent to a hybrid moral hazard model indexed by $\gamma > 0$ if $\gamma \in \overline{\Gamma}_2$.*

3.4. *Other Permutations*

Having characterized identification for the hybrid and pure moral hazard models when $(L_1^o, L_2^o) = (1, 1)$, we now briefly consider the six other permutations, formed from three pairs of $(L_1^o, L_2^o) \neq (1, 1)$. In three of them, the manager is compensated independently of abnormal returns in one state, in those states only the distribution for abnormal returns conditional on shirking is identified, and the methods developed for $(L_1^o, L_2^o) = (1, 1)$ can be adapted, simplifying identification of the remaining parameters. 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, thus identifying the realized value of (L_1^o, L_2^o). The data we analyze below strongly reject the null hypothesis of no effect in at least one state, explaining why we analyzed the most complicated case $(L_1^o, L_2^o) = (1, 1)$.

When $(L_1^o, L_2^o) = (0, 0)$ the wage outcomes of the pure and hybrid model are identical, and noninformative about the scope of moral hazard. The hybrid model with $(L_1^o, L_2^o) = (1, 0)$ is more parsimonious than the $(L_1^o, L_2^o) = (1, 1)$, 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. But upon algebraically manipulating the first order conditions to substitute out $g_2(x)$ we obtain:

$$v_{1t}(x)^{-1} = \Upsilon_{0t} + \Upsilon_{1t} h(x)\, v_{2t}(x)^{-1} - \Upsilon_{2t} h(x) \tag{3.17}$$

for the $(L_1^o, L_2^o) = (1, 0)$ permutation, where:

$$\begin{aligned} \Upsilon_{0t} &\equiv \textstyle\sum_{s=1}^{2} \varphi_s (\alpha_s)^{1/(b_t-1)} \\ \Upsilon_{1t} &\equiv \eta_2^{-1}\eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} \\ \Upsilon_{2t} &\equiv \eta_3 + \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} \Big[\eta_2^{-1} \textstyle\sum_{s=1}^{2} \varphi_s (\alpha_s)^{1/(b_t-1)} + (\alpha_2/\alpha_1)^{1/(b_t-1)} \\ &\qquad + \eta_2^{-1}\eta_3 + \eta_2^{-1}\eta_4\Big] \end{aligned} \tag{3.18}$$

The bond rate b_t is observed, the state probabilities (φ_1, φ_2) are identified from the series on s_n, and $h(x)$ is identified from data on (s_n, x_n). Hence the vector function $\Upsilon_t \equiv (\Upsilon_{0t}, \Upsilon_{1t}, \Upsilon_{2t})$ is essentially a mapping from the five dimensional parameter vector $(\eta_2, \eta_3, \eta_4, \alpha_1, \alpha_2,)$, while ρ 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 (3.17) holding exactly for all x.

For similar reasons we did not place restrictions on the primitives that render the sincerity constraint redundant making 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 2.3 that collectively ensure $\eta_4 = 0$. If that restriction is imposed on the data then we obtain from formulas for $g_1(x)$ and $g_2(x)$ given in Proposition 3.1:

$$\frac{f_1(x)}{f_2(x)} \equiv \frac{g_2(x)}{g_1(x)} = \frac{\eta_1}{\eta_2}\left[\frac{\overline{v}_t^{-1} - v_{2t}(x)^{-1}}{\overline{v}_t^{-1} - v_{1t}(x)^{-1} + \eta_3 h(\overline{x}) - \eta_3 h(x)}\right] \tag{3.19}$$

Noting $f_s(x)$ and $h(x)$ are identified from data on abnormal returns and their states, along with the quantities $\overline{w}$ and $h(\overline{x})$, this permutation places very strong restrictions on admissible values of ρ, η_1/η_2 and η_3, because Equation (3.19) 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 $(L_1, L_2) = (1, 1)$, imposing the regularity condition, that the the state does not affect the shirking distribution, has powerful implications from the perspective of inference which are likely to be rejected in empirical studies based on large data sets.

4. EMPIRICAL APPLICATION

The remainder of this paper applies the preceding analysis to the compensation of chief executive officers of publicly traded firms. From an empirical standpoint, trading by corporate insiders appears to be profitable. Seyhun (1986) finds that insiders tend to buy before an abnormal rise in stock prices and sell before an abnormal decline. Earlier studies by Lorie and Niederhoffer (1968), Jaffe (1974), and Finnerty (1976) draw similar conclusions. More recently, Seyhun (1992a) presents evidence showing that insider trading volume, frequency, and profitability all increases significantly during the 1980s. Over the decade, he documents that insiders earned over 5 percent abnormal returns on average. Seyhun (1992b) determines that insider trades predict up to 60 percent of the total variation in one-year-ahead returns. Drawing upon a 9 year panel beginning 1992, in Gayle and Miller (2008a) we showed that a manager's adjustments in the financial securities of his own firm's is a significant explanatory variable for predicting the abnormal returns of the firm next period. We also constructed a simple self-financing dynamic portfolio strategy based on changes in asset holdings by managers which significantly outperforms the market portfolio, realizing over 90 percent of the gains that could have been achieved with perfect foresight.

Our model shows that in the optimal contract managers are rewarded for truthfully reporting the prospects of the firm impounded in the state variables driving future profitability. However, as proved in Gayle and Miller (2008a), if there is only private information but no hidden actions, then the optimal contract is to pay the manager a fixed wage. Since existing institutional arrangements in the U.S. prevent managers from trading anonymously in their firm's shares, it is relatively straightforward for a compensation board representing shareholders to retrospectively neutralize, and indeed, penalize managers who attempt to benefit from insider trading. Taken together, the theoretical implications of our model, the mandatory reporting of trading by managers, and the evidence reviewed above, suggest there is both private information and hidden actions. Thus the purpose of our empirical investigation is to test the specification of the generalized moral hazard framework, investigate whether the pure moral hazard model can be rejected in favor of the hybrid model, and quantify the importance of moral hazard versus private information in optimal contract design, by applying the results derived in the previous sections.

The main data for the empirical portion of our study was compiled from 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.

4.1. *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, her 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 her 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 she holds. More specifically, let v_{nt} denote the value of firm n at time t on the stock market, and let $\widetilde{x}_{nt}$, 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:

$$x_{nt} \equiv \widetilde{x}_{nt} + \frac{w_{nt}}{V_{n,t-1}} \tag{4.1}$$

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

$$\widetilde{w}_{nt} = w_{nt} + \varepsilon_{nt} \tag{4.2}$$

Although $(\widetilde{w}_{nt}, \widetilde{x}_{nt})$ rather than (w_{nt}, x_{nt}) is observed for each (n, t), we can, however, construct consistent estimates of (w_{nt}, x_{nt}) from $(\widetilde{w}_{nt}, \widetilde{x}_{nt})$ 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 4.1 *If* $V(x_2 - x_1) \neq w(x_2) - w(x_1)$ *for all* $(x_1, x_2) \in R^2$*, then:*

$$w_{nt} = E[\widetilde{w}_{nt} | \widetilde{x}_{nt}, z_{nt}, s_{nt}, b_t, V_{n,t-1}] \tag{4.3}$$

This lemma implies that compensation schedule is the conditional expectation of measured compensation given net abnormal returns and lagged firm size. In our application we assumed that Z is a finite set, and in the optimal contract the manager also reveals the state s_{nt} which we assume econometricians observe retrospectively. Consequently pointwise consistent estimates of compensation w_{nt} can be obtained for each observation with Kernel estimators of the cross section taking the form:

$$w_{nt}^{(N)} = \frac{\sum_{m=1,m\neq n}^{N} w_{mt} I\{z_{mt} = z_{nt}, s_{mt} = s_{nt}\} K\left(\frac{x_{mt}-x_{nt}}{\delta_{xN}}, \frac{v_{m,t-1}-v_{n,t-1}}{\delta_{vN}}\right)}{\sum_{n=1,m\neq n}^{N} I\{z_{mt} = z_{nt}, s_{mt} = s_{nt}\} K\left(\frac{x_{mt}-x_{nt}}{\delta_{xN}}, \frac{v_{m,t-1}-v_{n,t-1}}{\delta_{vN}}\right)} \tag{4.4}$$

where $K(\cdot)$ is a bivariate probability density function with full support and $\delta_N \equiv (\delta_{xN}, \delta_{vN})$ is the bandwidth satisfying the convergence property $\delta_N \to 0$ as $N \to \infty$. Similarly a consistent estimator of the gross abnormal return is:

$$x_{nt}^{(N)} \equiv \widetilde{x}_{nt} + \frac{w_{nt}^{(N)}}{v_{n,t-1}} \tag{4.5}$$

and an estimate of the density $f_z(x)$ at x is:

$$f_{z,s}^{(NT)}(x) = \frac{\sum_{t=1}^{T} \sum_{n=1}^{N} I\{z_{nt} = z, s_{nt} = z\} K_x\left(\frac{x_{nt}^{(NT)}-x}{\delta_{x,NT}}\right)}{\delta_{x,NT} \sum_{t=1}^{T} \sum_{n=1}^{N} I\{z_{nt} = z, s_{nt} = z\}} \tag{4.6}$$

where $K_x(\cdot)$ is a univariate probability density function with full support and the bandwidth $\delta_{x,NT} \to 0$ as $NT \to \infty$.

4.2. *Data Summary*

Tables 1 and 2 respectively summarize the cross sectional and longitudinal features of our data. Table 1 shows there are 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.[2] 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.

We also assume that managers release information about the state of the firm through accounting statements, and exercise considerable discretion over their determination. There are many ways for managers to directly affect the firm's comprehensive income, defined as the difference between the change in assets and the changes in liabilities plus dividends. For example they can adjust the level of what is called over balanced sheet financing, choose among valuation methods for assets and liabilities, use discretionary timing when writing off nonperforming assets. Exercising such liberties provides a mechanism for managers to signal the state of the firm to shareholders. We adapt a commonly used accounting measure of the manager's accomplishments and firm's success called comprehensive income. Rather than use the definition of comprehensive income to directly measure changes in this state, we normalize for one measure of firm size, equity holdings, and define accounting return r_{nt} as:

$$r_{nt} = \frac{Assets_{nt} - Debt_{nt} + Dividend_{nt}}{Assets_{n,t-1} - Debt_{n,t-1}} \tag{4.7}$$

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.

Table 1 shows 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.[3] 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

[2]Findings of several studies, including our own, show this is indeed the case.

[3]Paranthetically 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.

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 1 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 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 1 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 3 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.

For convenience a finite partition was used to differentiate between firms each period when constructing the compensation and returns data $\left(w_{nt}^{(N)}, x_{nt}^{(N)}\right)$. We categorized each firm in each sector per year as small or large depending on whether they are below or above the median firm in assets, employees and debt to equity ratio. Let $J_n \in \{1, 2, 3\}$ denote the sector to which the n^{th} firm belongs, let $A_{n,t-1} \in \{S, L\}$ indicate whether the total assets of firm n lie above or below the median assets for its sector at the beginning of period t, let $W_{n,t-1} \in \{S, L\}$ indicate whether the number of employees (workers) at the firm is above or below the median assets, and let $D_{n,t-1} \in \{S, L\}$ indicate whether the debt to equity in firm n at the beginning of period t is above or below the median debt to equity. 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 between them on z_{nt} without resorting to constraints that induce truth telling.

Similarly we define the manger's announcement about the hidden state $s \in \{1, 2\}$ as an indicator variable, telling whether the firm's accounting return is higher or lower than the average for all firms with the same publicly observed state s_{nt} in that period

t :

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

As explained in the previous section, we would not expect s_{nt} to be meaningful unless the truth telling and sincerity constraints were satisfied by the contract, conditions we impose in testing and estimation.

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 4 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.

Table 4 displays the numbers in each cell (z_{nt}, s_{nt}). For the most part, the probability of being in the bad state is higher, implying the median of r_{nt} is less than its mean. However there are exceptions, such as $(A, W, D) = (S, S, L)$ in the primary and consumer sectors. Table 3 provides a cross sectional summary of the average abnormal returns and CEO compensation conditional on the publicly observable z_{nt} and the accounting report s_{nt} 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, the sample deviations are between recording applying the numbers of observations in each observed state from the first column in Table 4, and noting the independence of the observations that are used to form the sample means, 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. 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.

5. TESTING MORAL HAZARD AND HIDDEN INFORMATION

Because Γ_i is defined by a vector function of equalities and inequalities in ρ, the framework is amenable to testing whether a regular data generating process comes from generalized model of moral hazard or not. To mitigate the risk of spurious rejection, we develop tests that accommodate differences in contracts attributable to sources of retrospectively observed heterogeneity other than the potentially unobserved states we have been analyzing $s \in \{1, 2\}$. We assume the data contain information on a finite set of economic and financial characteristics $z \in \{1, \dots, Z\}$, and to allow for exogenous background influences z in estimation and testing, we now write $\Psi_{jt}(\rho, z)$ for $\Psi_{jt}(\rho)$ and $\Lambda_{kt}(\rho, z)$ for $\Lambda_{kt}(z)$. Armed with this expanded notation, the sets Γ_i can be expressed as $\Gamma_i = \{\rho > 0 : Q_i(\rho) = 0\}$ for $i \in \{1, 3\}$ where:

$$
\begin{aligned}
Q_1(\rho) &\equiv \sum_{t=1}^{T}\sum_{z=1}^{Z}\left[\sum_{j=1}^{3}\Psi_{jt}(\rho, z)^2 + \sum_{k=1}^{2}\Lambda_{kt}(\rho, z)^2\right] \\
Q_2(\rho) &\equiv \sum_{t=1}^{T}\sum_{z=1}^{Z}\left\{\sum_{j=1,2,5,6,\dots,9}\min\left[0, \Psi_{jt}(\rho, z)\right]^2 + \sum_{j=6,7}\left[\Psi_{5t}(\rho, z)\,\Psi_{jt}(\rho, z)\right]^2\right. \\
&\qquad \left. +\Psi_{4t}(\rho, z)^2 + \left[\Psi_{6t}(\rho, z)\,\Psi_{8t}(\rho, z)\right]^2 + \sum_{k=3,4,5}\Lambda_{kt}(\rho, z)^2\right\}
\end{aligned}
\tag{5.1}
$$

Appealing to Proposition 3.4, 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 $i \in \{1, 2\}$:

$$
\begin{aligned}
H_0^{(i)} &: Q_i(\rho) = 0 \text{ for some } \rho > 0 \\
H_A^{(i)} &: Q_i(\rho) > 0 \text{ for all } \rho > 0
\end{aligned}
$$

Empirically, estimation and testing in this paper is based on observing N firms over T periods, with generic observation denoted by (w_{nt}, z_{nt}, b_t). The asymptotic properties described here are for large N, but can easily be extended to handle large NT. To test the null hypothesis we define nonparametric estimators of $\Psi_{jt}(\rho, z)$ and $\Lambda_{kt}(\rho, z)$, respectively denoted by $\Psi_{jt}^{(N)}(\rho, z)$ and $\Lambda_{kt}^{(N)}(\rho, z)$, to form empirical analogues of $Q_i(\rho)$, denoted by:

$$
\begin{aligned}
Q_1^{(N)}(\rho) &\equiv \sum_{t=1}^{T}\sum_{z=1}^{Z}\left[\sum_{j=1}^{3}\Psi_{jt}^{(N)}(\rho, z)^2 + \sum_{k=1}^{2}\Lambda_{kt}^{(N)}(\rho, z)^2\right] \\
Q_2^{(N)}(\rho) &\equiv \sum_{t=1}^{T}\sum_{z=1}^{Z}\left\{\sum_{j=1,2,5,6,\dots,9}\min\left[0, \Psi_{jt}^{(N)}(\rho, z)\right]^2 + \sum_{j=6,7}\left[\Psi_{5t}^{(N)}(\rho, z)\,\Psi_{jt}^{(N)}(\rho, z)\right]^2\right. \\
&\qquad \left. +\Psi_{4t}^{(N)}(\rho, z)^2 + \left[\Psi_{6t}^{(N)}(\rho, z)\,\Psi_{8t}^{(N)}(\rho, z)\right]^2 + \sum_{k=3,4,5}\Lambda_{kt}^{(N)}(\rho, z)^2\right\}
\end{aligned}
\tag{5.2}
$$

and a confidence region for Γ_i, defined as $\Gamma_i^{(N)} \equiv \{\rho > 0 : A_N Q_i^{(N)}(\rho) \leq c_\delta\}$, where A_N is the asymptotic rate of convergence of $Q_i^{(N)}(\rho)$, 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 $\Psi_{jt}^{(N)}(\rho, z)$ and $\Lambda_{kt}^{(N)}(\rho, z)$ are formed from estimates of their components. In the previous sections we described our estimates of the compensation scheme, $w_s^{(N)}(x, z)$, 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 $\Psi_{jt}^{(N)}(\rho, z)$ for $j \in \{1, 2, 5\}$ using the definitions of $\Psi_{jt}(\rho, z)$ given in the previous section. However to estimate $\Psi_{3t}^{(N)}(\rho, z)$ we require an estimate of $\overline{v}_{st}(\rho, z) \equiv \exp[-\rho \overline{w}_s(z)/b_{t+1}]$. We use the fact that although $\overline{x}$ is unknown, $w_j(x)$ is a locally non-decreasing function in x. Following Brunk (1958), for each state $s \in \{1, \dots, S\}$, we rank the observations on returns in decreasing order by x_{s1}, $x_{s2}, \dots$ and so on, denoting by w_{s1}, $w_{s2}, \dots$ the corresponding compensations, and estimate $w_j(\overline{x}, z)$ with $\overline{w}_j^{(N)}(z)$ defined as:

$$\overline{w}_s^{(N)} \equiv \max_q \sum\nolimits_{r=1}^{q} \frac{w_{sr}}{q} \tag{5.3}$$

Finally to estimate $\Psi_{jt}^{(N)}(\rho, z)$ for $j \in \{4, 7, 8, 9\}$ and $\Lambda_{kt}^{(N)}(\rho, z)$ for $k \in \{1, \dots, 5\}$ we also require estimates of $g_s(x, z)$, which we denote by $g_s^{(N)}(\rho, x, z)$. Note from Proposition 3.1 that $g_2^{(N)}(\rho, x, z)$ can be directly found from $\overline{w}_j^{(N)}(z)$ but that $g_1^{(N)}(\rho, x, z)$ also requires an estimate of $\overline{h}(z)$. From the definition of a derivative:

$$\left[\frac{f_2(x, z)}{f_1(x, z)}\right] = \lim_{\Delta \to 0} \left[\frac{F_2(x + \Delta, z) - F_2(x, z)}{F_1(x + \Delta, z) - F_1(x, z)}\right] \tag{5.4}$$

it follows that:

$$\overline{h}(z) = \frac{\varphi_2(z)}{\varphi_1(z)} \left\{\lim_{x \to \infty} \left[\frac{f_2(x, z)}{f_1(x, z)}\right]\right\} = \frac{\varphi_2(z)}{\varphi_1(z)} \left\{\lim_{x \to \infty} \left[\frac{1 - F_2(x, z)}{1 - F_1(x, z)}\right]\right\} \tag{5.5}$$

Again, following Brunk (1958), we estimated $\overline{h}(z)$ with:

$$\overline{h}^{(N)}(z) \equiv \max_q \sum_{r=1}^{q} \left[\frac{\sum_{n=1}^{N} 1\{x_n \geq q, z_n = z, s_n = 2\}}{\sum_{n=1}^{N} 1\{x_n \geq q, z_n = z, s_n = 1\}}\right] \tag{5.6}$$

Under standard regularity conditions $\Psi_{jt}^{(N)}(\rho, z)$ and $\Lambda_{kt}^{(N)}(\rho, z)$ converge in probability to $\Psi_{jt}(\rho, z)$ and $\Lambda_{kt}(\rho, z)$. Although $w_j^{(N)}(x)$ and $f_j^{(N)}(x)$ 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:

$$N^{1/2}\left[\Psi_{jt}^{(N)}(\rho,z)-\Psi_{jt}(\rho,z)\right]\Longrightarrow N(0,\Omega_j(\rho,z)) \tag{5.7}$$

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

Suppose there exists a finite $\overline{x}_s$ such that $F_s(\overline{x}_s)<1$ and if $x>\overline{x}_s$, then $g_s(x)=0$. In that case the derivative of $w_s(x)$ at $\overline{x}_s$ is zero, following Parsons (1978) the norming constant is $N^{1/2}$, and hence (5.7) holds for all $j\in\{1,\dots,9\}$. An analogous result applies to $\Lambda_{kt}^{(N)}(\rho,z)$ too. However, if we relax the assumption about the existence of a finite $\overline{x}_s$, and assume, less restrictively, that $\lim_{x\to\infty}g_s(x)=0$ 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 (5.7).

Although the assumption about $\overline{x}_s$ 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 $\overline{x}_s$ such that $F_s(\overline{x}_s)<1$ with $g_s(x)=0$ for all $x>\overline{x}_s$, but where $a=2/3$ under the weaker assumption that $\lim_{x\to\infty}g_s(x)=0$. If a finite $\overline{x}_s$ does not exist and $A_N=N^{2/3}$, then asymptotic covariance matrix is driven by the tail observations. If a finite $\overline{x}_s$ does exist and $A_N=N$, then all the observations help determine the asymptotic properties.

Assuming the following condition (5.1) holds, then by Lemma 3.1 of Chernozhukov, Tamer and Hong (2007), c_δ is the $\delta-$quantile of the distribution of C.

Assumption 5.1 *For:*

$$C_i^{(N)}=\sup_{\rho\in\Gamma}\left[N^aQ_i^{(N)}(\rho)\right]$$

$P\{C^{(N)}\le c\}\to P\{C\le c\}$ *for each* $c\in[0,\infty)$*, where the probability distribution function for* C *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 ρ we consider away from zero.

Algorithm 1 (Subsampling) *Consider all subsets of the data with size* $N_b<N$*, where* $N_b\longrightarrow\infty$ *but* $N_b/N\longrightarrow 0$*, and denote the number of subsets by* B_N*. Define*

c_{0i} and $\Gamma_{0i}^{(N)}$ as:

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

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

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

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

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

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

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 $\rho \in [0.0037, 0.0042]$, there is no evidence that managers with different attitudes towards risk are sorting into different sectors.

6. PRIVATE INFORMATION AND MORAL HAZARD

To characterize the impact of private information and hidden actions we first compared the optimal contract under generalized moral hazard with a counterfactual compensation schedule that would apply in the analogous pure moral hazard model where the private information is made freely available to shareholders, which we denote by $w_{ms}(x)$. Then we estimated the losses shareholders would incur from letting the manager tend his own interests instead of maximizing the expected value of the firm, denoted by τ_1, and how much managers would gain from tending their own interests instead of their firm's, denoted by τ_2. Their difference, $\tau_1 - \tau_2$, 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 τ_3 and how much shareholders would pay to make the private information public rather than induce revelation through the manager's compensation schedule, τ_4. These parameters place lower bounds on what it would cost firms to rid itself of private information and hidden actions.

The estimated compensation schedule for the pure moral hazard model and the four measures are mappings of (ρ, θ). Because each element ρ induces a value for the remaining parameters θ through the mapping $\theta(\rho)$ we can write $\tau_k \equiv \tau_k(\rho)$ for $k \in \{1, \ldots, 4\}$. Our analysis of identification and the tests showed there is a set of values Γ for the risk aversion parameter such that its elements $\rho \in \Gamma$ are observationally equivalent. Consequently one approach to estimation is to delineate the set of the benefits and losses 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 $Q^{(N)}(\rho)$ in ρ to obtain a consistent estimator of one of the parameters in the identified set, denoted by $\widehat{\rho}$ and formed the estimator $\widehat{\tau}_k \equiv \tau_k(\widehat{\rho})$. Consistent estimators of the standard errors were computed for $\widehat{\rho}$ and hence $\widehat{\tau}_k$ by minimizing the criterion function $Q_2^{(j,N_b)}(\rho)$ in ρ for each subsample $j \in \{1, \ldots, B_N\}$ and computing the standard deviation of $\widehat{\rho}_j$.

The value obtained 0.0038 for $\widehat{\rho}$ is precisely estimated, with standard error 0.0005, and is comparable 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 ρ implies that a manager would pay \$185,600 to avoid an equal chance of losing versus gaining \$1,000,000.

The compensation schedule that would apply in the analogous pure moral hazard model where the private information is made freely available to shareholders is found by solving the pure moral hazard problem, that is setting $\eta_3^o = \eta_4^o = 0$ in the equations defining the first order conditions to obtain:

$$w_{ms}(x) = w_2^0 + \frac{b_{t+1}}{\rho} \log\left[1 + \eta_{mj}^0 (\alpha_2/\alpha_1)^{1/(b_t - 1)} - \eta_{mj}^o g_j(x)\right] \tag{6.1}$$

for $s \in \{1, 2\}$ where η_{ms}^o uniquely solves (2.13).

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 $v_{st}(x)^{-1}$ is a linear transformation of $v_{st}^{(m)}(x)^{-1}$ determined by the sum of the multipliers associated with hidden information $\eta_3 + \eta_4$ and the ratio of the multipliers for the incentive compatibility for the two cases η_2/η_{m2}^o. 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 is due to the terms $-\eta_3 h(x) - \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x)h(x)$, which appear in the first order condition for the hybrid model but drop out in the pure moral hazard case. 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 does not represent a theoretical implication or an internal inconsistency of our model. If the sincerity constraint did 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 $v_{st}(x)^{-1}$ is monotone increasing in $w_{ms}(x)$. 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 τ_1 through τ_3 depicted in the top three panels of Table 6, denoted $\widehat{\tau}_1$ through $\widehat{\tau}_3$ respectively, were computed from $\theta(\widehat{\rho})$ and numerically integrating over x where appropriate. The first measure, denoted τ_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 give that the state are reported truthfully. In symbols:

$$\tau_1 \equiv E\{x[1 - g_s(x)]\} = -E[xg_s(x)]$$

where the expectation following the second equality is over (x, j) 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 $\widehat{\tau}_1$ 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, $\widehat{\tau}_1$ is 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 τ_2. This second measure of generalized moral hazard can be expressed as the difference between w_2^0, the manager's reservation certainty equivalent wage to work under perfect monitoring, and w_1^0, the manager's reservation certainty equivalent wage to shirk. These certainty equivalents are derived from the participation constraint that w_1^0 and w_2^0 to obtain:

$$\tau_2 \equiv w_2^0 - w_1^0 = \frac{b_{t+1}}{\rho(b_t - 1)} \log\left(\alpha_2/\alpha_1\right) \tag{6.2}$$

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 $\widehat{\tau}_1$ and also the highest $\widehat{\tau}_2$. Similarly the qualitative effects of increasing assets, given firm employment and financial leverage, are almost identical for $\widehat{\tau}_2$ as for $\widehat{\tau}_1$; thus increasing assets lead to increases in $\widehat{\tau}_1$ and $\widehat{\tau}_2$ 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_2^0. Hence the expected value of a perfect monitor to shareholders, denoted τ_3, is the difference between expected compensation under the current optimal scheme and w_2^0, or:

$$\tau_3 \equiv E\left[w_s\left(x\right)\right] - w_2^0 = E\left[w_s\left(x\right)\right] - \frac{b_{t+1}}{\rho(b_t - 1)} \log\left(\alpha_2/\alpha_0\right)$$

Our estimates of τ_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:

$$\tau_4 \equiv E\left[w_s\left(x\right) - w_{ms}\left(x\right)\right]$$

where $w_{ms}\left(x\right)$ denotes the optimal compensation schedule for a pure moral hazard problem in state $s \in \{1, 2\}$ whose solution we described above.

Our estimates of τ_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 τ_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 τ_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. The difference between τ_3 and τ_4 :

$$\tau_3 - \tau_4 = w_{mj}(x) - w_2^o = \frac{b_{t+1}}{\rho} \log\left[1 + \eta_{mj}^0\left(\alpha_2/\alpha_1\right)^{1/(b_t-1)} - \eta_{mj}^o g_j(x)\right] > 0$$

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

This paper demonstrates the shape of the optimal contract for an important class of generalized moral hazard models embodies many restrictions that can be used in identification and testing, without imposing strong parametric assumptions on the conditional distributions of abnormal returns, or the functional form of the contract. We fully characterize the restrictions, and then apply the framework to a large panel data set on the compensation of chief executive officers, and the financial and accounting returns of their publicly trade firms. Our tests reject the pure moral hazard model, but we cannot reject the hybrid model, where there are hidden actions and hidden information. Finally we find that the benefits of contracting to deter managers from deviating from shareholder interests, and also the risk premium paid to executives for taking uncertain pay, are comparable to previous estimates obtained by estimating parametric models of pure moral hazard, and that the degree of private information varies considerably across sectors and over firm size.

Tepper School of Business, Carnegie Mellon University
ggayle@andrew.cmu.edu, http://www.andrew.cmu.edu/user/ggayle/

and

Tepper School of Business, Carnegie Mellon University
ramiller@andrew.cmu.edu, http://www.comlabgames.com/ramiller/

APPENDIX A: PROOFS

PROOF OF LEMMA 2.1: Let λ_r is the date t price of a contingent claim made on a consumption unit at date r, implying the bond price is defined as:

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

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

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

From Equation (15) of Margiotta and Miller (2000, page 680) the value of a manager with current wealth endowment e_{nt} announcing state s_t' in period t, choosing effort level l_{t1} in anticipation of compensation $w\left(s_t', x_{t+1}\right)$ at the beginning of period $t+1$ when he retires one period later is:

$$-b_t\alpha_2^{1/b_t}\alpha_0^{1-1/b_t}\left\{E_t\left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right)\right]\right\}^{1-1/b_t}\exp\left(-\frac{q_t+\rho e_{nt}}{b_{t+1}}\right)$$

the corresponding value from choosing effort level l_{t1} is:

$$-b_t\alpha_1^{1/b_t}\alpha_0^{1-1/b_t}\left\{E_t\left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right)\left[g_s\left(x_{t+1}\right)\right]\right]\right\}^{1-1/b_t}\exp\left(-\frac{q_t+\rho e_{nt}}{b_{t+1}}\right)$$

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

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

Dividing each expression through by the retirement utility it immediately follows that the manager chooses $l_t \equiv (l_{t0}, l_{t1}, l_{t2})$ to minimize the negative of expected utility:

$$\begin{aligned} & l_{t0} + \left(\frac{\alpha_j}{\alpha_0}\right)^{1/b_t} \left\{ E_t \left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right) \left[g_s\left(x_{t+1}\right) l_{t1} + l_{t2}\right] \right] \right\}^{1-1/b_t} \\ = \quad & l_{t0} + \left\{ \left(\frac{\alpha_j}{\alpha_0}\right)^{1/(b_t-1)} E_t \left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right) \left[g_s\left(x_{t+1}\right) l_{t1} + l_{t2}\right] \right] \right\}^{(b_t-1)/b_t} \end{aligned}$$

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

$$l_{t0} + \left(\frac{\alpha_j}{\alpha_0}\right)^{1/(b_t-1)} E_t \left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right) \left[g_s\left(x_{t+1}\right) l_{t1} + l_{t2}\right] \right]$$

Multiplying through by the factor $(\alpha_0/\alpha_j)^{1/(b_t-1)}$ yields the minimand in Lemma 2.1:

$$(\alpha_0/\alpha_j)^{1/(b_t-1)} l_{t0} + E_t \left[\exp\left(-\frac{\rho w\left(s_t', x_{t+1}\right)}{b_{t+1}}\right) \left[g_s\left(x_{t+1}\right) l_{t1} + l_{t2}\right] \right]$$

Q.E.D.

Proof of Lemma 2.2: Multiplying each first order equation in the text by $\varphi_s v_{st}(x) f_s(x)$, then summing and integrating over x yields:

$$1 = \eta_0 \left[\sum\nolimits_{s=1}^{2} \int_{\underline{x}}^{\infty} \varphi_s v_{st}(x) f_s(x)\, dx \right] \equiv \eta_0 E\left[v_{st}(x)\right]$$

where we make use of the complementary slackness conditions. Substituting for $\eta_0 = \{E\left[v_{st}(x)\right]\}^{-1}$ 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_0 we obtain:

$$1 = \{E\left[v_{st}(x)\right]\}^{-1} v_{2t}(x) + \eta_3 v_{2t}(x) + \eta_2 v_{2t}(x) \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x) \right] + \eta_4 v_{2t}(x)$$

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:

$$\begin{aligned} 1 \quad &= \quad \{E\left[v_{st}(x)\right]\}^{-1} E_2\left[v_2(x)\right] + \eta_3 E_2\left[v_{2t}(x)\right] + \eta_4 E_2\left[v_{2t}(x)\right] \\ &= \quad E_2\left[v_2(x)\right] \left(\{E\left[v_{st}(x)\right]\}^{-1} + \eta_3 + \eta_4 \right) \end{aligned}$$

Dividing through by $E_2\left[v_2(x)\right]$ proves the second numbered item in the lemma. *Q.E.D.*

Proof of Lemma 2.3: The second numbered item in Lemma 2.2, the incentive compatibility constraint for the second state, and the assumption of the lemma, successively imply:

$$\begin{aligned}\int v_2(x) f_2(x)\,dx &\leq \int v_1(x) f_1(x)\,dx \\ &\leq \int \left(\frac{\alpha_1}{\alpha_2}\right)^{1/(b_t-1)} v_1(x) f_1(x) g_1(x)\,dx \\ &= \int \left(\frac{\alpha_1}{\alpha_2}\right)^{1/(b_t-1)} v_1(x) f_2(x) g_2(x)\,dx\end{aligned}$$

Q.E.D.

Proof of Lemma 2.4: 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 T periods is simply a replication of the one period optimal contract. Suppose that for all $s \in \{t+1, t+2, \ldots, T-1\}$:

$$V_{sk}(e_s) = -b_s \alpha_k^{\frac{1}{b_s}} \alpha_0^{1-\frac{1}{b_s}} \{E_s[v_{k,s+1}]\}^{1-\frac{1}{b_s}} \exp\left(-\frac{a_s + \rho e_s}{b_s}\right)$$

Then from Lemma 1 the continuation of the optimal contract beginning in period $t+1$ yields a utility of:

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

because the participation constraint is satisfied with equality at that time. Therefore the problem of participating at time t and possibly continuing with the firm for more than one period reduces to the problem of participating at time t one period at most, solved in Lemma 2.1. *Q.E.D.*

Proof of Proposition 3.1: Writing $\theta^* \equiv (\alpha_1, \alpha_2, g_1(x), g_2(x))$ we prove $\theta^* = \theta_t(\rho^*)$. To conserve on notation, we write $v_{st}(x) \equiv \exp[-\rho^* w_s(x)/b_{t+1}]$ and $\overline{v}_{st} \equiv \exp[-\rho^* \overline{w}_s / b_{t+1}]$, and prove the proposition by successively treating each component of θ^*.

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

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

2. Proving $\eta_2 = \eta_{2t}(\rho^*)$ comes from substituting the solution for η_0 into the first order condition for the second state, which yields:

$$v_{2t}(x)^{-1} = \{E[v_{st}(x)]\}^{-1} + \eta_2[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)] + \eta_3 + \eta_4$$

Taking expectations we obtain:

$$E_2\left[v_{2t}(x)^{-1}\right] = \{E[v_{st}(x)]\}^{-1} + \eta_2[(\alpha_2/\alpha_1)^{1/(b_t-1)} - 1] + \eta_3 + \eta_4$$

Also:

$$\overline{v}_{2t}^{-1} = \{E[v_{st}(x)]\}^{-1} + \eta_2(\alpha_2/\alpha_1)^{1/(b_t-1)} + \eta_3 + \eta_4$$

Differencing the second two equations:

$$\eta_2 = \overline{v}_{2t}^{-1} - E_2\left[v_{2t}(x)^{-1}\right] = \eta_{2t}(\rho^*)$$

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

$$v_{2t}(x)^{-1} = \{E[v_{st}(x)]\}^{-1} + \eta_2[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)] + \eta_3 + \eta_4$$

from:

$$\overline{v}_{2t}^{-1} = \{E[v_{st}(x)]\}^{-1} + \eta_2(\alpha_2/\alpha_1)^{1/(b_t-1)} + \eta_3 + \eta_4$$

to give:

$$\overline{v}_{2t}^{-1} - v_{2t}(x)^{-1} = \eta_2 g_2(x)$$

Upon rearrangement, we appeal to the result in Item 2, that $\eta_2 = \eta_{2t}(\rho^*)$ to obtain:

$$g_2(x) = \eta_2^{-1}\left[\overline{v}_{2t}^{-1} - v_{2t}(x)^{-1}\right] = g_{2t}(x,\rho^*)$$

4. To show $\alpha_1 = \alpha_1(\rho^*)$ we substitute the solution for η_2 above into the first order condition for the second state evaluated at the limit $x \to \infty$ to obtain:

$$\overline{v}_{2t}^{-1} = \{E[v_{st}(x)]\}^{-1} + \left\{\overline{v}_{2t}^{-1} - E_2\left[v_{2t}(x)^{-1}\right]\right\}(\alpha_2/\alpha_1)^{1/(b_t-1)} + \eta_3 + \eta_4$$

or, upon appealing to Lemma 2.2:

$$\begin{aligned}(\alpha_2/\alpha_1)^{1/(b_t-1)} &= \frac{\overline{v}_{2t}^{-1} - \{E[v_{st}(x)]\}^{-1} - \eta_3 - \eta_4}{\overline{v}_{2t}^{-1} - E_2\left[v_2(x)^{-1}\right]} \\ &= \frac{\overline{v}_{2t}^{-1} - \{E[v_{2t}(x)]\}^{-1}}{\overline{v}_{2t}^{-1} - E_2\left[v_2(x)^{-1}\right]}\end{aligned}$$

Making α_1 the subject of the equation:

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

5. To prove $\eta_4 = \eta_{4t}(\rho^*)$ we first multiply the first order conditions for the first state by $v_{1t}(x)$,

after solving for η_0 to obtain:

$$\begin{aligned} & 1-\eta_1 v_{1t}(x)\left[(\alpha_2/\alpha_1)^{1/(b_t-1)}-g_1(x)\right] \\ = \quad & \{E[v_{st}(x)]\}^{-1} v_{1t}(x)-\eta_3 v_{1t}(x)\, h(x)-\eta_4(\alpha_1/\alpha_2)^{1/(b_t-1)} v_{1t}(x)\, g_2(x)\, h(x) \end{aligned}$$

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

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

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

$$\eta_3=\{E_2[v_{2t}(x)]\}^{-1}-\{E[v_{st}(x)]\}^{-1}-\eta_4$$

we obtained from lemma 2.2 reduces this expression to:

$$\begin{aligned} 1 \quad = \quad & \{E[v_{st}(x)]\}^{-1} E_1[v_{1t}(x)]-\eta_4(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x)\, g_2(x)\, h(x)] \\ & -\left[\{E_2[v_{2t}(x)]\}^{-1}-\{E[v_{st}(x)]\}^{-1}-\eta_4\right] E_1[v_{1t}(x)\, h(x)] \end{aligned}$$

Upon collecting terms:

$$\begin{aligned} & \eta_4\left\{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x)\, g_2(x)\, h(x)]-E_1[v_{1t}(x)\, h(x)]\right\} \\ = \quad & \{E[v_{st}(x)]\}^{-1} E_1[v_{1t}(x)]-E_1[v_{1t}(x)\, h(x)]\left[\{E_2[v_{2t}(x)]\}^{-1}-\{E[v_{st}(x)]\}^{-1}\right]-1 \end{aligned}$$

so solving for η_4 we now have:

$$\begin{aligned} \eta_4 \quad & = \quad \frac{\{E[v_{st}(x)]\}^{-1} E_1[v_{1t}(x)]-E_1[v_{1t}(x)\, h(x)]\left[\{E_2[v_{2t}(x)]\}^{-1}-\{E[v_{st}(x)]\}^{-1}\right]-1}{(\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[v_{1t}(x)\, g_2(x)\, h(x)]-E_1[v_{1t}(x)\, h(x)]} \\ & = \quad \eta_{4t}(\rho^*) \end{aligned}$$

6. Proving $\eta_3=\eta_{3t}(\rho^*)$ follows directly from the lemma above, which implies:

$$\eta_3 \equiv \{E_2[v_{2t}(x)]\}^{-1}-\eta_{4t}(\rho^*)-\{E[v_{st}(x)]\}^{-1}$$

7. To prove $\eta_1=\eta_{1t}(\rho^*)$, rewrite the first order condition for the first state as:

$$\eta_1\left[(\alpha_2/\alpha_1)^{1/(b_t-1)}-g_1(x)\right]=v_{1t}(x)^{-1}-\{E[v_{st}(x)]\}^{-1}+\eta_3 h(x)+\eta_4(\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x)\, h(x)$$

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

$$\eta_1(\alpha_2/\alpha_1)^{1/(b_t-1)}=\overline{v}_{1t}^{-1}-\{E[v_{st}(x)]\}^{-1}+\eta_3 \overline{h} \tag{A.1}$$

Making η_1 the subject of the equation now demonstrates $\eta_1=\eta_1(\rho^*)$.

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

$$\eta_1 g_1(x)=\overline{v}_{1t}^{-1}-v_{1t}(x)^{-1}+\eta_3\left[\overline{h}-h(x)\right]-\eta_4(\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x)\, h(x)$$

Dividing both sides by η_1 we thus establish $g_1(x)=g_{1t}(x, \rho^*)$ for all t.

Q.E.D.

PROOF OF PROPOSITION 3.2: We prove $\Psi_{1t}(\rho) = \Psi_{2t}(\rho) = 0$, derive the expression for $\Psi_{4t}(\rho)$ given by (3.7) and show the sincerity constraint implies $\Psi_{6t}(\rho) \geq 0$. Setting $\eta_3 = \eta_4 = 0$, and appealing to symmetry implies from the statement of Proposition 1 that:

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

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

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

$$\eta_1 \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - 1 \right] = E_1\left[v_{1t}(x)^{-1}\right] - \{E[v_{st}(x)]\}^{-1} + \eta_3 \frac{\varphi_2}{\varphi_1} + \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[g_2(x) h(x)]$$

Differencing this expression from (A.1) yields:

$$\eta_1 = \overline{v}_{1t}^{-1} + \eta_3 \overline{h} - E_1\left[v_{1t}(x)^{-1}\right] - \eta_3 \frac{\varphi_2}{\varphi_1} - \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} E_1[g_2(x) h(x)]$$

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

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

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

Q.E.D.

PROOF OF PROPOSITION 3.3: We prove each numbered item in order.

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

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

so:

$$1 - \overline{v}_{st}(\rho)\{E_2[v_{2t}(x,\rho)]\}^{-1} > 1 - \overline{v}_{st}(\rho) E_2\left[v_{2t}(x,\rho)^{-1}\right]$$

and consequently the quotient:

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

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

2. For $\rho \in \overline{\Gamma}_2$ the participation constraint implies:

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

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

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

4. Since $g_2(x) \to 0$ as $x \to \infty$ and $g_2(x) > 0$, it follows from the first order condition associated with the second state that $v_2(x) \geq \overline{v}_2$ for all $x \in R$, so from the definition of $v_2(x)$ we infer:

$$\overline{w}_2 \equiv \lim_{x\to\infty}\left[w_2(x)\right] = \sup_{x\in R}\left[w_2(x)\right]$$

Hence $v_{2t}(x,\rho) \geq \overline{v}_{2t}(\rho)$ for all $\rho > 0$ and $x \in R$, thus proving from its definition that $g_{2t}(x,\rho) \geq 0$. In the pure moral hazard model, $\eta_3 = \eta_4 = 0$, and hence the same logic shows $g_{1t}(x,\rho) \geq 0$ if $\rho \in \overline{\Gamma}_1$. Finally the inequality $\Psi_{9t}(\rho) \geq 0$ guarantees $g_{1t}(x,\rho) \geq 0$ for $\rho \in \overline{\Gamma}_2$.

5. In the proof of Item 2 of Proposition 1 we established $\eta_2 \equiv \overline{v}_{2t}^{-1} - E_2\left[v_{2t}(x)^{-1}\right]$, and also proved in Item 2 above that $\overline{v}_{2t} \leq v_{2t}(x)$. Hence $\eta_2 \geq 0$. The same arguments apply to $\eta_{1t}(\rho)$ when $\rho \in \overline{\Gamma}_1$ thus establishing $\eta_{1t}(\rho) \geq 0$ in that case.

Q.E.D.

Proof of Proposition 3.4: We show that any compensation schedule from a regular data generating process satisfying the restrictions implied by some $\widehat{\rho} \in \overline{\Gamma}_i$ for $i \in \{1,2\}$ solves the Kuhn Tucker formulation of an optimal contracting problem for a model which we abbreviate by the parameterization $\left(\widehat{\rho},\widehat{\theta}\right)$ where:

$$\begin{aligned}\widehat{\theta} &\equiv \left(\widehat{\alpha}_1,\widehat{\alpha}_2,\widehat{g}_1\left(x\right),\widehat{g}_2\left(x\right)\right)\\ &\equiv \left(\alpha_{1t}\left(\widehat{\rho}\right),\alpha_{2t}\left(\widehat{\rho}\right),g_{1t}\left(x,\widehat{\rho}\right),g_{2t}\left(x,\widehat{\rho}\right)\right)\\ &\equiv \theta_t\left(\widehat{\rho}\right)\end{aligned}$$

Similarly we abbreviate the mappings $\eta_{jt}\left(\widehat{\rho}\right)$ defined in Proposition 1 by $\widehat{\eta}_j$. Since the objective function for the underlying maximization problem is strictly concave, and the constraints are linear, there is a unique stationary point determined by the first order and complementary slackness conditions in the Kuhn Tucker formulation. As we demonstrate in this proof, by construction $w^*\left(x\right)$ satisfies the first order condition and the complementary slackness conditions for the parameterization $\left(\widehat{\rho},\widehat{\theta}\right)$ with multipliers $\widehat{\eta}\equiv\left(\widehat{\eta}_1,\widehat{\eta}_2,\widehat{\eta}_3,\widehat{\eta}_4\right)$. Consequently the compensation schedule for $\left(\widehat{\rho},\widehat{\theta}\right)$ is $w^*\left(x\right)$. It is convenient to treat the pure and hybrid moral hazard models separately.

We first consider models of pure moral hazard and suppose $\widehat{\rho}\in\overline{\Gamma}_1$. We show that the participation and incentive compatibility constraints are met with equality, that the first order condition is satisfied and that $\widehat{\eta}_j>0$ solve the equations defining the Kuhn Tucker multipliers.

1. Since $\Psi_{1t}\left(\widehat{\rho}\right)=\Psi_{2t}\left(\widehat{\rho}\right)=0$, the competitive selection constraints ensure:

$$\left\{E_1\left[v_{1t}\left(x,\widehat{\rho}\right)\right]\right\}^{1-b_t}=\left\{E_2\left[v_{2t}\left(x,\widehat{\rho}\right)\right]\right\}^{1-b_t}$$

so

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

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 $\widehat{\alpha}_1$, $\widehat{\alpha}_2$ and $\widehat{g}_s\left(x\right)$:

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

Multiplying both sides by:

$$\widehat{\eta}_s\equiv\overline{v}_{st}\left(\widehat{\rho}\right)^{-1}-E_s\left[v_{st}\left(x,\widehat{\rho}\right)^{-1}\right]$$

we obtain the first order condition for the state s as:

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

3. The equation above implies:

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

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

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

Since $\widehat{\eta}_s>0$ the incentive compatibility condition is satisfied with equality.

4. Since the first order condition is satisfied by $\widehat{\eta}_s$ for each x we may write:

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

and substitute $v_{st}(x,\widehat{\rho})^{-1}$ into the expression:

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

to obtain:

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

the equality following from the incentive compatibility constraint, which we proved in Item 3 above is satisfied by $\widehat{\eta}_{3s}$. Consequently $\widehat{\eta}_s$ is a solution to the equation:

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

in η, which define the Kuhn Tucker multipliers. This completes the proof for the pure moral hazard case.

We now consider the hybrid case, and suppose $\widehat{\rho}\in\overline{\Gamma}_2$. 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 $\widehat{\eta}$ solves the equations defining the Kuhn Tucker multipliers.

1. The definition of $\widehat{\alpha}_2\equiv\{E[v_{st}(x,\widehat{\rho})]\}^{1-b_t}$ directly implies the participation constraint is met with equality. The truth telling and sincerity constraints are directly imposed from $\widehat{\rho}$ belonging to $\overline{\Gamma}_2$ through the equality $\Psi_{7t}(\widehat{\rho})\Psi_{8t}(\widehat{\rho})=0$ and the two inequalities $\Psi_{7t}(\widehat{\rho})\geq 0$, $\Psi_{8t}(\widehat{\rho})\geq 0$. This only leaves the three tasks of establishing $\left(\widehat{\rho},\widehat{\theta}\right)$ satisfies the first order conditions, that the incentive compatibility conditions are met with equality, and that $\widehat{\eta}$ solves the equations defining the Kuhn Tucker multipliers.

2. Noting the definitions of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{g}_2(x)$ and $\widehat{\eta}_2$ are identical to their counterparts in the pure

moral hazard model we can appeal to Item 2 in the moral hazard case to establish:

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

From the definition of $\widehat{\eta}_3$ we have:

$$\{E[v_{st}(x,\widehat{\rho})]\}^{-1}+\widehat{\eta}_3+\widehat{\eta}_4=\{E_2[v_{2t}(x,\widehat{\rho})]\}^{-1}$$

Subtracting the first equation from the second:

$$\{E[v_{st}(x,\widehat{\rho})]\}^{-1}+\widehat{\eta}_3+\widehat{\eta}_2\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}-\widehat{g}_2(x)\right]+\widehat{\eta}_4=v_{2t}(x,\widehat{\rho})^{-1}$$

we obtain the first order condition for the second state in the hybrid model.

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

$$\begin{aligned}\widehat{\eta}_1\widehat{g}_1(x) &= \overline{v}_{1t}(\widehat{\rho})^{-1}-v_{1t}(x,\widehat{\rho})^{-1}+\widehat{\eta}_3\left[\overline{h}-h(x)\right]\\ &\quad -\widehat{\eta}_4(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}\widehat{g}_2(x)h(x)\end{aligned}$$

From the definition of $\widehat{\eta}_1$:

$$\eta_3\overline{h}=\widehat{\eta}_1(\alpha_2/\alpha_1)^{1/(b_t-1)}-\{E[v_{st}(x,\widehat{\rho})]\}^{-1}-\overline{v}_{1t}(\widehat{\rho})^{-1}$$

Substituting for $\eta_3\overline{h}$ in to the expression above for $\widehat{\eta}_1\widehat{g}_1(x)$, now yields the first order condition in the first state upon rearrangement.

$$v_{1t}(x,\widehat{\rho})^{-1}=\{E[v_{st}(x,\widehat{\rho})]\}^{-1}+\widehat{\eta}_1\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}-\widehat{g}_1(x)\right]-\widehat{\eta}_3h(x)-\widehat{\eta}_4(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}\widehat{g}_2(x)h(x)$$

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 $\widehat{\eta}_2$ are identical to their counterparts in the pure moral hazard model, which implies from Item 2 in the moral hazard case that:

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

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

$$\begin{aligned}&E_2\left\{\widehat{\eta}_2v_{2t}(x,\widehat{\rho})\left[\widehat{g}_2(x)-(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}\right]\right\}\\ =\ &E_2\left\{\widehat{\eta}_2v_{2t}(x,\widehat{\rho})\left[\{E_2[v_{2t}(x,\widehat{\rho})]\}^{-1}-v_{2t}(x,\widehat{\rho})^{-1}\right]\right\}\\ =\ &0\end{aligned}$$

Multiplying the expression we derived for the first order condition in Item 2 above by $v_{1t}(x,\widehat{\rho})$ and taking the expectation conditional on the first state yields implies:

$$\begin{aligned}&\widehat{\eta}_1E_1\left\{v_{1t}(x,\widehat{\rho})\left[\widehat{g}_1(x)-(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}\right]\right\}\\ =\ &E_1[v_{1t}(x,\widehat{\rho})]\{E[v_{st}(x,\widehat{\rho})]\}^{-1}-\widehat{\eta}_3E_1[v_{1t}(x,\widehat{\rho})h(x)]\\ &-\widehat{\eta}_4(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}E_1[v_{1t}(x,\widehat{\rho})\widehat{g}_2(x)h(x)]-1\end{aligned}$$

Successively substituting the definitions of $\widehat{\eta}_3$ and $\widehat{\eta}_4$ into the right side of the equation:

$$\begin{aligned}
& E_1\left[v_{1t}(x,\widehat{\rho})\right]\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}-\widehat{\eta}_3 E_1\left[v_{1t}(x,\widehat{\rho})h(x)\right] \\
& -\widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{1/(b_t-1)} E_1\left[v_{1t}(x,\widehat{\rho})\widehat{g}_2\left(x\right)h\left(x\right)\right]-1 \\
= & E_1\left[v_{1t}(x,\widehat{\rho})\right]\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}-\left(\left\{E_2\left[v_{2t}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}-\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}-\widehat{\eta}_4\right) E_1\left[v_{1t}(x,\widehat{\rho})h(x)\right] \\
& -\widehat{\eta}_4\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{1/(b_t-1)} E_1\left[v_{1t}(x,\widehat{\rho})\widehat{g}_2\left(x\right)h\left(x\right)\right]-1 \\
= & E_1\left[v_{1t}(x,\widehat{\rho})\right]\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}-\left(\left\{E_2\left[v_{2t}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}-\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}\right) E_1\left[v_{1t}(x,\widehat{\rho})h(x)\right]-1 \\
& +\widehat{\eta}_4\left\{E_1\left[v_{1t}(x,\widehat{\rho})h(x)\right]-\left(\widehat{\alpha}_1/\widehat{\alpha}_2\right)^{1/(b_t-1)} E_1\left[v_{1t}(x,\widehat{\rho})\widehat{g}_2\left(x\right)h\left(x\right)\right]\right\} \\
= & 0
\end{aligned}$$

Therefore:

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

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 $\widehat{v_{st}^{-1}}(x,\widehat{\rho})$, $\widehat{g}_s\left(x\right)$ and $\widehat{\alpha}_s$ for $s\in\{1,2\}$ and the defined elements $\widehat{\eta}_2$ through $\widehat{\eta}_4$ that:

$$\begin{aligned}
\widehat{v}_{1t}(x,\widehat{\rho})^{-1} \quad = \quad & \left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}+\widehat{\eta}_1\left[\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}-\widehat{g}_1\left(x\right)\right] \\
& -\widehat{\eta}_3 h\left(x\right)-\widehat{\eta}_4\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}\widehat{g}_2\left(x\right)h\left(x\right)
\end{aligned}$$

and

$$\widehat{v_{2t}^{-1}}(x,\widehat{\rho})=\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}+\widehat{\eta}_2[\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}-\widehat{g}_2\left(x\right)]+\widehat{\eta}_3+\widehat{\eta}_4$$

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

$$\begin{aligned}
0 \quad = \quad & E_2\left[\widehat{v}_{1t}(x,\widehat{\rho})-\widehat{v}_{1t}(x,\widehat{\rho})\right] \\
= \quad & E_2\left\{\left[\begin{array}{c}\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}+\widehat{\eta}_1\left[\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}-\widehat{g}_1\left(x\right)\right] \\ -\widehat{\eta}_3 h\left(x\right)-\widehat{\eta}_4\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}\widehat{g}_2\left(x\right)h\left(x\right)\end{array}\right]^{-1}\right\} \\
& -E_2\left\{\left[\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}+\widehat{\eta}_2[\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}-\widehat{g}_2\left(x\right)]+\widehat{\eta}_3+\widehat{\eta}_4\right]^{-1}\right\}
\end{aligned}$$

Similarly, since the incentive compatibility and first order conditions are satisfied with equality in each state by $\widehat{\rho}$, $\widehat{\theta}$ and $\widehat{\eta}$:

$$\begin{aligned}
0 \quad = \quad & E_1\left\{v_{1t}(x,\widehat{\rho})\left[\widehat{g}_1\left(x\right)-\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}\right]\right\} \\
= \quad & E_1\left[\frac{\widehat{g}_1\left(x\right)-\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}}{\left(\begin{array}{c}\left\{E\left[v_{st}\left(x,\widehat{\rho}\right)\right]\right\}^{-1}+\widehat{\eta}_1\left[\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}-\widehat{g}_1\left(x\right)\right] \\ -\widehat{\eta}_3 h\left(x\right)-\widehat{\eta}_4\left(\widehat{\alpha}_2/\widehat{\alpha}_1\right)^{1/(b_t-1)}\widehat{g}_2\left(x\right)h\left(x\right)\end{array}\right)}\right]
\end{aligned}$$

and:

$$\begin{aligned} 0 &= E_2\left\{\widehat{v_{2t}^{-1}}(x,\widehat{\rho})\left[\widehat{g}_2(x)-(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}\right]\right\} \\ &= E_2\left[\frac{\widehat{g}_2(x)-(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}}{\{E[v_{st}(x,\widehat{\rho})]\}^{-1}+\widehat{\eta}_2[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}-\widehat{g}_2(x)]+\widehat{\eta}_3+\widehat{\eta}_4}\right]\end{aligned}$$

From its definition $\widehat{\eta}_4=0$ when $\Psi_{6t}(\widehat{\rho})>0$, and when $\Psi_{6t}(\widehat{\rho})=0$:

$$\begin{aligned} 0 &= E_2[v_{2t}(x,\widehat{\rho})]-(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}E_2[v_{1t}(x,\widehat{\rho})\widehat{g}_2(x)] \\ &= E_2\left\{\frac{1}{1+\widehat{\eta}_2[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}-\widehat{g}_2(x)]+\widehat{\eta}_3+\widehat{\eta}_4}\right\} \\ &\quad -E_2\left\{\frac{(\widehat{\alpha}_1/\widehat{\alpha}_2)^{1/(b_t-1)}\widehat{g}_2(x)}{\left(\begin{array}{c}\{E[v_{st}(x,\widehat{\rho})]\}^{-1}+\widehat{\eta}_1\left[(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}-\widehat{g}_1(x)\right] \\ -\widehat{\eta}_3h(x)-\widehat{\eta}_4(\widehat{\alpha}_2/\widehat{\alpha}_1)^{1/(b_t-1)}\widehat{g}_2(x)h(x)\end{array}\right)}\right\}\end{aligned}$$

This equations demonstrate that $\widehat{\eta}_1$ through $\widehat{\eta}_4$ solve the equations defining the Lagrange multipliers for the parameterization defined by $\left(\widehat{\rho},\widehat{\theta}\right)$ in the hybrid moral hazard model, thus completing the proof.

Q.E.D.

Proof of Lemma 4.1: Without loss of generality we suppress the dependence of w_{nt} on (z_{nt},s_{nt},b_t), let $\widetilde{x}$ denote net abnormal return, and let V denote the value of the firm at the beginning of the period. For any net abnormal return $\widetilde{x}$, and for any value of the firm at the beginning of the period V, we denote by:

$$w=\Lambda\left(\widetilde{x}+\frac{w}{V}\right) \tag{A.2}$$

for each $w\in W\equiv\{w:w=w(x)\text{ for some }x\in X\}$ any relation that satisfies:

$$w(x)=\Lambda\left(\widetilde{x}+\frac{w(x)}{V}\right)$$

for all x and:

$$\widetilde{x}=x-w(x)/V$$

We remark that $w(x)$ is one solution to the relation defined by Λ. Suppose that for some pair $(\widetilde{x},V)$ there exists two distinct values of $w\in W$, denoted $w_1\equiv w(x_1)$ and $w_2\equiv w(x_2)$ satisfying the relation:

$$w_i=w(x_i)=\Lambda\left(\widetilde{x}+\frac{w_i}{V}\right)$$

for $i\in\{1,2\}$. From the definition of $\widetilde{x}$ we obtain:

$$\widetilde{x}=x_i-w(x_i)/V$$

which implies:

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

Therefore $w(x)$ is the unique solution to the relation defined by Λ for each pair $(\widetilde{x}, V)$ if:

$$V(x_2 - x_1) \neq w(x_2) - w(x_1)$$

for all $(x_1, x_2) \in R^2$. We denote that unique solution by $\Lambda_1(\widetilde{x}, V)$. Having proved $w(x) = \Lambda_1(\widetilde{x}, V)$, the lemma now follows because the measurement error on compensation is assumed to independent of $(\widetilde{x}, V)$ so $E[\widetilde{w}\,|\widetilde{x}, V] = \Lambda_1(\widetilde{x}, V)$. *Q.E.D.*

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" J*ournal of Accounting Research,* 23 pp.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 pp.1-39.

[4] Berle, A. A., and G. C. Means (1932): *The Modern Corporation and Private Propert*y. Harcourt, Brace, & World Inc., New York.

[5] Brunk, H. D. (1958): " On the Estimation of Parameters Restricted by Inequalities", *Annals of Mathematical Statistics*, Vol. 29, 437-454.

[6] Chernozhukov, V., H. Hong and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, Vol 75. No. 5, pp. 1243-1284.

[7] Chesher, A. (2007): "Instrumental Values," *Journal of Econometrics*, 139, 15-34.

[8] Ciliberto, F. and E. Tamer (2007): "Market Structure and Multiple Equilibria in Airline Markets", Department of Economics, Northwestern University.

[9] Dubois P. and, T. Vukina (2005): "Optimal Incentives under Moral Hazard and Heterogeneous Agents: Evidence from Production Contracts Data", Working paper

[10] Duffo E. , R. Hanna, and S. Ryan (2007): "Monitoring Works: Getting Teachers to Come to School", Working paper, MIT

[11] Ferrall C. and Shearer B. (1999): "Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records," *Review of Economic Studies*, 66, 2, 309-338.

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

[13] Fudenberg, D., B. Holmstrom and P. Milgrom (1990): " Short-Term Contracts and Long-Term Agency Relationships." *Journal of Economic Theory*, Vol. 50, pp. 1-31.

[14] Gayle, G.-L. and R. A. Miller (2008a). " The Paradox of Insider Information and Performance Pay" Forthcoming, *CESifo Economic Studies* .

[15] Gayle, G.-L. and R.A. Miller, (2008b): "Has Moral Hazard become a More important Factor in Managerial Compensation?", Forthcoming *American Economic Review*

[16] Gayle, G.-L., L. Golan and R.A. Miller (2008): "Promotion, Turnover and Compensation in the Executive Market", Tepper School of Business, Carnegie Mellon University

[17] Grossman, S. and O. Hart (1983): "An Analysis of the Principal-Agent Problem" *Econometrica, 1983,* 51 pp. 7-46.

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

[19] Ho, K. (2008); "Insurer-Provider Networks in the Medical Care Market" , Forthcoming *American Economic Review* .
[20] Ho, K., J. Ho, and J. Mortimer (2008): "The Effects of Full-Line Forcing Contracts" Department of Economics, Columbia University.
[21] Huang Y. ,I. Perrigne, and Q. Vuong:(2007): " Nonlinear Pricing in Yellow Pages", Department of Economics, Pennsylvania State University.
[22] Jaffe, J.F. (1974) : " Special information and Insider Trading", *Journal of Business*, ,Vol. 47, pp. 410-428.
[23] Levine, A. (2008) : "Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry", Department of Economics, Stanford University.
[24] Lorie, J.H. and V. Niederhoffer (1968): " Predictive and Statistical Properties of Insider Trading," *Journal of Law and Economics*, vol. 11, pp.35-51.
[25] Malcomson, J. M. and F. Spinnewyn (1988):" The Multiperiod Principal-Ageny Problem", *Review of Economic Studies*, Vol. 55, No. 3 pp 391-407.
[26] Margiotta, M. and R.A. Miller (*2000) :* "Managerial Compensation and The Cost of Moral Hazard" *International Economic Review, August* , 41 (3) pp. 669-719.
[27] Myerson, R.(1982): "Optimal coordination mechanisms in generalized principalagent problem". *Journal of Mathematical Economics*, 10:67–81.
[28] 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.
[29] Pakes, A. , J. Porter, K. Ho and J. Ishii (2206): " Moment Inequalities and Their Application", CEMMAP Working Paper.
[30] Parsons, V. L. (1978): The Limiting Distribution of the Isotonic Estimator at a Point", Unpublished Manuscript.
[31] Perrigne, I. and Q. Vuong:(2007): "Nonparametric Identification of Incentive Regulation Models", CAPCP Working Paper.
[32] Rey, P. and B. Salanie (1990): "Long-term, Short-term and Renegotiation: On the Value of Commitment in Contracting", *Econometrica*, vol. 58, No. 3, pp. 597-619.
[33] Romano, J.P., Shaikh, A.M., 2006. "Inference for Identifiable Parameters in Partially Identified Econometric Models" working paper.
[34] Seyhun, H. N. (1986): " Insiders' Profits, Cost of Trading, and Market Efficiency," *Journal of Financial Economics*, 1986, vol.16, pp. 189-212.
[35] Seyhun, H. N (1992a): " The Effectiveness of Insider Trading Sanctions," *Journal of Law and Economics*, vol. 35, pp. 149-182.
[36] Seyhun, H. N (1992b): " Why Does Aggregate Insider Trading Predict Future Stock Returns?" *The Quarterly Journal of Economics*, Vol. 107, pp. 1303-1331.
[37] Shearer, B. (2004) : " Piece Rates, Fixed Wages and Incentives: Evidence From a Field Experiment", *Review of Economic Studies*, 71, 2, 513-534.
[38] Wright, F.T. (1981): " The Asymptotic Behavior of Monotone Regression Estimates", *The Annals of Statistics*, Vol. 9, No 2, 443-448.

Table 1: Time Series Summary
(Assets in millions of 2000 $US, Employees in Thousand, Compensation in thousand of 2000 $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

Standard deviation in parentheses. Earnings in thousands of year-2000 US$

Table 2: Cross Section Summary
(Assets in millions of 2000 $US, Employees in Thousand, Compensation in thousand of 2000 $US)

Variable	Primary	Consumer	Services
Observation	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	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 3: Cross Section Summary
(Compensation in thousand of 2000 $US)

	Variable		Primary		Consumer		Services
S_{1t}		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 4: Estimate $\varphi(j)$

		Primary			Consumer			Services			Total
S_{1t}		N	Good	Bad	N	Good	Bad	N	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 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]
Nunber replication	100	100	100
Point	Estimates	from overall	Sample
ρ			0.0038
standard error			0.0005

Table 6: Structural Estimates and Simulations
τ_2, τ_3 and τ_2 are measured in US100,000 of dollars
τ_1 is measured in percentage per year

		Primary		Consumer		Services	
	S_{1t}	Estimates	S.E.	Estimates	S.E.	Estimates	S.E.
τ_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**
τ_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**
τ_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**
τ_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

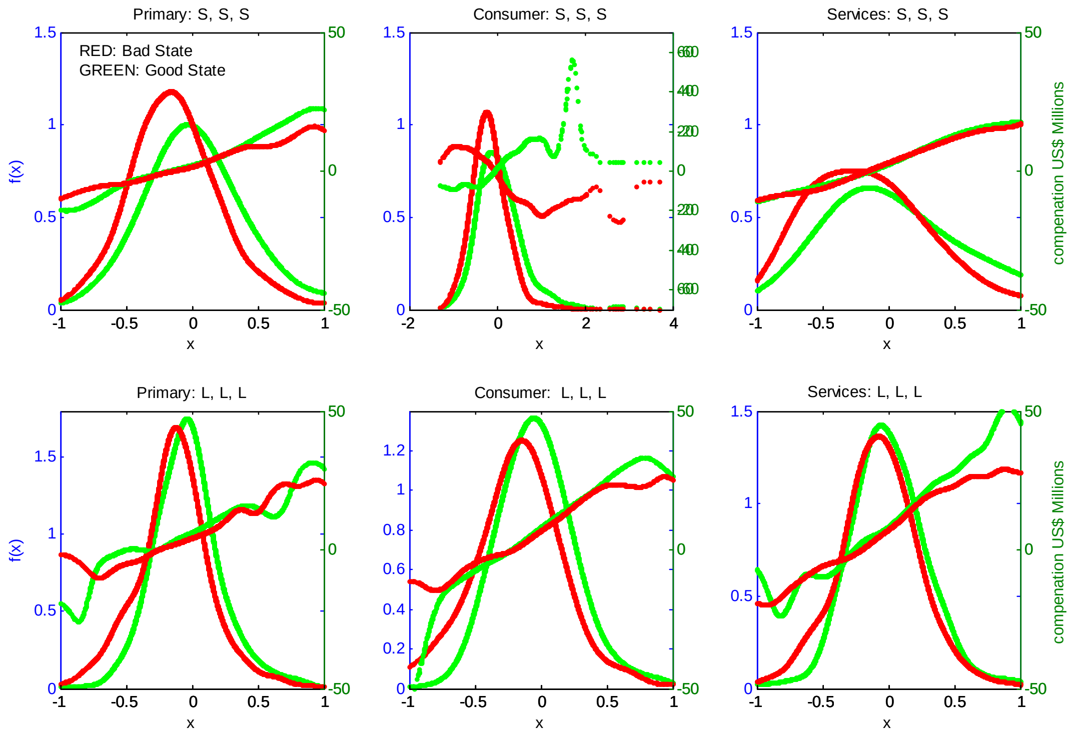

Figure 1: Compensation Schedules and Financial Return Densities

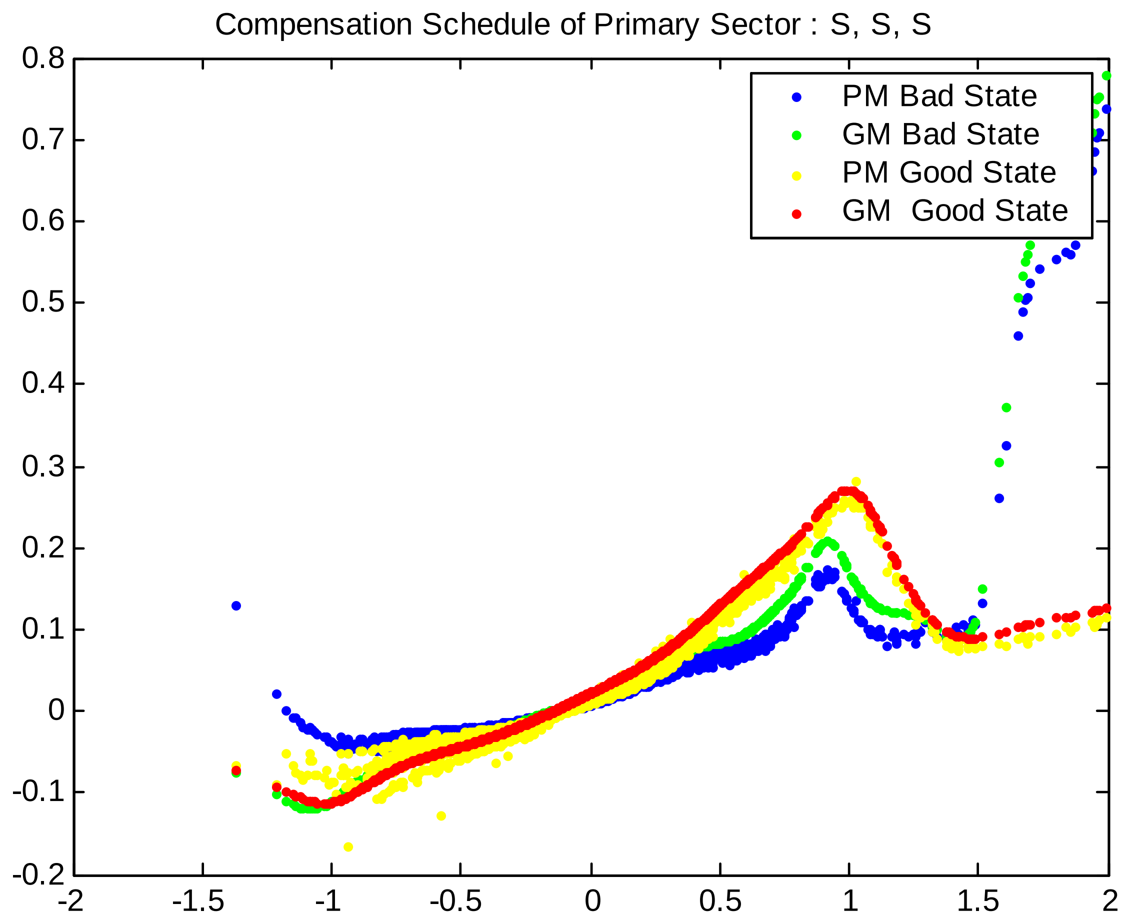

Figure 2: Predicted Pure versus Actual Generalized Compensation Schedules