Identifying and Testing Generalized Moral Hazard Models of Managerial Compensation*

George-Levi Gayle and Robert A. Miller
Tepper School of Business, Carnegie Mellon University
June 24, 2008

Abstract

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

1 Introduction

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

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

*Preliminary and Incomplete. Do not quote without expressed permission of the authors.
of executives in large corporations necessarily make them privy to information about their firm’s performance that is not easily accessible by stockholders.

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 Niederhofer (1968), Jaffe (1974), and Finnerty (1976) draw similar conclusions. More recently, Seyhun (1992a) finds compelling evidence that insider trading volume, frequency, and profitability all increase 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 of the total variation in one-year-ahead returns. To summarize, hidden information is an economically important phenomenon in executive compensation.

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

Section 2 presents a theoretical framework for exploring moral hazard and hidden information. In Section 3 we characterize the set of feasible facing shareholders and the solution to their cost minimization problem is derived in Section 4. We describe the sources of data on chief executive officers of publicly traded companies and explain how they were compiled in Section 5, while Section 6 summarizes their main features. Our analysis of identification in Section 7 builds upon our previous work of identification in moral hazard Gayle and Miller (2008). In this paper we lay out necessary and sufficient conditions for identifying generalized models of moral hazard of the form we describe in Section 2. We develop test statistics for both the pure and hybrid models reports the results of our testing on the CEO compensation data in Section 6, and conclude in Section 7 by estimating the various components to the costs of hidden information and moral hazard.

2 The Model

In our model the manager of a firm is subject to moral hazard, but also has private information about the firm’s future returns at the beginning of each period. Shareholders do not observe the prospects 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 man-
ger, 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 firms prospects, 
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 se-
quentially 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_t \in \{0, 1\} , where l_t = 1 if the manager chooses to be engaged outside the firm and l_t = 0 
if she chooses to be engaged inside the firm.

If l_t = 0, the prospects of the firm are then fully revealed to the manager but partially 
hidden to the shareholders. Without loss of generality we index shareholders information 
sets in the empirical portions of the analysis only. 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 (s_t, x_{t+1}) 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

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

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 \( F_s (x_{t+1}) \), and assume it is differentiable with 
density \( f_s (x_{t+1}) \). Similarly, let \( f_s (x_{t+1}) g_s (x_{t+1}) \) denote the probability density function for
abnormal returns in period \( t \) when the manager shirks. Since \( f_s(x) g_s(x) \) is a density, \( g_s(x) \) must be a positive mapping with \( E_s [g_s(x)] = 1 \), where the expectation is taken with respect to \( f_s(x) \). Compensation to the manager is denoted by \( w_{t+1} \equiv w(s'_t, x_{t+1}) \). We also assume there is an upper range of returns that, conditional on the state \( s \), might be achieved with diligence, but is extremely unlikely to occur if the manager shirks. Formally we assume

\[
\lim_{x \to \infty} [g_s(x)] = 0
\]

for each \( s \in \{1, 2\} \).

We assume there are a complete set of markets for all publicly disclosed events, with price measure \( \Lambda_t \) defined on \( F_t \) and derivative \( \lambda_t \) and denoted 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:

\[
- \sum_{t=0}^{T} \sum_{j=0}^{J} \beta^t \alpha_j l_{tj} \exp (-\rho c_t)
\]

where \( \beta \) is the constant subjective discount factor, \( \rho \) is the constant absolute level of risk aversion, and \( \alpha_j \) is a utility parameters with consumption equivalent \(-\rho^{-1} \log (\alpha_j)\) that measures the distaste from working at level \( j \in \{0, 1, 2\} \). We assume \( \alpha_2 > \alpha_1 \) 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.\(^1\) 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 \( w(s'_t, x_{t+1}) \), 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

---
\(^1\)Malcomson and Spinnewyn (1988), Fudenberg, Homstrom and Milgrom (1990) and Rey and Salanie (1990) have independently established conditions under which long term optimal contracts can be implemented via a sequence of one period contracts in dynamic models of generalized moral hazard, and the proof of Lemma 2 in the Appendix draws extensively upon their results.
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 off 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 1** If the manager anticipating a contract of $w(s_t', x_{t+1})$ retires in period $t$ or period $t+1$ by setting $(1-l_{t0})(1-l_{t+1,0}) = 0$, she optimally chooses $(l_t, s_t')$ to minimize

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

(1)

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

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

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 (1) and substituting in $v_s(x_{t+1})$, the participation constraint is thus:

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

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 (1) 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

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

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

\[ 0 \leq \int [v_1(x) - v_2(x)] f_2(x) \, dx \equiv E_2 [v_1(x) - v_2(x)] \]

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

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

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

This leads to a formulation of the cost minimization problem shareholders solve. Shareholders maximize the value of the firm, inducing the manager to make choices that serve
their interests. It is straightforward to show from the participation constraint that the cost
minimizing contract for employing the manager to shirk is a constant wage of

\[ w_{l+1}^o = \frac{b_t+1}{\rho(b_t-1)} \log \left( \frac{\alpha_1}{\alpha_0} \right) \]

that just offsets the value of leaving the firm and consequently does not depend on the state.
It pays shareholders to induce the manager to distinguish between pairs of states if and only
if it is more profitable to create incentives that motivate her to work diligently in at least one
of the states than to shirk in both. Denote by \(w_t^o(x)\) the optimal contract that induces truth
telling, sincerity and diligence in the \(s^{th}\) state. Our discussion implies that the manager will
choose \((l_1, l_2)\) for \(s \in \{1, 2\}\) to maximize the value of the firm, namely

\[ \sum_{s=1}^{2} \int_{x}^{\infty} \varphi_s \left\{ [zx - w_t^o(x)] l_2 - [zx g_s(x) - w_t^o] l_1 \right\} f_s(x) \, dx \]

We now derive the cost minimizing contract that induces diligence in at least one state. Our formulation satisfies the Kuhn Tucker conditions, permitting us to use Lagrangian methods to characterize the optimal short term contract. Deriving \(w(s_t, x_t+1)\) to minimize expected compensation subject to the three constraints is equivalent to choosing \(v_s(x_{t+1})\) to maximize

\[ \sum_{s=1}^{2} \int_{x}^{\infty} \varphi_s \log v_s(x_{t+1}) f_s(x_{t+1}) \, dx_{t+1} \equiv E [\log v_s(x)] \]
subject to the same three 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 \( x_{t+1} \). Appealing to the Kuhn-Tucker theorem, we formulate the problem as maximizing a Lagrangian with respect to \( v_s(x_{t+1}) \).

Let \( h(x) \) denote the weighted density

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

Recalling from the definition of \( v_s(x) \) that \( \log v_s(x) \) is proportional to \( -w_s(x) \), the shareholders maximize:

\[
\sum_{s=1}^{2} \varphi_s \int_{\mathbb{R}} \left\{ \log v_s(x) + \eta_0 \left[ (\alpha_0/\alpha_2)^{1/(b_t-1)} - v_{st} \right] \right\} f_s(x) \, dx
\]

\[
+ \sum_{s=1}^{2} \varphi_s \int_{\mathbb{R}} \left\{ \eta_s v_s(x) \left[ (g_s(x) - (\alpha_2/\alpha_1)^{1/(b_t-1)}) \right] \right\} f_s(x) \, dx
\]

\[
+ \varphi_2 \eta_3 \int_{\mathbb{R}} [v_1(x) - v_2(x)] f_2(x) \, dx
\]

\[
+ \varphi_2 \eta_4 \int \left[ (\alpha_1/\alpha_2)^{1/(b_t-1)} v_1(x) g_2(x) - v_2(x) \right] f_2(x) \, dx
\]

where \( \eta_0 \) through \( \eta_4 \) are the shadow values assigned to the linear constraints.

The solution involves solving the Lagrangian multipliers associated with the minimization problem. The proof of Lemma 2 below shows that \( \eta_0 = (\alpha_2/\alpha_0)^{1/(b_t-1)} \). That still leaves several cases to analyze. The value of the firm is computed in each case, by relaxing different combinations of constraints to see whether which ones are satisfied by the constraint inequalities. For example if neither the truth telling nor the sincerity constraints are binding, then \( \eta_3 = \eta_4 = 0 \) and the optimization problem reduces to the pure moral hazard problem solved in Margiotta and Miller (2000). If at least one constraint is violated in this solution, then \( \eta_3 + \eta_4 > 0 \), and at least one of the constraints is binding. When the sincerity constraint does not bind, the truth telling constraint binds, and diligence in both states is induced, we substitute the first order condition into the incentive compatibility and truth telling constraints give use the following system of three equations in the remaining three unknowns \( \eta_1, \eta_2, \) and \( \eta_3 \) when \( \eta_4 = 0 \). They are:

\[
\frac{\int_{\mathbb{R}} f_2(x) \, dx}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_3 h(x) + \eta_1 \left[ (\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x) \right]}
\]

\[
= \int_{\mathbb{R}} \frac{f_2(x)}{(\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_3 + \eta_2 \left[ (\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x) \right]} \, dx
\]
0 \begin{align*}
&= \int_{\mathbb{R}} \frac{\left[ g_1(x) - \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} \right] f_1(x)}{(\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_3 h(x) + \eta_1 \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} - \eta_1 g_1(x)} \, dx \\
&= \int_{\mathbb{R}} \frac{\left[ g_2(x) - \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} \right] f_2(x)}{(\alpha_2/\alpha_0)^{1/(b_t-1)} - \eta_3 + \eta_2 \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} - \eta_2 g_2(x)} \, dx
\end{align*}

The Kuhn Tucker theorem guarantees there is a unique positive solution to this equation system. If the sincerity constraint is met by the unconstrained solution \((v_1(x), v_2(x))\), then we conclude that \(\eta_3 = \eta_4 = 0\). Alternatively if the sincerity condition is not satisfied by the unconstrained solution, then \(\eta_4 > 0\) and the new set of conditions solving the remaining four unknowns \(\eta_1\) through \(\eta_4\) are:

\begin{align*}
\int_{\mathbb{R}} \frac{f_2(x) \, dx}{(\alpha_2/\alpha_0)^{1/(b_t-1)}} &+ \eta_3 + \eta_4 + \eta_2 \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} - \eta_2 g_2(x) \\
\int_{\mathbb{R}} \frac{f_2(x) \, dx}{(\alpha_2/\alpha_0)^{1/(b_t-1)}} &- \eta_3 h(x) + \eta_1 \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} - \eta_1 g_1(x) + \eta_4 \left( \frac{\alpha_1}{\alpha_2} \right)^{1/(b_t-1)} g_2(x) h(x)
\end{align*}

These equations can also be used to solve the other cases. Setting \(\eta_3 = 0\), we check whether the truth telling constraint is satisfied by an inequality, in which case the other multipliers are found by ignoring the first equation in the groups of three or four, the less constrained solution yielding a lower cost than the case in which \(\eta_3 > 0\). Setting \(\eta_4 = 0\) yields \(l_{s1} = 1\), in which case \(w_s(x) = w_s^0\).

Substituting these values back into the two first order equations yields the solution to the compensation schedule as a function of the states. As before we check which constraints are satisfied with inequalities by relaxing the truth telling and sincerity constraints. Lemma 2 characterizes its solution and then verifies the assumptions of Fudenberg, Holmstrom 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 The long term optimal contract can be implemented by a $T$ period replication of the short term optimal contract, which is defined by the shadow values $\eta_1$ through $\eta_4$ defined by the equation system, along with compensation schedules $v_1(x)$ and $v_2(x)$

$$v_1(x)^{-1} = (\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_1 \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_1(x)\right] - \eta_3 h(x) - \eta_4 (\alpha_1/\alpha_2)^{1/(b_t-1)} g_2(x) h(x)$$

$$v_2(x)^{-1} = (\alpha_2/\alpha_0)^{1/(b_t-1)} + \eta_2 \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g_2(x)\right] + \eta_3 + \eta_4$$

In the special case of pure moral hazard, whether state is revealed before or after the contract is made is immaterial to both managers and shareholders. If the contract is made before the state is revealed, 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 (??) and Lemma 2. Similarly two of the Kuhn Tucker equations drop out, leaving the solution to $\eta_s$ uniquely defined by:

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

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

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

In this case, however, the first order conditions simplify to the other case where there is only one participation constraint, because following the same logic as the proof to the second Lemma it is straightforward to show that $\eta_{01} = \eta_{02}$. Consequently the expected utility of the manager is identical in both states, which we demonstrate in the next section provides an equation for the pure moral hazard model with multiple states that is violated by the hybrid model.

3 Compiling the Data

The main data for our empirical 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 COMPSTAT North America database and monthly stock price data from the Center for Securities Research (CSP) 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.

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 (2008). 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 $\tilde{x}_{nt}$, net abnormal returns, denote the financial return on its stock net of the financial return on the market portfolio in period $t$. Gross abnormal returns for the $n^{th}$ firm in period $t$ attributable to the manager’s actions are defined as net abnormal returns plus compensation as a ratio of firm equity

$$x_{nt} \equiv \tilde{x}_{nt} + \frac{w_{nt}}{v_{n,t-1}} \quad (3)$$

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

$$\tilde{w}_{nt} = w_{nt} + \varepsilon_{nt} \quad (4)$$

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

**Lemma 3 (3)** If

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

for all $(x_1, x_2) \in \mathbb{R}^2$, then

$$w_{nt} = E[\bar{w}_{nt} | x_{nt}, z_{nt}, s_{nt}, b_t, V_{n,t-1}]$$  \hspace{1cm} (5)

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

$$w_{nt}^{(N)} = \frac{\sum_{m=1, m \neq n}^N w_{mt} I \{ z_{mt} = z_{nt}, s_{mt} = s_{nt} \} K \left( \frac{x_{mt} - x_{nt}}{\delta_{x,N}}, \frac{b_m - b_{n,t-1}}{\delta_{b,N}} \right)}{\sum_{n=1, m \neq n}^N I \{ z_{mt} = z_{nt}, s_{mt} = s_{nt} \} K \left( \frac{x_{mt} - x_{nt}}{\delta_{x,N}}, \frac{b_m - b_{n,t-1}}{\delta_{b,N}} \right)}$$

where $K(\cdot)$ is a bivariate probability density function with full support and $\delta_N \equiv (\delta_{x,N}, \delta_{b,N})$ 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

$$x_{nt}^{(N)} \equiv \bar{x}_{nt} + \frac{w_{nt}^{(N)}}{v_{n,t-1}}$$  \hspace{1cm} (6)

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

$$f_{z,s}^{(NT)}(x) = \frac{\sum_{t=1}^T \sum_{n=1}^N \prod \{ z_{nt} = z, s_{nt} = s \} K_x \left( \frac{x_{nt}^{(NT)} - x}{\delta_{x,NT}} \right)}{\delta_{x,NT} \sum_{t=1}^T \sum_{n=1}^N \prod \{ z_{nt} = z, s_{nt} = s \}}$$  \hspace{1cm} (7)

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 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 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. 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

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

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 2 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. In the sample period, financial returns from the stock market to diversified shareholders ranged from a yield to 45 percent in one year to a loss of 14 percent returns in another. Far greater is the variation around the market return by individual firms. As explained above, this latter variation in abnormal returns, rather than variability due to aggregate factors, is critical to explaining managerial compensation. The collective signal managers send about business, average accounting returns, is highly correlated with financial returns, almost without exception rising and falling together. Note though that accounting returns have a considerably higher standard deviation, in part attributable to fixed effects across firms, but also to higher idiosyncratic variability over time.

The term structure of interest rates underlying the bond price series were constructed from data on Treasury bills of varying maturities, and the prices were derived using methods

---

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

3Paranthetically 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.
described in Gayle and Miller (2008). Table 2 shows that over this period, year to year bond price fluctuations are in the order of 5 to 10 percent, but there is no discernible trend in this aggregate variable. Because the optimal contract depends on bond prices, in principle, variation in bond prices can be used to identify the risk aversion parameter from the participation constraint. In practice, the variation over this period is too small to exploit in estimation.

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

From Table 2 we see that the mean compensation of managers fluctuates much more than real wages for professional employees, the trough of $1.7 million for the 12 years occurring only 2 years after the peak of $4.7 million and just one year before the second highest, $4.6 million. Variation in CEO compensation between firms within years is greater than the average variation over the 12 years, with a standard deviation of approximately 3 to 10 times the mean, depending on the year, although this feature of the data is partly due to individual variation, reflected in the sectorial differences evident in Table 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 \( \left( u_{nt}^{(N)}, x_{nt}^{(N)} \right) \). 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 manager’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 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 $\rho$. 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 $\alpha_j$ values the nonpecuniary features of engaging in activity $j \in \{1, 2\}$ within the firm relative to the total util 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 $\alpha_1, \alpha_2$ and $\rho$.

To facilitate the discussion we partition the parameter space into $\rho$, 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 $(\rho^*, \theta^*)$ the generalized model of moral hazard generating the data process $(x_t, s_t, w_t)$, where $\theta^* \equiv (\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$. Following Chesher (2007) we say the structural parameter $(\rho, \theta) \in R^+ \times \Theta$ with true value $(\rho^*, \theta^*)$ is identified if $(\rho^*, \theta^*)$ can be written as a functional of the conditional distribution of the observed variables. If no such functional exists, then the correspondence from the conditional distribution of the observed variables defines an equivalence class.

We first prove that if $\rho^*$ is known, then $(\alpha_1^*, \alpha_2^*, g_1^*(x), g_2^*(x))$ 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 $\theta^*$ can be written as $(\rho^*, \theta_t(\rho^*))$ for all $t$. In models of pure moral hazard, we set the shadow values $\eta_{3t}(\rho) = \eta_{4t}(\rho) = 0$ for all $\rho$ and $t$, and the functions defined in the proposition below reduce to the equations derived in Gayle and Miller (2008).

**Proposition 4** For each $(s, t,)$, let

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

and

$$\overline{v}_{st}(\rho) \equiv \exp \left[ -\rho \overline{w}_{st} / b_{t+1} \right]$$

Then $\theta^* = \theta_t(\rho^*)$ for all $t$ in the hybrid model 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

\[
\begin{align*}
\alpha_{2t}(\rho) & \equiv \left\{ E\left[v_{st}(x, \rho)\right]\right\}^{1-b_t} \\
\eta_{2t}(\rho) & \equiv \frac{1}{\eta_{2t}(\rho)^{-1} - E_2\left[v_{2t}(x, \rho)^{-1}\right]}
\end{align*}
\]

\[
\begin{align*}
g_{2t}(x, \rho) & \equiv \frac{\eta_{2t}(\rho)^{-1}\left[\eta_{2t}(\rho)^{-1} - v_{2t}(x, \rho)^{-1}\right]}{1 - v_{2t}(\rho)^{-1}} \\
\alpha_{1t}(\rho) & \equiv \alpha_{2t}(\rho) \left\{ \frac{1 - v_{2t}(\rho)^{-1}}{1 - v_{2t}(\rho)^{-1} E_2\left[v_{2t}(x, \rho)^{-1}\right]} \right\}^{1-b_t}
\end{align*}
\]

\[
\begin{align*}
\eta_{4t}(\rho) & = \frac{\left\{ E\left[v_{st}(x, \rho)\right]\right\}^{-1} E_1\left[v_{1t}(x, \rho)\right] - 1}{\left(\alpha_1/\alpha_2\right)^{1/(b_1-1)} E_1\left[v_{1t}(x, \rho)\right] g_{2t}(x, \rho) h(x) - E_1\left[v_{1t}(x, \rho) h(x)\right]} \\
\eta_{3t}(\rho) & \equiv \left\{ E_2\left[v_{2t}(x, \rho)\right]\right\}^{-1} - \eta_{4t}(\rho) - \left\{ E\left[v_{st}(x, \rho)\right]\right\}^{-1} \\
\eta_{1t}(\rho) & = \left[\alpha_{1t}(\rho)/\alpha_{2t}(\rho)\right]^{1/(b_1-1)} \left[\eta_{1t}(\rho)^{-1} \left\{ E\left[v_{st}(x, \rho)\right]\right\}^{-1} + \eta_{3t}(\rho)\right]^{-1} \\
g_{1t}(x, \rho) & \equiv \eta_{1t}(\rho)^{-1} \left\{ \frac{\eta_{1t}(\rho)^{-1} - v_{1t}(x, \rho)^{-1}}{-\eta_{4t}(\rho) \eta_{1t}(\rho)} \left[\eta_{1t}(\rho)^{-1} \eta_{1t}(\rho) + \eta_{3t}(\rho)\right]^{-1} g_{2t}(x, \rho) h(x)\right\}
\end{align*}
\]

Since our test statistics and estimators are based on sample analogues to the population moments that define these parameters as a function of \( \rho \), we briefly describe the intuition to interpret explaining relating the. The certainty equivalent of a lottery that yields \( w(x) \) for a person with absolute risk aversion parameter \( \rho \) is MMM. In our the optimal contract for our model the manager is indifferent between accepting the job which provides nonpecuniary benefits of 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 \( \left\{E_2\left[v_{2t}(x, \rho)\right]\right\}^{-1} \) and \( E_2\left[v_{2t}(x, \rho)^{-1}\right] \). By definition \( b_t > 1 \) and the inverse function \( v^{-1} \) is convex decreasing, 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 \( v_{2t}(\rho)^{-1} = \exp[\rho \eta_{1t}(\rho)] \) 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 \( \rho \), 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.

Proposition 1 implies the set of identified parameters can be indexed by a Borel set of values for the risk aversion parameter \( \rho \in \mathbb{R}^+ \) that are observationally equivalent. Next,
we analyze some restrictions derived from optimizing behavior in the model that limit the admissible values of \( \rho \). Within the class of pure moral models we denote the set of admissible values by \( \Gamma_1 \). For the class of hybrid models the corresponding set is denoted by \( \Gamma_2 \). The restrictions on \( \rho \) defining these sets arise from several sources.

One source is the optimal choice of effort, a fixed wage to induce shirking, versus an incentivized scheme to induce truth telling, sincerity and diligence. Recalling \( V_{st} \) denotes the value of the firm in state \( s \) at time \( t \), it is optimal to select a contract inducing diligence if and only if

\[
E_s \left[ V_{st} x - w_{st} (x) \right] \geq V_{st} E_s \left[ x g_s (x) \right] - w_t^{(1)}
\]

where \( w_t^{(1)} \) denotes the shirking wage. For each \( t \) and \( s \in \{1, 2\} \) define

\[
\Psi_{st} (\rho) = \frac{b_{t+1} \log [\alpha_{st} (\rho)]}{\rho (b_t - 1)} - E_s [w_{st} (x)] - V_{st} E_s [x g_s (x)]
\]

Then optimally selecting diligent work implies \( \Psi_{st} (\rho) \geq 0 \) because \( E_s [x] = 0 \) and the optimal wage for shirking is

\[
w_t^{(1)} = \frac{b_{t+1} \log (\alpha_{st}^*)}{\rho^* (b_t - 1)}
\]

Optimality conditions that determine the compensation schedule for diligent work also provide a source for culling values of \( \rho \) that do not belong to \( \Gamma_1 \). Here it is convenient to discuss the pure and hybrid models separately. For models of pure moral hazard, we have shown elsewhere, Gayle and Miller (2008), that competitive selection constraints of the form

\[
\Psi_{3t} (\rho) = \Psi_{4t} (\rho) = 0
\]

hold for all \( t \), where \( \Psi_{3t} (\rho) \) and \( \Psi_{4t} (\rho) \) are respectively defined by

\[
\Psi_{s+2,t} (\rho) \equiv \left\{ E_1 \left[ v_{11} (x, \rho) \right] \right\}^{1-b_t} - \left\{ E_s \left[ v_{st} (x, \rho) \right] \right\}^{1-b_t}
\]

for \( s \in \{1, 2\} \). Defining the set

\[
\Gamma_1 \equiv \{ \rho > 0 : \Psi_{st} (\rho) \geq 0 \text{ and } \Psi_{s+2,t} (\rho) = 0 \text{ for } s \in \{1, 2\} \text{ and all } t \}
\]

our discussion implies \( \rho^* \in \Gamma_1 \) or equivalently \( \Gamma_1 \subseteq \Gamma_1 \).

The truth telling constraint in the hybrid moral hazard framework yields an equation in \( \rho \) that is similar to the competitive selection equations in pure moral hazard models. Defining

\[
\Psi_{5t} (\rho) \equiv E_2 \left[ v_{2t} (x, \rho) - v_{1t} (x, \rho) \right]
\]

\( \Psi_{5t} (\rho) = 0 \) for all \( t \) and \( \rho \in \Gamma_2 \). We also require \( \Psi_{6t} (\rho) \equiv \eta_{3t} (\rho) \), the multiplier associated with truth telling, to be strictly positive in the hybrid model.

The sincerity constraint in the hybrid model yields restrictions on the data too, regardless of whether it binds or not. Letting

\[
\Psi_{6t} (\rho) \equiv E_2 \left[ v_{1t} (x, \rho) \frac{[v_{2t} (\rho)]^{-1} - [v_{2t} (x, \rho)]^{-1}}{[v_{2t} (\rho)]^{-1} - E_2 [v_{2t} (x, \rho)]^{-1} - v_{2t} (x, \rho)} \right]
\]
the proof to the next proposition shows that the sincerity constraint can be expressed by the inequality \( \Psi_{st} (\rho) \geq 0 \) for admissible values of \( \rho \). Furthermore if the sincerity constraint is not binding, that is \( \Psi_{st} (\rho) > 0 \), then its associated multiplier, now denoted \( \Psi_{7t} (\rho) \equiv \eta_{4t} (\rho) \), is zero.

Finally we impose a restriction on the space of functions containing \( g_1 (x) \), the likelihood ratio in the hybrid model. To ensure \( g_{1t} (x, \rho) \geq 0 \) with unit mass we require

\[
\Psi_{st} (\rho) \equiv E_1 [1 \{g_{1t} (x, \rho)\} - 1] \geq 0
\]

Summarizing the restrictions directly applied to the hybrid model, \( \rho^* \) must satisfy the two equalities relating to effort selection, \( \Psi_{1t} (\rho) \geq 0 \) and \( \Psi_{2t} (\rho) \geq 0 \), the truth telling constraint, \( \Psi_{3t} (\rho) = 0 \), the complementary slackness condition for sincerity, \( \Psi_{4t} (\rho^*) \Psi_{7t} (\rho^*) = 0 \), two inequalities that the multipliers for sincerity and truth telling are positive, \( \Psi_{7t} (\rho^*) \geq 0 \) and \( \Psi_{9t} (\rho^*) \geq 0 \), and the restriction that the likelihood ratio \( g_{1t} (x, \rho) \) be nonnegative for all \( x \), namely \( \Psi_{st} (\rho) \geq 0 \). Defining the set of risk aversion parameters as

\[
\Gamma_2 \equiv \left\{ \rho > 0: \quad \Psi_{jt} (\rho) \geq 0 \text{ for } j \in \{1, 2, 6, \ldots, 9\} \text{ and } \Psi_{5t} (\rho) = \Psi_{6t} (\rho) \Psi_{7t} (\rho) = 0 \right\}
\]

our discussion implies \( \rho^* \in \Gamma_2 \) in moral hazard models with hidden information.

**Proposition 5** \( \Gamma_2 \subseteq \Gamma_2 \).

Are there any other valid restrictions on the sets of admissible parameters to further shrink \( \Gamma_i \) for \( i \in \{1, 2\} \) when \( \Gamma_i \) 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 \( \Gamma_1 \) and \( \Gamma_2 \), 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 \( \Gamma_i \).

**Proposition 6** Supposing \( \rho \in \Gamma_i \) 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 \( \rho \in \Gamma_1 \) and \( \sum_{s=1}^{2} \varphi_s E_s [w_s (x)] > w^{(2)} \) for \( \rho \in \Gamma_2 \)
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. \( \eta_{st}(\rho) \geq 0 \)

It remains to show that if \( \rho \in \bar{\Gamma}_i \) for \( i \in \{1, 2\} \), then \( \rho \) is observationally equivalent to \( \rho^* \). 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 \( s \), impose the regularity condition that

\[
\lim_{x \to \infty} [w_s(x)] = \sup_{x \in \mathbb{R}} [w_s(x)] \equiv \bar{w}_s
\]

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 \( t \).\(^4\) In this broader context we entertain four possibilities, tested in the next section. Letting \( \phi \) 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 \( \gamma \) that obeys the restrictions described above. Consequently \( \Gamma_i = \bar{\Gamma}_i \) for \( i \in \{1, 2\} \) and the bounds we have constructed are tight.

**Proposition 7** A regular data generating process is observationally equivalent to a pure moral hazard model indexed by \( \gamma > 0 \) if \( \gamma \in \bar{\Gamma}_1 \), and is observationally equivalent to a hybrid moral hazard model indexed by \( \gamma > 0 \) if \( \gamma \in \bar{\Gamma}_2 \).

\(^4\)Gayle and Miller (2008) discuss the additional competitive selection conditions that arise from shifts in the bond price over time when there are no secular changes in the nonpecuniary benefits from management.
Propositions 1 through 4 imply that the small number of restrictions the pure moral hazard model imposes on regular data generating processes arise solely from the effort selection and competitive selection conditions. The representation defined in Proposition 1 that maps the risk aversion parameter into the remaining parameters of the problem embodies all the other features that characterize preference parameters, likelihood ratios and Lagrange multipliers in the optimizing behavior in pure moral hazard models. Furthermore it is easy to see that the competitive selection constraint is robbed of any empirical content if we replace the single parameter $\alpha_2$ with two parameters $\alpha_{2s}$ for $s \in \{1, 2\}$ to reflect dependence of utility on the state and the nonpecuniary benefits from diligent work are permitted to vary across the different states in an unrestricted manner. Consequently the assumptions limiting the variation in the nonpecuniary benefits of diligent work across different states provide necessary and sufficient conditions to identify and estimate models of pure moral hazard.

In the hybrid model the effort selection constraint is the same as in the pure moral hazard model. The sincerity constraint, not relevant for the model of pure moral hazard, provides additional information about the hybrid model, whether it is binding or not, and the interaction between both those constraints and incentive compatibility also provides empirical content. The competitive selection constraint associated with the pure moral hazard model is replaced by a similar, but not identical, truth telling constraint. Supposing the nonpecuniary utility is a hidden parameter the manager privately observes, then the truth telling constraint, which relates to the second state, limits admissible values of $\rho$, because it does not depend on how much nonpecuniary utility would be received in the first state, but in a model of pure moral hazard has no empirical content, as we explained. In this case the hybrid model formal nests the pure moral hazard model, thus demonstrating that the hybrid model has more empirical content than the pure moral hazard model.

6 Testing Moral Hazard and Hidden Information

Because $\Gamma_i$ is defined by a vector function of equalities and inequalities in $\rho$, 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 $z \in \{1, \ldots, Z\}$, and to allow for exogenous background influences $z$ in estimation and testing, we now write $\Psi_{jt}(\rho, z)$ for $\Psi_{jt}(\rho)$. Armed with this expanded notation, the sets $\Gamma_i$ can be expressed as

$$\Gamma_i = \{\rho > 0 : Q_i(\rho) = 0\}$$
for \( i \in \{1, 2\} \)

\[
Q_1(\rho) = \sum_{t=1}^{T} \sum_{z=1}^{Z} \sum_{s=1}^{2} \left\{ \min \left[0, \Psi_{st}(\rho, z)\right]^2 + \Psi_{s+2,t}(\rho, z)^2 \right\}
\]

(8)

\[
Q_2(\rho) = \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ \sum_{j=1,2,6,7} \min \left[0, \Psi_{jt}(\rho, z)\right]^2 + \Psi_{5t}(\rho, z)^2 + \left[\Psi_{6t}(\rho, z) \Psi_{7t}(\rho, z)\right]^2 \right\}
\]

(9)

Appealing to Proposition X, we would 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 \( \Gamma_1 \) is empty. Similarly the hybrid model would be rejected if and only if \( \Gamma_2 \) is empty. For \( i \in \{1, 2\} \)

\[
H^{(i)}_0 : Q_i(\rho) = 0 \text{ for some } \rho > 0
\]

\[
H^{(i)}_A : Q_i(\rho) > 0 \text{ for all } \rho > 0
\]

Empirically, estimation and testing in this paper is based 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 \( NT \) more generally. To test the null hypothesis we respectively define nonparametric estimators of \( \Psi_{jt}(\rho, t, z) \), denoted by \( \Psi^{(N)}_{jt}(\rho, t, z) \), empirical analogues of \( Q_i(\rho) \), denoted by

\[
Q^{(N)}_1(\rho) = \sum_{t=1}^{T} \sum_{z=1}^{Z} \sum_{s=1}^{2} \left\{ \min \left[0, \Psi^{(N)}_{st}(\rho, z)\right]^2 + \Psi^{(N)}_{s+2,t}(\rho, z)^2 \right\}
\]

(10)

\[
Q^{(N)}_2(\rho) = \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ \sum_{j=1,2,6,7} \min \left[0, \Psi^{(N)}_{jt}(\rho, z)\right]^2 + \Psi^{(N)}_{5t}(\rho, z)^2 + \left[\Psi^{(N)}_{6t}(\rho, z) \Psi^{(N)}_{7t}(\rho, z)\right]^2 \right\}
\]

(11)

and a confidence region for \( \Gamma_i \), defined as

\[
\Gamma^{(N)}_i \equiv \{ \rho > 0 : A_N Q^{(N)}_i(\rho) \leq c_\delta \}
\]

(12)

where \( A_N \) is the asymptotic rate of convergence of \( Q^{(N)}_i(\rho) \), and \( c_\delta \) is the \( \delta \) critical value of the test statistic. We reject the pure moral hazard model at level \( \delta \) if \( \Gamma^{(N)}_1 \) is empty, and interpret \( \Gamma^{(N)}_2 \) in a similar manner.

The estimated functions, \( \Psi^{(N)}_{jt}(\rho, z) \), are formed from estimates of the components to \( \Psi_{jt}(\rho, z) \). In the previous sections we have already described our estimates of the compensation scheme, \( w^{(N)}_s(x, z) \), the probability densities, \( f^{(N)}_s(x, z) \), and the probabilities, \( \varphi^{(N)}_s(z) \). From these estimated functions, we directly form the estimated weighted ratio \( h^{(N)}(x, z) \), as well as \( \Psi^{(N)}_{jt}(\rho, z) \) for \( j \in \{3, 4, 5, 7\} \) using the definitions of \( \Psi_{jt}(\rho, z) \) given in the previous section. However to estimate \( \Psi^{(N)}_{6t}(\rho, z) \) we require an estimate of \( \tau_{st}(\rho, z) = \exp \left[-\rho \bar{w}_s(z)/b_{t+1}\right] \). We use the fact that although \( \bar{w} \) is unknown, \( w_j(x) \) is a
locally non-decreasing function in \( x \). Following Brunk (1958), for each state \( s \in \{1, \ldots, S\} \), we rank the observations on returns in decreasing order by \( x_{s1}, x_{s2}, \ldots \) and so on, denoting by \( w_{s1}, w_{s2}, \ldots \) the corresponding compensations, and estimate \( w_{j}(\beta, z) \) with \( \overline{w}_{j}(z) \) defined as
\[
\overline{w}_{s}^{(N)} \equiv \max_{q} \sum_{r=1}^{q} \frac{w_{sr}}{q}
\]
Finally to estimate \( \Psi_{st}^{(N)}(\rho, z) \) for \( s \in \{1, 2\} \) we also require estimates of \( g_{s}(x, z) \), which we denote by \( g_{s}^{(N)}(\rho, x, z) \). Note from Proposition 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
\[
\begin{bmatrix} f_{2}(x, z) \\ f_{1}(x, z) \end{bmatrix} = \lim_{\Delta \to 0} \left[ \frac{F_{2}(x + \Delta, z) - F_{2}(x, z)}{F_{1}(x + \Delta, z) - F_{1}(x, z)} \right]
\]
it follows that
\[
\overline{h}(z) = \frac{\varphi_{2}(z)}{\varphi_{1}(z)} \left\{ \lim_{x \to \infty} \left[ \frac{f_{2}(x, z)}{f_{1}(x, z)} \right] \right\} = \frac{\varphi_{2}(z)}{\varphi_{1}(z)} \left\{ \lim_{x \to \infty} \left[ \frac{1 - F_{2}(x, z)}{1 - F_{1}(x, z)} \right] \right\}
\]
Again, following Brunk (1958), we estimated \( \overline{h}(z) \) with
\[
\overline{h}_{s}^{(N)}(z) \equiv \max_{q} \sum_{r=1}^{q} \left[ \frac{\overline{w}_{s}^{(N)}(z) \cdot \overline{f}_{s}^{(N)}(z)}{\overline{w}_{s}^{(N)}(z) + \overline{f}_{s}^{(N)}(z)} \right] \left( \sum_{n=1}^{N} \frac{1}{n} \left[ x_{n} \geq q, s_{n} = 2 \right] \right)
\]
Under standard regularity conditions \( \Psi_{jt}^{(N)}(\rho, z) \) converges in probability to \( \Psi_{jt}(\rho, z) \) for \( j \in \{3, 4, 5, 7\} \). Although \( w_{j}^{(N)}(x) \) and \( f_{j}^{(N)}(x) \) are estimated nonparametrically in the first stage and therefore converge at a slower rate than \( N^{\frac{1}{2}} \), appealing to results in Newey and MacFadden(1998), under standard regularity conditions
\[
\sqrt{N} \left[ \Psi_{st}^{(N)}(\rho, z) - \Psi_{st}(\rho, z) \right] \to N(0, \Omega_{j}(\rho)) \quad (13)
\]
for a given \( \rho > 0 \) and \( j \in \{3, 4, 5, 7\} \). Their results do not necessarily extend to \( \Psi_{2}^{(N)}(\rho) \) for \( j \in \{1, 2, 6\} \) because of \( \overline{w}_{j}^{(N)}(z) \) and \( \overline{h}_{s}^{(N)}(z) \). Here the choice of regularity condition on the upper bound \( \overline{\tau}_{s} \) plays a role. If there exist a finite \( \overline{\tau}_{s} \) such that \( \overline{F}_{s}(\overline{\tau}_{s}) < 1 \) and if \( x > \overline{\tau}_{s} \), then \( g_{s}(x) = 0 \). In that case the derivative of \( w_{s}(x) \) at \( \overline{\tau}_{s} \) is zero, following Parsons (1978) the norming constant is \( \sqrt{N} \), and hence
\[
\sqrt{N}(\Psi_{2}^{(N)}(\rho) - \Psi_{2}(\rho)) \to N(0, \Omega_{2}(\rho)) \quad (14)
\]
However, if we relax this assumption to the less restrictive assumption that \( \lim_{x \to \infty} g_{s}(x) = 0 \) then as Wright (1981) shows the norming constant will be \( N^{\frac{1}{2}} \) which implies that
\[
N^{\frac{1}{2}}(\Psi_{2}^{(N)}(\rho) - \Psi_{2}(\rho)) \to N(0, \Omega_{2}(\rho)) \quad (15)
\]
Although the assumption does not affect the estimation of the model or the identification results it does affects the rate of convergence of the estimates and the asymptotic covariance
matrix. Thus in our model $A_N = N^a$, where $a = 1$ if there exists a finite $\bar{x}_s$ such that $F_s(\bar{x}_s) < 1$ with $g_s(x) = 0$ for all $x > \bar{x}_s$, but where $a = 2/3$ under the weaker assumption that $\lim_{\bar{x} \to \infty} g_s(x) = 0$. If a finite $\bar{x}_s$ does not exist and $A_N = N^{2/3}$, then asymptotic covariance matrix is driven by the tails observations. If a finite $\bar{x}_s$ does exist and $A_N = N$, then all the observations help determine the asymptotic properties.

Assuming the following condition 8 holds, then by Lemma 3.1 of Chernozhukov, Tamer and Hong(2007), $c_d$ is the $\delta$–quantile of the distribution of $C$.

**Condition 8** For

$$C^{(N)}_i = \sup_{\rho \in \Gamma} \left[ N^a Q^{(N)}_i (\rho) \right]$$

(16)

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

Since $c_0$ is unknown, the test cannot be implemented as stated, but a 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.

**Algorithm 9 (Subsampling)** Consider all subsets of the data with size $N_b < N$, where $N_b \to \infty$ but $N_b/N \to 0$, and denote the number of subsets by $B_N$. Let

$$c_0 = \inf_{\rho > 0} \left[ N^a Q^{(N)}_i (\rho) \right] + \kappa_N$$

and

$$\Gamma^{(N)}_{0i} = \{ \rho > 0 : N^a Q^{(N)}_i (\gamma) \leq c_0 \}$$

where $\kappa_N \propto \ln N$. For each subset $j \in \{1, \ldots, B_N\}$ of size $N_b$ define

$$C^{(j,N_b)}_i \equiv \sup_{\rho \in \Gamma^{(N)}_{0i}} \left[ N^a Q^{(j,N_b)}_i (\rho) \right]$$

Denoting the $\delta$–quantile of the sample $\left\{ C^{(1,N_b)}_i, \ldots, C^{(B_N,N_b)}_i \right\}$ by $\widehat{\delta}$, let

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

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

We conducted this test separately for each sector, 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. Table 5 depicts the results. We did not reject the hybrid 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 \( \rho \in [0.037, 0.042] \), there is no evidence that managers with different attitudes towards risk are sorting into different sectors.

7 Private Information and Moral Hazard

There are many ways to characterize the impact of private information and hidden actions in models of management. In this paper we estimate the losses shareholders would incur from letting the manager tend his own interests instead of maximizing the expected value of the firm, how much managers would gain from tending their own interests instead of their firm’s. These two measures illustrate the gains from internalizing the conflicting goals of managers and shareholders. We also estimate how much shareholders would pay to make the private information public rather than induce revelation through the manager’s compensation schedule, and how much shareholders would pay to rid the firm of moral hazard problem altogether. These measures place lower bounds on what it would cost firms to rid itself of private information and hidden actions.

The first measure, denoted \( \tau_1 \), 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

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

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

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 \( \tau_2 \). 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

\[
\tau_2 = w_2^0 - w_1^0 = \frac{b_{t+1}}{\rho (b_t - 1)} \log \left( \frac{\alpha_2}{\alpha_1} \right)
\]

The net benefits from incentivizing the manager are therefore \( \tau_1 - \tau_2 \).

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 $\tau_3$, is the difference between expected compensation under the current optimal scheme and $w_0^2$, or:

$$
\tau_3 \equiv E[w_s(x)] - w_0^2 = E[w_s(x)] - \frac{b_{t+1}}{\rho (b_t - 1)} \log (\alpha_2/\alpha_0)
$$

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

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

where $w_{ms}(x)$ denotes the optimal compensation schedule for a pure moral hazard problem in state $s \in \{1, 2\}$. The solution to the pure moral hazard problem is found by setting $\eta_3^0 = \eta_4^0 = 0$ in the equations defining the pure moral hazard model to obtain

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

for $s \in \{1, 2\}$ where $\eta_{ms}^0$ uniquely solves (Equation at end of Section 4). It follows that

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

the final inequality following from the fact that whenever the truth telling and sincerity constraints are binding, they necessarily modify the optimal contract that would have been written in a pure moral hazard setup, thereby increasing the expected cost to shareholders.

Our analysis of identification and the tests showed there is a set of values $\Gamma$ for the risk aversion parameter such that its elements $\rho \in \Gamma$ are observationally equivalent. Since each element $\rho$ induces a value for the remaining parameters $\theta$ through the mapping $\theta(\rho)$, the set of the benefits and losses defined above could be estimated as a mapping from the risk aversion parameter, thus bounding them. Rather than pursue that approach, we minimized the sum of squared criterion function $Q^{(N)}(\rho)$ in $\rho$ to obtain a consistent estimator of one of the parameters in the identified set and generated a consistent estimator of its standard error using subsampling. The value obtained 0.0038 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 (2008), who applied a fully parametric estimator to data on U.S. managers from other industries over different time periods. For example our estimate of $\rho$ implies that a manager would pay $$$ to avoid an equal chance of losing versus gaining $100,000.

Our estimates of $\tau_1$ through $\tau_3$ depicted in Table 6, denoted $\hat{\tau}_1$ through $\hat{\tau}_3$ respectively, were computed from $\theta(\hat{\rho})$ and replacing population moments with their corresponding sample analogues. Concerning $\hat{\tau}_4$ we remark that $\eta_{mj}^0$ is not part of the solution for the generalized model of moral hazard, but a consistent estimate of it can be directly. We note first that from the compensation function and the definition of a logarithm

$$
1 + \eta_{ms}^0 (\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_{ms}^0 \left\{ \max_x [g_s(x)] \right\} \geq 1 + \eta_{ms}^0 (\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_{ms}^0 g_s(x) > 0
$$
for all \( x \) and \( s \in \{1, 2\} \). Combining the implied inequality for \( \eta_{ms}^0 \) with the Kuhn Tucker condition that \( \eta_{ms}^0 \) is positive yields the bounds

\[
0 < \eta_{ms}^0 < \left\{ \max_x [g_s(x)] - \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t - 1)} \right\}^{-1}
\]

Substituting our estimate for \( \rho \) into the mapping \( \theta(\rho) \) we take the sample analogue to the expectation and solve the one dimensional equation in \( \eta_{ms}^o \) to obtain a consistent estimate of that parameter

\[
\sum \left[ \frac{g_s(x) - \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t - 1)}}{1 + \eta_{ms}^o \left( \frac{\alpha_2}{\alpha_1} \right)^{1/(b_t - 1)} - \eta_{ms}^o g_s(x)} \right] = 0
\]

The estimated standard errors were computed by minimizing the criterion function \( Q_2^{(j, N_1)}(\rho) \) in \( \rho \) for each subsample \( j \in \{1, \ldots, B_N\} \) and computing the standard deviation of \( \hat{\rho}_j \). Aside from its computational tractability in our application, another advantage of this approach, compared to derivative based methods, is that it frees us from taking a stand on the regularity conditions concerning \( g_s(x) \) that determine the rate of convergence in order to construct confidence intervals and hypotheses tests; the procedure is robust to whether \( N^{1/2} \) or \( N^{1/3} \) applies.

Our estimate of \( \tau_1 \) is the sample analogue of the second line

Table 6 shows the estimated annual expected losses per year by firm type as a percentage of firm value along with the estimated standard errors. Except for 3 estimates out of the 24, they range from just under 1 percent of the firm’s value to just under 5 percent, all with standard errors of a lower order magnitude. Again the estimates are similar to those found by Margiotta and Miller (2000) and Gayle and Miller (2008). At the interest rates which applied to this period our estimates imply the summed discounted loss amounts for shirking lie between XXX and YYY of the total shareholder value.

Table 6 reports our estimates of \( \tau_2 \) for the 24 subsectors. They range from less than $1 million to about $15 million per year, with about half of the estimates being significantly different from 0.

Table 7 reports the estimated differences along with consistent estimates of their standard errors. They show that a large fraction of the expected compensation to managers is the risk premium they receive to offset the uncertain compensation. Eliminating private information would also reduce the cost of motivating the manager, but not by as much as installing a perfect monitor.

Comparing our estimates of \( E[w_{ms}(x)] \) with those for \( E[w_s(x)] \) in Table 7 we see wrongs signs in 6 of the 24 subsectors, while comparing our estimates of \( E[w_{ms}(x)] \) with those for \( w_2 \), we obtain wrong signs for 4. Overall these estimates suggest the existence of private information increases about ZZZZ of the extra cost suggest to over what is attributable to pure moral hazard alone.
8 Appendix

Proof of Lemma 1. Let \( \lambda_r \) is the date \( t \) price of a contingent claim made on a consumption unit at date \( r \), implying the bond price is defined as

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

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

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

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

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

the corresponding value from choosing effort level \( l_{t1} \) is

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

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

\[
-b_t \alpha_0 \exp \left( -\frac{q_t + \rho c_{nt}}{b_{t+1}} \right)
\]

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

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

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

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

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

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

\[
(\alpha_0/\alpha_j)_{1/(b_t-1)} l_{t0} + E_t \left[ \exp \left( -\frac{\rho w(s_t', x_{t+1})}{b_{t+1}} \right) [g_s(x_{t+1}) l_{t1} + l_{t2}] \right]
\]
Proof on Lemma 2. Differentiating the objective function in the text, the first order equations are

\[
\begin{align*}
v_{1t}(x)^{-1} &= \eta_0 + \eta_1 \left[ \left( \alpha_2 / \alpha_1 \right)^{(b_1 - 1)} - g_1(x) \right] - \eta_3 h(x) - \eta_4 \left( \alpha_1 / \alpha_2 \right)^{(b_1 - 1)} g_2(x) h(x) \\
v_{2t}(x)^{-1} &= \eta_0 + \eta_2 \left[ \left( \alpha_2 / \alpha_1 \right)^{(b_1 - 1)} - g_2(x) \right] + \eta_3 + \eta_4
\end{align*}
\]

Multiplying each equation by \( v_{st}(x) f_s(x) \), then summing and integrating over \( x \) yields

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

where we make use of the complementary slackness conditions. Substituting for \( \eta_0 = \{ E[v_{st}(x)] \}^{-1} \) gives the equations in the lemmas.

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 \( s \in \{ t + 1, t + 2, \ldots, T - 1 \} \)

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

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 1.

Proof of Lemma 3. Without loss of generality we suppress the dependence of \( w_{nt} \) on \( (z_{nt}, s_{nt}, b_t) \), let \( \tilde{x} \) denote net abnormal return, and let \( V \) denote the value of the firm at the beginning of the period. For any net abnormal return \( \tilde{x} \), and for any value of the firm at the beginning of the period \( V \), we denote by

\[
w = \Lambda \left( \tilde{x} + \frac{w}{V} \right) \quad \text{(w relation)}
\]

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

\[
w(x) = \Lambda \left( \tilde{x} + \frac{w(x)}{V} \right)
\]

for all \( x \) and

\[
\tilde{x} = x - w(x) / V
\]

28
We remark that \( w(x) \) is one solution to the relation defined by \( \Lambda \). Suppose that for some pair \((\bar{x}, V)\) 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

\[
w_i = w(x_i) = \Lambda \left( \bar{x} + \frac{w_i}{V} \right)\]

for \( i \in \{1, 2\} \). From the definition of \( \bar{x} \) we obtain

\[
\bar{x} = x_i - w(x_i) / V
\]

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 \( \Lambda \) for each pair \((\bar{x}, V)\) 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 \( \Lambda_1(\bar{x}, V) \). Having proved \( w(x) = \Lambda_1(\bar{x}, V) \), the lemma now follows because the measurement error on compensation is assumed independent of \((\bar{x}, V)\) so \( E[\tilde{w} | \bar{x}, V] = \Lambda_1(\bar{x}, V) \). ■

The following lemma is used in the proof of Proposition 1.

**Lemma 10** In the optimal contract

\[
\eta_3 + \eta_4 = \{E_2[v_2(x)]\}^{-1} - \{E[v_{st}(x)]\}^{-1}
\]

**Proof of Lemma.** Multiply the first order conditions for the second state by \( v_{2t}(x) \), after solving for \( \eta_0 \) to obtain

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

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

\[
1 = \{E[v_{st}(x)]\}^{-1} E_2[v_2(x)] + \eta_3 E_2[v_{2t}(x)] + \eta_4 E_2[v_{2t}(x)]
\]

\[
= E_2[v_2(x)] \left( \{E[v_{st}(x)]\}^{-1} + \eta_3 + \eta_4 \right)
\]

Dividing through by \( E_2[v_2(x)] \) proves the lemma. ■

**Proof of Proposition 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 \( v_{st} \equiv \exp[-\rho^* \bar{w}_s / b_{t+1}] \), and prove the proposition by successively treating each component of \( \theta^* \).

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 $\eta_0$ into the first order condition for the second state, which yields

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

Taking expectations we obtain

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

Also

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

Differenting 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 differenting

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

from

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

to give

$$\overline{v}_{2t}^{-1} - v_{2t}(x)^{-1} = \eta_2g_2(x)$$

Upon rearrangement, we appeal to the result in Item 1, that $\eta_{32} = \eta_{32t}(\rho^*)$ to obtain

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

4. To show $\alpha_1 = \alpha_1(\rho^*)$ we substitute the solution for $\eta_2$ above into the first order condition for the second state evaluated at the limit $x \to \infty$ to obtain

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

or, upon appealing to Lemma 4

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

Making $\alpha_1$ the subject of the equation

$$\alpha_1 = \alpha_2 \left[ \frac{\overline{v}_{2t}^{-1} - \left\{ E [v_{2t}(x)] \right\}^{-1}}{\overline{v}_{2t}^{-1} - E_2 \left[v_{2t}(x)^{-1} \right]} \right]^{1-b_t} = \alpha_1(\rho^*)$$
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 \( \eta_0 \) to obtain

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

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 the lemma reduces this expression to

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

Upon collecting terms

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

so solving for \( \eta_4 \) we now have

\[
\eta_4 = \frac{\{ E [v_{st} (x)] \}^{-1} E_1 [v_{1t} (x)] - E_1 [v_{1t} (x) h(x)] \left[ \{ E_2 [v_{2t} (x)] \}^{-1} - \{ E [v_{st} (x)] \}^{-1} \right] - 1}{(\alpha_1 / \alpha_2)^{1/(b_t - 1)} E_1 [v_{1t} (x) g_2 (x) h (x)] - E_1 [v_{1t} (x) h(x)]}
\]

\[
= \eta_{4t} (\rho^*)
\]

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 \( x \to \infty \) we have

\[
\eta_1 (\alpha_2 / \alpha_1)^{1/(b_t - 1)} = v_{1t}^{-1} - \{ E [v_{st} (x)] \}^{-1} + \eta_3 \bar{h}
\]

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

8. Differentiating the first order condition for the first state and its limit as \( x \to \infty \) gives

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

Dividing both sides by \( \eta_1 \) we thus establish \( g_1 (x) = g_{1t} (x, \rho^*) \) for all \( t \).
Proof of Proposition 2. We show the sincerity constraint implies $\Psi_{6t}(\rho) \geq 0$. 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

$$0 \leq E_2 \left[ (\alpha_{1t}(\rho)/\alpha_{2t}(\rho))^{1/(b_t-1)} v_{1t}(x, \rho) g_2(x, \rho) - v_{2t}(x, \rho) \right]$$

$$= E_2 \left[ \frac{1 - \bar{\nu}_{st}(\rho)}{1 - \bar{\nu}_{st}(\rho)} \frac{E_2 [v_{2t}(x, \rho)]^{-1}}{E_2 [v_{2t}(x, \rho)]^{-1}} \right] \left[ \frac{1 - \bar{\nu}_{2t}(\rho)/v_{2t}(x, \rho)}{1 - \bar{\nu}_{2t}(\rho)/v_{2t}(x, \rho)} \right] v_{1t}(x, \rho) - v_{2t}(x, \rho)$$

$$\equiv \Psi_{6t}(\rho)$$

Proof of Proposition 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 Jensen's inequality

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

so

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

and consequently the quotient

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

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

2. For $\rho \in \Gamma_2$ the participation constraint implies

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

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

$$< (1 - b_t) b_{t+1} \left\{ \sum_{s=1}^{2} \varphi_s E_s [-\rho w_s(x)/b_{t+1}] \right\}$$

$$= -\rho (1 - b_t) \left\{ \sum_{s=1}^{2} \varphi_s E_s [w_s(x)] \right\}$$

$$= \rho (b_t - 1) \left\{ \sum_{s=1}^{2} \varphi_s E_s [w_s(x)] \right\}$$

The proof for $\rho \in \Gamma_1$ is found by replacing $\sum_{s=1}^{2} \varphi_s E_s [\exp [-\rho w_s(x)/b_{t+1}]]$ with $E_s [\exp [-\rho w_s(x)/b_{t+1}]]$ for $s \in \{1, 2\}$. 

32
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 $E_s[g_{st}(x, \rho)] = 1$.

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 $v_2(x) \geq \bar{v}_2$ for all $x \in R$, so from the definition of $v_2(x)$ we infer

$$\bar{v}_2 \equiv \lim_{x \to \infty} [w_2(x)] = \sup_{x \in R} [w_2(x)]$$

Hence $v_{2t}(x, \rho) \geq \bar{v}_{2t}(\rho)$ 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 $\rho \in \Gamma_1$. Finally the inequality $\Psi_{st}(\rho) \geq 0$ guarantees $g_{1t}(x, \rho) \geq 0$ for $\rho \in \Gamma_2$.

5. In the proof of Item 2 of Proposition 1 we established $\eta_{2} \equiv \bar{v}_{2t}^{-1} - E_2[v_{2t}(x)^{-1}]$, and also proved in Item 2 above that $\bar{v}_{2t} \leq v_{2t}(x)$. Hence $\eta_2 \geq 0$. The same arguments apply to $\eta_{1t}(\rho)$ when $\rho \in \Gamma_1$ thus establishing $\eta_{1t}(\rho) \geq 0$ in that case.

**Proof of Proposition 4.** We show that any compensation schedule from a regular data generating process satisfying the restrictions implied by some $\hat{\rho} \in \Gamma_i$ for $i \in \{1, 2\}$ solves the Kuhn Tucker formulation of an optimal contracting problem for a model which we abbreviate by the parameterization $\left(\hat{\rho}, \hat{\theta}\right)$ where

$$\hat{\theta} \equiv (\hat{\alpha}_1, \hat{\alpha}_2, \hat{g}_1(x), \hat{g}_2(x))$$

$$\equiv (\alpha_{1t}(\hat{\rho}), \alpha_{2t}(\hat{\rho}), g_{1t}(x, \hat{\rho}), g_{2t}(x, \hat{\rho}))$$

$$\equiv \theta_{i}(\hat{\rho})$$

Similarly we abbreviate the mappings $\eta_{jt}(\hat{\rho})$ defined in Proposition 1 by $\hat{\eta}_j$. 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 $w^*(x)$ satisfies the first order condition and the complementary slackness conditions for the parameterization $\left(\hat{\rho}, \hat{\theta}\right)$ with multipliers $\hat{\eta} \equiv (\hat{\eta}_1, \hat{\eta}_2, \hat{\eta}_3, \hat{\eta}_4)$. Consequently the compensation schedule for $\left(\hat{\rho}, \hat{\theta}\right)$ is $w^*(x)$. It is convenient to treat the pure and hybrid moral hazard models separately.

We first consider models of pure moral hazard and suppose $\hat{\rho} \in \Gamma_1$. We show that the participation and incentive compatibility constraints are met with equality, that the first order condition is satisfied and that $\hat{\eta}_j > 0$ solve the equations defining the Kuhn Tucker multipliers.

1. Since $\Psi_{3t}(\hat{\rho}) = \Psi_{3t}(\hat{\rho}) = 0$, the competitive selection constraints ensure

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

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

Multiplying both sides by
\[
\hat{\eta}_s \equiv \frac{1}{v_{st} (\hat{\rho})} - E_s \left[ v_{st} (x, \hat{\rho}) \right]^{-1}
\]
we obtain the first order condition for the state \( s \) as
\[
\hat{\eta}_s \left[ \hat{g}_s (x) - (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)} \right] = \left[ \frac{1}{v_{st} (\hat{\rho})} - v_{st} (x, \hat{\rho}) \right] - \left[ \frac{1}{v_{st} (\hat{\rho})} - E_s \left[ v_{st} (x, \hat{\rho}) \right]^{-1} \right]
\]
\[
= \{ E_s [v_{st} (x, \hat{\rho})] \}^{-1} - v_{st} (x, \hat{\rho})^{-1}
\]

3. The equation above implies
\[
\left[ \hat{g}_s (x) - (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)} \right] = \frac{1}{\hat{\eta}_s} \left\{ E_s [v_{st} (x, \hat{\rho})] \right\}^{-1} - v_{st} (x, \hat{\rho})^{-1}
\]

Multiplying through by \( v_{st} (x, \hat{\rho}) \) and taking the expectation with respect to \( x \) conditional on the \( s^{th} \) state yields
\[
E_2 \left\{ \left[ \hat{g}_s (x) - (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)} \right] v_{st} (x, \hat{\rho}) \right\} = \frac{1}{\hat{\eta}_s} E_s \left\{ \left\{ E_s [v_{st} (x, \hat{\rho})] \right\}^{-1} - v_{st} (x, \hat{\rho})^{-1} \right\} v_{st} (x, \hat{\rho})
\]
\[
= \frac{1}{\hat{\eta}_s} E_s \left\{ v_{st} (x, \hat{\rho}) / E_s [v_{st} (x, \hat{\rho})] - 1 \right\}
\]
\[
= 0
\]

Since \( \hat{\eta}_s > 0 \) the incentive compatibility condition is satisfied with equality.

4. Since the first order condition is satisfied by \( \hat{\eta}_s \) for each \( x \) we may write
\[
v_{st} (x, \hat{\rho})^{-1} = \hat{\eta}_s \left[ \hat{g}_s (x) - (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)} \right] - E_s [v_{st} (x, \hat{\rho})]^{-1}
\]

and substitute \( v_{st} (x, \hat{\rho})^{-1} \) into the expression
\[
E_2 \left\{ \frac{\hat{g}_s (x) - (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)}}{\hat{\eta}_s \hat{g}_s (x) - \hat{\eta}_s (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)} - E_s [v_{st} (x, \hat{\rho})]^{-1}} \right\}
\]
to obtain
\[
E_2 \left\{ v_{st} (x, \hat{\rho}) \left[ \hat{g}_s (x) - (\hat{\alpha}_2 / \hat{\alpha}_1)^{1/(b_t-1)} \right] \right\} = 0
\]
the equality following from the incentive compatibility constraint, which we proved in Item 3 above is satisfied by \( \hat{\eta}_3 \). Consequently \( \hat{\eta}_s \) is a solution to the equation

\[
E_2 \left\{ \frac{\hat{g}_s (x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_1-1)}}{\eta \hat{g}_s (x) - \eta (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_1-1)} - \{E_s [v_{st} (x, \hat{\rho})]\}^{-1}} \right\} = 0
\]

in \( \eta \), which define the Kuhn Tucker multipliers. This completes the proof for the pure moral hazard case.

We now consider the hybrid case, and suppose \( \hat{\rho} \in \Gamma_2 \). We show that the participation and incentive compatibility constraints are met with equality, that the first order conditions hold, that the truth telling and sincerity constraints are satisfied, and that \( \hat{\eta} \) solves the equations defining the Kuhn Tucker multipliers.

1. The definition of \( \hat{\alpha}_2 \equiv \{E [v_{st} (x, \hat{\rho})]\}^{-1-\beta} \) directly implies the participation constraint is met with equality. The truth telling and sincerity constraints are directly imposed from \( \hat{\rho} \) belonging to \( \Gamma_2 \) through the equality \( \Psi_{st} (\hat{\rho}) = 0 \) and the inequality \( \Psi_{st} (\hat{\rho}) \geq 0 \). This only leaves the three tasks of establishing \( (\hat{\rho}, \hat{\theta}) \) satisfies the first order conditions, that the incentive compatibility conditions are met with equality, and that \( \hat{\eta} \) solves the equations defining the Kuhn Tucker multipliers.

2. Noting the definitions of \( \hat{\alpha}_1, \hat{\alpha}_2, \hat{g}_2 (x) \) and \( \hat{\eta}_2 \) are identical to their counterparts in the pure moral hazard model we can appeal to Item 2 in the moral hazard case to establish

\[
\hat{\eta}_2 \left[ \hat{g}_2 (x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_1-1)} \right] = \{E_2 [v_{2t} (x, \hat{\rho})]\}^{-1} - v_{st} (x, \hat{\rho})^{-1}
\]

From the definition of \( \hat{\eta}_3 \) we have

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

Subtracting the first equation from the second

\[
\{E [v_{st} (x, \hat{\rho})]\}^{-1} + \hat{\eta}_3 + \hat{\eta}_4 \left[ (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_1-1)} - \hat{g}_2 (x) \right] + \hat{\eta}_4 = v_{st} (x, \hat{\rho})^{-1}
\]

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

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

\[
\hat{\eta}_1 \hat{g}_1 (x) = \nu_{1t} (\hat{\rho})^{-1} - v_{1t} (x, \hat{\rho})^{-1} + \hat{\eta}_3 \left[ \bar{h} - h (x) \right] - \hat{\eta}_4 (\hat{\alpha}_1/\hat{\alpha}_2)^{1/(b_1-1)} \hat{g}_2 (x) h (x)
\]

From the definition of \( \hat{\eta}_1 \)

\[
\eta_3 \bar{h} = \hat{\eta}_1 (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_1-1)} - \{E [v_{st} (x, \hat{\rho})]\}^{-1} - \nu_{1t} (\hat{\rho})^{-1}
\]

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

\[
v_{1t} (x, \hat{\rho})^{-1} = \{E [v_{st} (x, \hat{\rho})]\}^{-1} + \hat{\eta}_1 \left[ (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_1-1)} - \hat{g}_1 (x) \right] - \hat{\eta}_3 h (x) - \hat{\eta}_4 (\hat{\alpha}_1/\hat{\alpha}_2)^{1/(b_1-1)} \hat{g}_2 (x) h (x)
\]
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 \( \hat{\alpha}_1, \hat{\alpha}_2, \hat{g}_2(x) \) and \( \hat{\eta}_2 \) are identical to their counterparts in the pure moral hazard model, which implies from Item 2 in the moral hazard case that

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

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

\[
E_2 \left\{ \hat{\eta}_2 v_{2t}(x, \hat{\rho}) \left[ \hat{g}_2(x) - \frac{\hat{\alpha}_2}{\hat{\alpha}_1} \right]^{1/(b_t-1)} \right\} = E_2 \left\{ \hat{\eta}_2 v_{2t}(x, \hat{\rho}) \left\{ \{ E_2 [v_{2t}(x, \hat{\rho})] \}^{-1} - v_{2t}(x, \hat{\rho})^{-1} \right\} \right\} = 0
\]

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

\[
\hat{\eta}_1 E_1 \left\{ v_{1t}(x, \hat{\rho}) \left[ \hat{g}_1(x) - \frac{\hat{\alpha}_2}{\hat{\alpha}_1} \right]^{1/(b_t-1)} \right\} = E_1 \left\{ v_{1t}(x, \hat{\rho}) \{ E [v_{st}(x, \hat{\rho})] \}^{-1} - \hat{\eta}_3 E_1 [v_{1t}(x, \hat{\rho}) h(x)] \right. - \hat{\eta}_4 \left( \frac{\hat{\alpha}_1}{\hat{\alpha}_2} \right)^{1/(b_t-1)} E_1 [v_{1t}(x, \hat{\rho}) \hat{g}_2(x) h(x)] - 1
\]

Successively substituting the definitions of \( \hat{\eta}_3 \) and \( \hat{\eta}_4 \) into the right side of the equation

\[
E_1 \left\{ v_{1t}(x, \hat{\rho}) \{ E [v_{st}(x, \hat{\rho})] \}^{-1} - \hat{\eta}_3 E_1 [v_{1t}(x, \hat{\rho}) h(x)] \right. - \hat{\eta}_4 \left( \frac{\hat{\alpha}_1}{\hat{\alpha}_2} \right)^{1/(b_t-1)} E_1 [v_{1t}(x, \hat{\rho}) \hat{g}_2(x) h(x)] - 1
\]

\[
= E_1 \left\{ v_{1t}(x, \hat{\rho}) \{ E [v_{st}(x, \hat{\rho})] \}^{-1} - \left( \{ E_2 [v_{2t}(x, \hat{\rho})] \}^{-1} - \{ E [v_{st}(x, \hat{\rho})] \}^{-1} - \hat{\eta}_4 \right) E_1 [v_{1t}(x, \hat{\rho}) h(x)] \right. - \hat{\eta}_4 \left( \frac{\hat{\alpha}_1}{\hat{\alpha}_2} \right)^{1/(b_t-1)} E_1 [v_{1t}(x, \hat{\rho}) \hat{g}_2(x) h(x)] - 1
\]

\[
= E_1 \left\{ v_{1t}(x, \hat{\rho}) \{ E [v_{st}(x, \hat{\rho})] \}^{-1} - \left( \{ E_2 [v_{2t}(x, \hat{\rho})] \}^{-1} - \{ E [v_{st}(x, \hat{\rho})] \}^{-1} \right) E_1 [v_{1t}(x, \hat{\rho}) h(x)] - 1
\]

\[
\left. + \hat{\eta}_4 \left\{ E_1 [v_{1t}(x, \hat{\rho}) h(x)] - \left( \frac{\hat{\alpha}_1}{\hat{\alpha}_2} \right)^{1/(b_t-1)} E_1 [v_{1t}(x, \hat{\rho}) \hat{g}_2(x) h(x)] \right\} \right\}
\]

\[
= 0
\]

Therefore

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

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 \( \hat{v}_{st}^{-1}(x, \hat{\rho}), \hat{g}_s(x) \) and \( \hat{\alpha}_s \) for \( s \in \{1, 2\} \) and the defined elements \( \hat{\eta}_2 \) through \( \hat{\eta}_4 \) that

\[
\hat{v}_{1t}(x, \hat{\rho})^{-1} = \left\{ E [v_{st}(x, \hat{\rho})] \right\}^{-1} + \hat{\eta}_1 \left[ \left( \frac{\hat{\alpha}_2}{\hat{\alpha}_1} \right)^{1/(b_t-1)} - \hat{g}_1(x) \right] - \hat{\eta}_3 h(x) - \hat{\eta}_4 \left( \frac{\hat{\alpha}_2}{\hat{\alpha}_1} \right)^{1/(b_t-1)} \hat{g}_2(x) h(x)
\]

36
and 
\[ \hat{v}_{2t}^{-1}(x, \hat{\rho}) = \{ E[v_{st}(x, \hat{\rho})] \}^{-1} + \hat{\eta}_2[(\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} - \hat{g}_2(x)] + \hat{\eta}_3 + \hat{\eta}_4 \]

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

\[
0 = E_2 [\hat{v}_{1t}(x, \hat{\rho}) - \hat{v}_{1t}(x, \hat{\rho})] \\
= E_2 \left\{ \{ E[v_{st}(x, \hat{\rho})] \}^{-1} + \hat{\eta}_1 \left[ (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} - \hat{g}_1(x) \right] \right\}^{-1} \\
- E_2 \left\{ \{ E[v_{st}(x, \hat{\rho})] \}^{-1} + \hat{\eta}_2 \left[ (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} - \hat{g}_2(x) \right] + \hat{\eta}_3 + \hat{\eta}_4 \right\}^{-1}
\]

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

\[
0 = E_1 \left\{ v_{1t}(x, \hat{\rho}) \left[ \hat{g}_1(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] \right\} \\
= E_1 \left[ \frac{\hat{g}_1(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)}}{\{ E[v_{st}(x, \hat{\rho})] \}^{-1} + \hat{\eta}_1 \left[ (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} - \hat{g}_1(x) \right] - \hat{\eta}_3 h(x) - \hat{\eta}_4 (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \hat{g}_2(x) h(x)} \right]
\]

and

\[
0 = E_2 \left\{ \hat{v}_{2t}^{-1}(x, \hat{\rho}) \left[ \hat{g}_2(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} \right] \right\} \\
= E_2 \left[ \frac{\hat{g}_2(x) - (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)}}{\{ E[v_{st}(x, \hat{\rho})] \}^{-1} + \hat{\eta}_2 \left[ (\hat{\alpha}_2/\hat{\alpha}_1)^{1/(b_t-1)} - \hat{g}_2(x) \right] + \hat{\eta}_3 + \hat{\eta}_4} \right]
\]

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

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

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


Table 1: Time Series Summary

<table>
<thead>
<tr>
<th>Year</th>
<th>Bond</th>
<th>Assets</th>
<th>Employees</th>
<th>Debt/Equity</th>
<th>$x_t$</th>
<th>$\frac{C_t}{C_{t-1}}$</th>
<th>Compensation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>15.9</td>
<td>8896</td>
<td>18.02</td>
<td>2.83</td>
<td>1.19</td>
<td>1.18</td>
<td>1854</td>
<td>1574</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(26269)</td>
<td>(46.15)</td>
<td>(7.24)</td>
<td>(0.45)</td>
<td>(0.51)</td>
<td>(12412)</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>13.72</td>
<td>7770</td>
<td>16.18</td>
<td>2.87</td>
<td>0.97</td>
<td>1.07</td>
<td>2714</td>
<td>1876</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(25284)</td>
<td>(43.41)</td>
<td>(5.04)</td>
<td>(0.29)</td>
<td>(2.52)</td>
<td>(10909)</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>14.00</td>
<td>8187</td>
<td>16.43</td>
<td>3.45</td>
<td>1.26</td>
<td>1.18</td>
<td>1781</td>
<td>1867</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(28650)</td>
<td>(44.41)</td>
<td>(33.4)</td>
<td>(0.47)</td>
<td>(0.64)</td>
<td>(13252)</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>13.79</td>
<td>8357</td>
<td>17.31</td>
<td>2.41</td>
<td>1.16</td>
<td>1.17</td>
<td>3257</td>
<td>1926</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(29029)</td>
<td>(45.92)</td>
<td>(17.2)</td>
<td>(0.38)</td>
<td>(0.87)</td>
<td>(14824)</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>13.67</td>
<td>8770</td>
<td>17.94</td>
<td>2.76</td>
<td>1.30</td>
<td>1.22</td>
<td>4691</td>
<td>1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(31797)</td>
<td>(47.96)</td>
<td>(41.4)</td>
<td>(0.48)</td>
<td>(3.06)</td>
<td>(17791)</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>15.00</td>
<td>9486</td>
<td>17.67</td>
<td>3.91</td>
<td>1.05</td>
<td>1.20</td>
<td>2726</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(40415)</td>
<td>(45.91)</td>
<td>(71.3)</td>
<td>(0.53)</td>
<td>(1.11)</td>
<td>(18530)</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>13.97</td>
<td>10303</td>
<td>18.34</td>
<td>2.84</td>
<td>1.14</td>
<td>1.31</td>
<td>1652</td>
<td>1970</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(43087)</td>
<td>(45.75)</td>
<td>(11.57)</td>
<td>(0.76)</td>
<td>(8.27)</td>
<td>(21631)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>13.18</td>
<td>10484</td>
<td>19.59</td>
<td>2.64</td>
<td>1.14</td>
<td>1.18</td>
<td>4624</td>
<td>1865</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45936)</td>
<td>(54.08)</td>
<td>(8.31)</td>
<td>(0.68)</td>
<td>(1.5)</td>
<td>(21641)</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>14.16</td>
<td>12015</td>
<td>20.10</td>
<td>2.69</td>
<td>1.08</td>
<td>1.17</td>
<td>3314</td>
<td>1851</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(52064)</td>
<td>(56.50)</td>
<td>(14.9)</td>
<td>(0.54)</td>
<td>(1.86)</td>
<td>(18842)</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>14.32</td>
<td>12115</td>
<td>19.47</td>
<td>4.69</td>
<td>0.86</td>
<td>0.996</td>
<td>3165</td>
<td>1877</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(57166)</td>
<td>(54.51)</td>
<td>(105)</td>
<td>(0.42)</td>
<td>(2.43)</td>
<td>(16077)</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>14.87</td>
<td>13869</td>
<td>19.15</td>
<td>2.51</td>
<td>1.45</td>
<td>1.53</td>
<td>3151</td>
<td>1814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(66331)</td>
<td>(52.85)</td>
<td>(35.2)</td>
<td>(0.64)</td>
<td>(16.1)</td>
<td>(18830)</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>14.17</td>
<td>14429</td>
<td>21.05</td>
<td>2.77</td>
<td>1.16</td>
<td>1.11</td>
<td>4069</td>
<td>1687</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(70812)</td>
<td>(64.83)</td>
<td>(9.39)</td>
<td>(0.37)</td>
<td>(1.38)</td>
<td>(17195)</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>13.89</td>
<td>20925</td>
<td>22.19</td>
<td>2.63</td>
<td>1.07</td>
<td>1.16</td>
<td>4397</td>
<td>751</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(89832)</td>
<td>(52.34)</td>
<td>(12.27)</td>
<td>(0.36)</td>
<td>(1.63)</td>
<td>(19992)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Cross Section Summary  

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>8,980</td>
<td>6,762</td>
<td>11,144</td>
</tr>
<tr>
<td>Assets</td>
<td>6,322</td>
<td>5,277</td>
<td>17,776</td>
</tr>
<tr>
<td></td>
<td>(27773)</td>
<td>(22124)</td>
<td>(67133)</td>
</tr>
<tr>
<td>Employees</td>
<td>15.8</td>
<td>32.23</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>(40.8)</td>
<td>(78.75)</td>
<td>(26.59)</td>
</tr>
<tr>
<td>Debt/Equity</td>
<td>2.07</td>
<td>1.94</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>(40.9)</td>
<td>(26.21)</td>
<td>(50.63)</td>
</tr>
<tr>
<td>(\frac{CH_t}{CH_{t-1}})</td>
<td>1.15</td>
<td>1.13</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>(4.54)</td>
<td>(1.68)</td>
<td>(7.26)</td>
</tr>
</tbody>
</table>
## Table 3: Cross Section Summary
(Compensation in thousand of 2000 $US)

<table>
<thead>
<tr>
<th>$S_{it}$</th>
<th>Variable</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>(S,S,S)</td>
<td>1</td>
<td>0.11</td>
<td>-0.097</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.56)</td>
<td>(0.43)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>(S,L,S)</td>
<td>2</td>
<td>-0.03</td>
<td>-0.14</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.40)</td>
<td>(0.34)</td>
<td>(0.31)</td>
</tr>
<tr>
<td>(S,L,L)</td>
<td>3</td>
<td>-0.03</td>
<td>-0.11</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.36)</td>
<td>(0.37)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>(S,S,L)</td>
<td>4</td>
<td>0.07</td>
<td>-0.11</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.52)</td>
<td>(0.44)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>(L,S,S)</td>
<td>5</td>
<td>-0.005</td>
<td>-0.10</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.34)</td>
<td>(0.39)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>(L,L,S)</td>
<td>6</td>
<td>-0.07</td>
<td>-0.11</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.29)</td>
<td>(0.33)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>(L,L,L)</td>
<td>7</td>
<td>-0.03</td>
<td>-0.13</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td>(0.30)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>(L,S,L)</td>
<td>8</td>
<td>0.02</td>
<td>-0.13</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.30)</td>
<td>(0.40)</td>
<td>(0.47)</td>
</tr>
</tbody>
</table>

### Compensation

<table>
<thead>
<tr>
<th>$S_{it}$</th>
<th>Variable</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>(S,S,S)</td>
<td>1</td>
<td>3889</td>
<td>670</td>
<td>3397</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14651)</td>
<td>(10779)</td>
<td>(19178)</td>
</tr>
<tr>
<td>(S,L,S)</td>
<td>2</td>
<td>4384</td>
<td>2339</td>
<td>4922</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9381)</td>
<td>(14243)</td>
<td>(30677)</td>
</tr>
<tr>
<td>(S,L,L)</td>
<td>3</td>
<td>3742</td>
<td>521</td>
<td>9194</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(11903)</td>
<td>(15710)</td>
<td>(19898)</td>
</tr>
<tr>
<td>(S,S,L)</td>
<td>4</td>
<td>2522</td>
<td>721</td>
<td>3977</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9855)</td>
<td>(8851)</td>
<td>(14844)</td>
</tr>
<tr>
<td>(L,S,S)</td>
<td>5</td>
<td>3079</td>
<td>-850</td>
<td>4235</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20381)</td>
<td>(15773)</td>
<td>(20107)</td>
</tr>
<tr>
<td>(L,L,S)</td>
<td>6</td>
<td>4154</td>
<td>2422</td>
<td>4727</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(13375)</td>
<td>(16220)</td>
<td>(20989)</td>
</tr>
<tr>
<td>(L,L,L)</td>
<td>7</td>
<td>5781</td>
<td>2200</td>
<td>6897</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12807)</td>
<td>(12208)</td>
<td>(19288)</td>
</tr>
<tr>
<td>(L,S,L)</td>
<td>8</td>
<td>4396</td>
<td>-3729</td>
<td>4742</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14831)</td>
<td>(18890)</td>
<td>(19288)</td>
</tr>
<tr>
<td>$S_{1t}$</td>
<td>Sector1</td>
<td>Sector2</td>
<td>Sector 3</td>
<td>Total</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Good</td>
<td>Bad</td>
<td>N</td>
</tr>
<tr>
<td>(S,S,S)</td>
<td>1</td>
<td>2598</td>
<td>0.0917</td>
<td>0.1975</td>
</tr>
<tr>
<td>(S,L,S)</td>
<td>2</td>
<td>319</td>
<td>0.0141</td>
<td>0.0214</td>
</tr>
<tr>
<td>(S,L,L)</td>
<td>3</td>
<td>469</td>
<td>0.0257</td>
<td>0.0266</td>
</tr>
<tr>
<td>(S,S,L)</td>
<td>4</td>
<td>1326</td>
<td>0.0763</td>
<td>0.0713</td>
</tr>
<tr>
<td>(L,S,S)</td>
<td>5</td>
<td>541</td>
<td>0.0272</td>
<td>0.0331</td>
</tr>
<tr>
<td>(L,L,S)</td>
<td>6</td>
<td>1105</td>
<td>0.0635</td>
<td>0.0595</td>
</tr>
<tr>
<td>(L,L,L)</td>
<td>7</td>
<td>2398</td>
<td>0.1118</td>
<td>0.1552</td>
</tr>
<tr>
<td>(L,S,L)</td>
<td>8</td>
<td>224</td>
<td>0.0127</td>
<td>0.0123</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8980</td>
<td>0.423</td>
<td>0.577</td>
<td>6762</td>
</tr>
</tbody>
</table>
### Table 5: Subsampling Results

<table>
<thead>
<tr>
<th></th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>7796</td>
<td>5600</td>
<td>8536</td>
</tr>
<tr>
<td>$N_b$</td>
<td>3000</td>
<td>3000</td>
<td>3000</td>
</tr>
<tr>
<td>5% critical value</td>
<td>15</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>5% Confidence Interval $\rho$</td>
<td>[2.1e-05, 0.0026] ∪ [0.0037, 0.0042]</td>
<td>[2.1e-05, 0.0013] ∪ [0.0019, 0.0057]</td>
<td>[0.0027, 0.0053]</td>
</tr>
<tr>
<td>Number replication</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

- **Point Estimates from overall Sample**: $\rho = 0.0038$, standard error $= 0.0005$
Table 6: Structural Estimates and Simulations

\( w_1 \) and \( w_2 \) are measured in US\$100,000 of dollars

\( E[x(1-g(x))] \) is measured in percentage per year

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimates</th>
<th>S.E.</th>
<th>Estimates</th>
<th>S.E.</th>
<th>Estimates</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sector 1</td>
<td></td>
<td>Sector 2</td>
<td></td>
<td>Sector 3</td>
<td></td>
</tr>
<tr>
<td>( w_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S,S,S)</td>
<td>1 0.72</td>
<td>0.05</td>
<td>3.5</td>
<td>0.12</td>
<td>2.21</td>
<td>0.13</td>
</tr>
<tr>
<td>(S,L,S)</td>
<td>2 3.93</td>
<td>0.02</td>
<td>7.2</td>
<td>0.22</td>
<td>0.83</td>
<td>0.06</td>
</tr>
<tr>
<td>(S,L,L)</td>
<td>3 1.26</td>
<td>0.04</td>
<td>2.23</td>
<td>0.12</td>
<td>2.05</td>
<td>0.06</td>
</tr>
<tr>
<td>( E[x(1-g(x))] )</td>
<td>4 0.17</td>
<td>0.01</td>
<td>3.12</td>
<td>0.09</td>
<td>2.42</td>
<td>0.14</td>
</tr>
<tr>
<td>(L,S,S)</td>
<td>5 7.58</td>
<td>0.16</td>
<td>2.43</td>
<td>0.16</td>
<td>4.12</td>
<td>0.13</td>
</tr>
<tr>
<td>(L,L,S)</td>
<td>6 1.60</td>
<td>0.08</td>
<td>4.8</td>
<td>0.17</td>
<td>0.93</td>
<td>0.05</td>
</tr>
<tr>
<td>(L,L,L)</td>
<td>7 1.65</td>
<td>0.05</td>
<td>1.84</td>
<td>0.10</td>
<td>3.81</td>
<td>0.19</td>
</tr>
<tr>
<td>(L,S,L)</td>
<td>8 4.15</td>
<td>0.42</td>
<td>1.29</td>
<td>0.09</td>
<td>3.83</td>
<td>0.15</td>
</tr>
<tr>
<td>( w_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S,S,S)</td>
<td>1 5.52</td>
<td>3.5</td>
<td>-16.31</td>
<td>5.7</td>
<td>3.62</td>
<td>6.09</td>
</tr>
<tr>
<td>(S,L,S)</td>
<td>2 19.18</td>
<td>9.9</td>
<td>-5.19</td>
<td>28.2</td>
<td>-1.22</td>
<td>39.2</td>
</tr>
<tr>
<td>(S,L,L)</td>
<td>3 10.24</td>
<td>8.2</td>
<td>32.84</td>
<td>10.7</td>
<td>30.41</td>
<td>6.4</td>
</tr>
<tr>
<td>(S,S,L)</td>
<td>4 8.72</td>
<td>3.2</td>
<td>2.72</td>
<td>6.4</td>
<td>5.11</td>
<td>8.34</td>
</tr>
<tr>
<td>(L,S,S)</td>
<td>5 0.25</td>
<td>15.7</td>
<td>5.57</td>
<td>19.2</td>
<td>5.87</td>
<td>14.5</td>
</tr>
<tr>
<td>(L,L,S)</td>
<td>6 29.42</td>
<td>6.8</td>
<td>10.82</td>
<td>13.14</td>
<td>45.32</td>
<td>19.7</td>
</tr>
<tr>
<td>(L,L,L)</td>
<td>7 32.49</td>
<td>4.3</td>
<td>36.33</td>
<td>6.7</td>
<td>61.49</td>
<td>7.46</td>
</tr>
<tr>
<td>(L,S,L)</td>
<td>8 -4.15</td>
<td>13.4</td>
<td>-1.10</td>
<td>13.5</td>
<td>24.51</td>
<td>14.9</td>
</tr>
</tbody>
</table>
Table 7: Compensation Under Generalized Moral Hazard Vs Pure Moral Hazard
All estimates are scaled to be in hundreds of thousand of dollars

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std</th>
<th>Mean</th>
<th>Std</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S,S,S) 1</td>
<td>13.46</td>
<td>54.02</td>
<td>0.87</td>
<td>81.82</td>
<td>27.47</td>
<td>82.95</td>
</tr>
<tr>
<td>(S,L,S) 2</td>
<td>43.16</td>
<td>67.59</td>
<td>0.56</td>
<td>135.86</td>
<td>42.76</td>
<td>122.39</td>
</tr>
<tr>
<td>(S,L,L) 3</td>
<td>29.30</td>
<td>86.57</td>
<td>40.24</td>
<td>75.19</td>
<td>43.13</td>
<td>104.58</td>
</tr>
<tr>
<td>(S,L,L) 3</td>
<td>37.61</td>
<td>76.96</td>
<td>26.81</td>
<td>105.40</td>
<td>62.07</td>
<td>133.46</td>
</tr>
<tr>
<td>(S,L,L) 3</td>
<td>42.38</td>
<td>70.06</td>
<td>44.89</td>
<td>122.28</td>
<td>73.43</td>
<td>139.32</td>
</tr>
<tr>
<td>(L,S,L) 8</td>
<td>0.64</td>
<td>69.52</td>
<td>13.06</td>
<td>62.13</td>
<td>43.57</td>
<td>91.67</td>
</tr>
<tr>
<td>(S,S,S) 1</td>
<td>6.83</td>
<td>22.53</td>
<td>-2.25</td>
<td>118.81</td>
<td>15.16</td>
<td>67.79</td>
</tr>
<tr>
<td>(S,L,S) 2</td>
<td>46.57</td>
<td>119.69</td>
<td>-0.31</td>
<td>73.16</td>
<td>14.81</td>
<td>132.27</td>
</tr>
<tr>
<td>(S,L,L) 3</td>
<td>18.86</td>
<td>42.35</td>
<td>22.86</td>
<td>146.76</td>
<td>21.33</td>
<td>136.87</td>
</tr>
<tr>
<td>(S,L,L) 3</td>
<td>10.99</td>
<td>44.30</td>
<td>17.11</td>
<td>101.13</td>
<td>8.75</td>
<td>133.92</td>
</tr>
<tr>
<td>(S,S,L) 4</td>
<td>3.09</td>
<td>287.14</td>
<td>8.02</td>
<td>93.99</td>
<td>5.58</td>
<td>355.98</td>
</tr>
<tr>
<td>(S,S,L) 4</td>
<td>42.05</td>
<td>123.67</td>
<td>23.87</td>
<td>474.24</td>
<td>48.36</td>
<td>59.72</td>
</tr>
<tr>
<td>(L,L,L) 7</td>
<td>35.72</td>
<td>143.17</td>
<td>19.49</td>
<td>243.40</td>
<td>67.25</td>
<td>138.35</td>
</tr>
<tr>
<td>(L,L,L) 7</td>
<td>1.53</td>
<td>39.42</td>
<td>0.90</td>
<td>16.37</td>
<td>39.94</td>
<td>153.50</td>
</tr>
</tbody>
</table>

46
Figure 1: Selected Compensation Schedule