

# Identifying and Testing Models of Managerial Compensation

George-Levi Gayle and Robert A. Miller\*

Tepper School of Business, Carnegie Mellon University

November 30, 2011

## Abstract

This paper analyzes identification and estimation in principal agent models, and applies the approach to data on executive compensation. We investigate the optimal contract between a risk neutral principal and a risk adverse agent in two closely related models, one with pure moral hazard, and another in which the agent also has hidden information. We show that, with data on profits and compensation, such models are point identified up to the coefficient of absolute risk aversion, but that the risk aversion parameter is only partially identified from the principal's profit maximization conditions. Our empirical application focuses on the role of accounting information. In the pure moral hazard model the information is treated as independently verifiable; in the hybrid model we treat accounting data as unverifiable information that shareholders value only if the manager has incentives to reveal his knowledge. We find the pure moral hazard model does explain the correlation between financial returns and compensation, but can only jointly reconcile financial returns, accounting returns and managerial compensation by resorting to a taste based explanation where managerial preferences depend on the outlook of his firm relative to those of his rivals; indeed the CEO would be willing to pay for the privilege of working for the firm in the bad state under the pure moral hazard specification. The hybrid model yields more plausible findings: we do not reject the null hypothesis that managerial preferences are homogeneous across accounting states; moreover the estimates of the nonparametric bounds for the risk parameter, and the risk premium derived from the optimal contract for the hybrid specification, are comparable to previous estimates obtained for pure moral hazard models not exploiting accounting data.

---

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

# 1 Introduction

To a large extent the theory of optimal contracting in the presence of asymmetric information revolves around principal agent models. The first half of this paper analyzes the identification of a canonical class of such models with a risk neutral principal and a risk averse agent. We show the models are point identified up to the coefficient of absolute risk aversion, but that in the absence of further exclusion restrictions, the risk aversion parameter is only partially identified from profit or value maximization optimization conditions with data on profits or returns and compensation. The second half exploits our results on identification in an application of optimal contracting under moral hazard and hidden information to executive compensation. We use data on executive compensation, financial returns and accounting returns, to nonparametrically estimate bounds on the social costs of moral hazard in executive management and other measures of asymmetric information. Our estimation techniques is based on our identification results; by concentrating out the other parameters of the model the risk aversion parameter is obtained as a mapping of the underlying data generating process. This only leaves us with one parameter to estimate, the coefficient of absolute risk aversion.

Our framework incorporates two forms of asymmetric information, a pure moral hazard model where only the actions of the agent are hidden, and a hybrid model where there are both hidden actions and also information about the environment that is privy to the agent but not the principal. In the pure moral hazard model, the expected utility of a manager does not depend on the state of the firm at the beginning of the period but only on his outside alternative employment option. In a hybrid model managers are provided incentives to divulge the true state of the firm at the beginning of the period, their expected utility from truth telling typically differs across states, and the optimal contract equates the expected utility of revealing the true state of the firm with the expected utility of lying about it. We fully characterize the restrictions from the optimal contracts for both models on data measuring information provided by the agent about the state of the environment, gross profits (or more generally gross financial returns). We show that the pure and hybrid models are not nested, and the restrictions of optimal contract on the data generating process in equilibrium are testable. Our identification analysis establishes sharp and tight bounds on the identified set for both types of models. Our tests and estimators are based on these bounds.

The empirical work in the second half of this paper applies the principal agent models to a large panel data set with information about the financial and accounting returns to firms, compensation to their managers, background descriptors, and aggregate economic conditions. In the hybrid model we treat accounting information as discretionary, in other words information that shareholders cannot verify and in principle can be withheld by the manager through discretionary reporting. We find the pure moral hazard model does explain the correlation between financial returns and compensation, but can only jointly reconcile financial returns, accounting returns and managerial compensation by resorting to a taste based explanation characterized by heterogeneity in managerial preferences. In contrast, the hybrid model with stable homogeneous preferences is not rejected by the data. Our nonparametric bounds show that, overall, the total cost of moral hazard, and the loss from ignoring the incentive problem entirely, is not very sensitive to whether accounting returns

and hidden information are explicitly incorporated into the estimation framework; confirming previous parametric estimates, the cost of the latter dwarfs the cost of the former.

In the next section we develop our approach using a prototype static model of pure moral hazard. First we lay out the model and solve the optimal contract. Then we show how much of it can be identified from a cross section of data on identical firms with data on their profits and managerial compensation. This leads directly to procedures for testing and estimating the structural parameters of the model. We conclude this section with a discussion of how exclusion restrictions can be imposed to further narrow the identified set of parameters. In Section 3 we relax the assumption that shareholders and managers are equally well informed about the state of the firm, and analyze a static model in which the agent has hidden information and engages in actions. We set up the Lagrangian that yields the optimal contract, showing how the constraints for the optimization problem and the first order conditions are affected by the addition of hidden information. These changes work through the identification analysis of the hybrid model, where we compare the differences in the identified sets of both models.

Our empirical application is described in Section 4. We explain how the data were compiled and summarize its main features. The panel of roughly 27,000 observations covering the period 1993 to 2005 contains data on compensation to about 4,700 chief executive officers of publicly traded companies compiled from Standard and Poor's ExecuComp data base, financial and accounting information of the 2,600 publicly trade firms they manage, taken from the Center for Securities Research (CRSP), as well as background characteristics on the sector and size of the firms. Then we discuss how our model was modified to accommodate heterogeneity in firms, measurement error in the compensation, and consumption smoothing of executives, which gives the model dynamics. Our test results and estimates are reported in Section 5.

This work draws from, and is related to, several literatures. Our least cost approach to optimal contracting to moral hazard models with hidden information, or hybrid models, extends the two step procedure for solving pure moral hazard models pioneered by Grossman and Hart (1980). The closest papers to our work on nonparametric identification are the independent analyses of Perrigne and Vuong (2011), who also exploit predictions from principal agent theory to analyze nonparametric identification in models of moral hazard and adverse selection, and Huang, Perrigne and Vuong (2007), who identify and estimate a nonlinear pricing model of advertising.<sup>1</sup> The identified set of parameters in our models are defined by nonlinear inequalities that involved moments of the population from which the data is drawn. Other applications of moment inequalities in the parametric estimation of structural models, include Andrews, Berry and Jia (2004), Bajari, Bankard and Levin (2007), Ciliberto and Tamer (2009), Ho (2009), Ho, Ho and Mortimer (2011), Levine (2007), and Pakes, Porter, Ho, and Ishii (2006). To test the specification we appeal to results in Chernozhukov, Hong and Tamer (2007) and Romano and Shaikh (2006).

Murphy (1999), Prendergast (1999) and Chiappori and Salanie (2000) survey the large body of empirical work that investigates how well managerial compensation can be rational-

---

<sup>1</sup>Our work differs from theirs in many respects. The differences range from the theoretical structures considered, to the results obtained, and the nature of empirical problems to which the identification analysis applies. For example their theoretical framework models adverse selection, whereas we analyze a model with hidden actions and another with both hidden actions and hidden information.

ized by moral hazard. A much smaller number of recently published papers seek to quantify the economic significance of incentives in the labor market. They can be divided into those that exploit data on the institutional structure of incentives and compensation, to investigate how people respond to incentives, and those that estimate the relationship between firm returns and compensation, to investigate the role of asymmetric information in contracting problems. The first group of papers includes Ferrall and Shearer (1999), Shearer (2004), Dubois and Vukina (2009), Dufflo, Hanna and Ryan (2011) and Todd and Wolpin (2010); the second Haubrich (1994), Margiotta and Miller (2000), and Gayle and Miller (2009a, 2009b). We contribute to the latter body of work, by showing how chief executives use accounting records to keep shareholders abreast of the state of their firm relative to their competitors' for the purpose of setting appropriate incentives.

Finally, our identification, estimation, and empirical results add to the literature on estimating risk preferences. Previous studies have incorporated observed or unobserved heterogeneity, using data from laboratory experiments (e.g. Holt and Laury (2002, 2005), Harrison, Johnson, McInnes, and Rutstrom (2005), and Harrison, List, and Towe (2007)), field experiments (e.g. Harrison, List, and Towe (2007), Anderson, Harrison, Lau, Rutstrom (2008), and Dohmen, Falk, Huffman, and Sunde (2010)) and individual behavior in actual markets (e.g. Chetty (2006) and Cohen and Einav (2007)). Our results show that whether heterogeneity in risk preferences is necessary to rationalize observed behavior may depend on the information structure of the model and the degree of heterogeneity in the other parts of the model specification.

## 2 Pure Moral Hazard

To explain our approach to identification, estimation and testing, we first analyze a simple principal agent model. We set up a static model of pure moral hazard and derive the cost minimizing optimal contract of the principal for alternative effort levels by the agent, working diligently or shirking. Comparing the expected net revenue of both contracts yields the profit maximizing contract. Then we derive the set of structural parameters that are observationally equivalent for data on compensation and profits from identical profit maximizing principals. This leads to a procedure for testing whether or not the data generating process comes from the class of pure moral hazard models we analyze, and for constructing a confidence region for the identified set of parameters. Last, we show how exclusion restrictions can be imposed to further restrict the identified set of parameters in the presence of observed heterogeneity that affects some but not all of the parameters.

### 2.1 A Static Model

At the beginning of the period, a risk neutral principal, proposes a compensation plan to a risk averse agent, which depends on the future realization of the gross revenue to the principal. The plan may be an explicit contract or an implicit agreement. The agent decides whether to accept or reject the principal's (implicit) offer. If he rejects the offer he receives a fixed utility from an outside option. If he accepts the offer, the agent chooses between pursuing the principal's objectives of value maximization, called working diligently, versus

accepting employment for the principal but following the objectives he would pursue if he was paid a fixed wage, called shirking. The decision to accept or reject the offer is observed by the principal, but his work routine is not. After revenue is realized at the end of the period, the agent receives compensation according to the explicit contract or implicit agreement, and the remainder is profit to the principal. We introduce the notation, write down the model, and then solve for the cost minimizing contracts that elicit diligence and shirking.

### 2.1.1 Notation

We denote the workplace employment decision of the agent by an indicator  $l_0 \in \{0, 1\}$ , where  $l_0 = 1$  means the agent rejects the principal's offer. We denote the effort level choices by  $l_j \in \{0, 1\}$  for  $j \in \{1, 2\}$ , where work is defined by setting  $l_2 = 1$ , and shirking is defined by setting  $l_1 = 1$ . Since taking the outside option, working and shirking are mutually exclusive activities,  $l_0 + l_1 + l_2 = 1$ . Gross revenue to the principal is denoted by  $x$ , a random variable drawn from a probability distribution that is determined by the agent's work routine. After  $x$  is revealed the both the principal and the agent at the end of the period, the agent receives compensation according to the contract or implicit agreement. To reflect its potential dependence on (or measurability with respect to)  $x$ , we denote compensation by  $w(x)$ . The principal's profit is revenue less compensation,  $x - w(x)$ .

Denote by  $f(x)$  the probability density function for revenue conditional on the agent working, and let  $f(x)g(x)$  denote the probability density function for revenue when the agent shirks. We assume:

$$E[xg(x)] \equiv \int xf(x)g(x)dx < \int xf(x)dx \equiv E[x]$$

The inequality reflects the preference of principal for diligent work over shirking. Since  $f(x)$  and  $f(x)g(x)$  are densities,  $g(x)$ , the ratio of the two densities, is a likelihood ratio. That is  $g(x)$  is nonnegative for all  $x$  and:

$$E[g(x)] \equiv \int g(x)f(x)dx = 1$$

We assume there is an upper range of revenue that might be achieved from working, but is extremely unlikely to occur if the agent shirks. Formally:

$$\lim_{x \rightarrow \infty} [g(x)] = 0 \tag{1}$$

Intuitively this assumption states that a truly extraordinary performance by the principal can only be attained if the agent works. We assume that  $g(x)$  is bounded, an assumption that rules out the possibility of setting a contract that is arbitrarily close to the first best resource allocation, first noted by Mirrlees (1975), by severely punishing the agent when  $g(x)$  takes an extremely high value.

We assume the agent is an expected utility maximizer and utility is exponential in compensation, taking the form:

$$-l_0 - l_1\alpha_1 E[e^{-\gamma w(x)}g(x)] - l_2\alpha_2 E[e^{-\gamma w(x)}] \tag{2}$$

where without further loss of generality we normalize the utility of the outside option to negative one. Thus  $\gamma$  is the coefficient of absolute risk aversion, and  $\alpha_j$  is a utility parameter with consumption equivalent  $-\gamma^{-1} \log(\alpha_j)$  that measures the distaste from working at level  $j \in \{1, 2\}$ . We assume  $\alpha_2 > \alpha_1$  meaning that shirking gives more utility to the agent, than being diligent. A conflict of interest arises between the principal and the agent because he prefers shirking, meaning  $\alpha_1 < \alpha_2$ , yet the principal prefers the agent to work since  $E[xg(x)] < E[x]$ .

### 2.1.2 Optimal contracting

To induce the agent to accept the principal's offer and engage in his preferred activity, shirking, it suffices to propose a contract that gives the agent an expected utility of at least minus one. In this case we require  $w(x)$  to satisfy the inequality:

$$\alpha_1 E [e^{-\gamma w(x)} g(x)] \leq 1 \quad (3)$$

To elicit work from the agent, the principal must offer a contract that gives the agent a higher expected utility than the outside option provides, and a higher expected utility than shirking provides. In this case we require:

$$\alpha_2 E [e^{-\gamma w(x)}] \leq 1 \quad (4)$$

and:

$$\alpha_2 E [e^{-\gamma w(x)}] \leq \alpha_1 E [e^{-\gamma w(x)} g(x)] \quad (5)$$

To attain expected revenue of  $E[x]$  at minimal expected cost, the principal choose a schedule  $w(x)$  to minimize expected compensation, denoted by  $E[w(x)]$ , subject to Inequalities (4) and (5). Alternatively, the more limited expected revenue target of  $E[xg(x)]$  can be reached by minimizing  $E[w(x)]$  subject to Inequality (3). In the proof of Lemma 2.1 we show both problems have a Kuhn Tucker formulation that yields the following characterization of the solution to the two cost minimizing contracts.

**Lemma 2.1** *The minimal cost of employing a agent to shirk is  $\gamma^{-1} \log(\alpha_1)$ . To minimize the cost of inducing the agent to accept employment and work diligently the board offers the contract:*

$$w^o(x) \equiv \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[ 1 + \theta \left( \frac{\alpha_2}{\alpha_1} \right) - \theta g(x) \right] \quad (6)$$

where  $\theta$  is the unique positive solution to the equation:

$$E \left[ \frac{g(x)}{\alpha_2 + \theta[(\alpha_2/\alpha_1) - g(x)]} \right] = E \left[ \frac{(\alpha_2/\alpha_1)}{\alpha_2 + \theta[(\alpha_2/\alpha_1) - g(x)]} \right] \quad (7)$$

In the proof we show that the participation constraint is met with equality in both cases, pinning down the certainty equivalent wage. There is no point exposing the manager to uncertainty in a shirking contract by tying compensation to revenue. Hence a agent paid to shirk is offered a fixed wage that just offsets his nonpecuniary benefits,  $\gamma^{-1} \ln \alpha_1$ . The certainty equivalent of the cost minimizing contract that induces diligent work is  $\gamma^{-1} \ln \alpha_2$ ,

higher than the optimal shirking contract to compensate for the lower nonpecuniary benefits because  $\alpha_2 > \alpha_1$ . Moreover the agent is paid a positive risk premium of  $E[w^o(x)] - \gamma^{-1} \ln \alpha_2$ .<sup>2</sup> In this model of pure moral hazard these two factors, that diligence is less enjoyable than shirking, and more certainty in compensation is preferable, explains why compensating an agent to align his interests with the principal is more expensive than merely paying them enough to accept employment.

Profit maximization by the principal determines which cost minimizing contract the principal should offer the agent. The profits from inducing the agent to work diligently are  $x - w^o(x)$ , while the profits from employing the agent to shirk are  $xg(x) - \gamma^{-1} \log(\alpha_1)$ . Thus work is preferred by the principal if and only if:

$$\max\{0, \gamma E[xg(x)] - \log(\alpha_1)\} \leq \gamma E[x - w^o(x)] \quad (8)$$

while a shirking contract is offered if and only if:

$$\max\{0, \gamma E[x - w^o(x)]\} \leq \gamma E[xg(x)] - \log(\alpha_1) \quad (9)$$

Otherwise no contract is offered.

## 2.2 Identification in the Pure Moral Hazard Model

The parameters of the model are characterized by  $f(x)$  and  $g(x)$ , which together define the probability density functions of gross profits,  $(\alpha_1, \alpha_2)$ , the preference parameters for shirking and diligent work (relative to the normalized utility from taking the outside option), as well as the risk aversion parameter  $\gamma$ . For the purposes of this introductory example, we assume the data comprise independent draws of profits and compensation,  $(x_n, w_n)$  for a sample of  $N$  observations generated in equilibrium.

When the principal induces shirking and (9) holds, the density  $f(x)g(x)$  can be estimated from observations on profits, the wage is constant at  $w_n \equiv \gamma^{-1} \log(\alpha_1)$  for all  $n$ , but nothing more can be gleaned from the data about the structure of the model. Our analysis focuses on cases when work is induced, (8) holds, and compensation  $w_n$ , depends nontrivially on revenue  $x_n$ . Hence  $f(x)$  is identified, along with  $N$  points on the compensation schedule  $w_n \equiv w^o(x_n)$ . Under the assumptions of the model  $f(x)$  can be estimated with a nonparametric density estimator. From the compensation equation, the regularity condition on  $g(x)$  given by (1), and the fact that  $g(x)$  is nonnegative, the maximum compensation the agent can receive is:

$$\lim_{x \rightarrow \infty} w^o(x) = \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[ 1 + \theta \left( \frac{\alpha_2}{\alpha_1} \right) \right] \equiv \bar{w} \quad (10)$$

Thus  $\bar{w}$  is identified, and consistently estimated by the maximum compensation observed in the data. This essentially leaves  $\gamma$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $g(x)$  to identify from  $f(x)$ ,  $w^o(x)$ , and  $\bar{w}$ .

Our analysis proceeds in three steps. First we show that if  $\gamma$  is known, then  $\alpha_1$ ,  $\alpha_2$ , and  $g(x)$  are identified from the cost minimization problem. This means that the the set of

---

<sup>2</sup>To prove  $E[w^o(x)]$  is greater than its certainty equivalent,  $\gamma^{-1} \ln \alpha_2$ , we note that in the cost minimizing contract inducing diligence, (4) is met with equality. This implies  $\alpha_2 E[e^{-\gamma w(x)}] = 1$  or  $\gamma^{-1} \ln \alpha_2 = -\gamma^{-1} \ln E[e^{-\gamma w(x)}]$ . Thus  $E[w^o(x)]$  exceeds  $\gamma^{-1} \ln \alpha_2$  if and only if  $\exp E[\gamma w^o(x)]$  exceeds  $E[e^{-\gamma w(x)}]$  which is true by Jensen's inequality if and only if  $\gamma > 0$ .

observationally equivalent parameters can be indexed by the positive real number  $\gamma$ , the risk aversion parameter. Second, we show that the firm's preference for working over shirking provides an additional inequality that helps delineate the values of observationally equivalent  $\gamma$ . Third, we prove that the set of restrictions we have derived in the first two steps fully characterize the identified set.

### 2.2.1 Restrictions from cost minimization

Suppose  $\gamma$  is known, and define the mappings  $g(x, \gamma)$ ,  $\alpha_1(\gamma)$ , and  $\alpha_2(\gamma)$  as:

$$g(x, \gamma) \equiv \frac{e^{\gamma\bar{w}} - e^{\gamma w^o(x)}}{e^{\gamma\bar{w}} - E[e^{\gamma w^o(x)}]} \quad (11)$$

$$\alpha_1(\gamma) \equiv \frac{1 - E[e^{\gamma w^o(x) - \gamma\bar{w}}]}{E[e^{-\gamma w^o(x)}] - e^{-\gamma\bar{w}}} \quad (12)$$

$$\alpha_2(\gamma) \equiv \{E[e^{-\gamma w^o(x)}]\}^{-1} \quad (13)$$

All three mappings inherit the basic structure of the model for any positive value of  $\gamma$ . Integrating (11) over  $x$ , we that  $E[g(x, \gamma)] = 1$  for all  $\gamma > 0$ . Also by definition  $\bar{w} \geq w$ , so  $e^{\gamma\bar{w}} \geq E[e^{\gamma w^o(x)}]$  and  $e^{\gamma\bar{w}} \geq e^{\gamma w^o(x)}$  for all  $\gamma > 0$ . Therefore  $g(x, \gamma) \geq 0$  for all  $\gamma > 0$ . Furthermore as  $x \rightarrow \infty$ , from (10) we see that  $w(x) \rightarrow \bar{w}$ , and hence  $g(x, \gamma) \rightarrow 0$ , as stipulated by the regularity condition. This proves  $g(x, \gamma)$  can be interpreted as a likelihood ratio satisfying (1) for any  $\gamma > 0$ .

Next consider  $\alpha_1(\gamma)$  and  $\alpha_2(\gamma)$ . Clearly  $\alpha_2(\gamma) > 0$  because  $e^{-\gamma w^o(x)} > 0$ . Similarly the numerator and denominator of the equation for  $\alpha_1(\gamma)$  have the same sign for all  $\gamma$ , so  $\alpha_1(\gamma)$  is also positive. Rearranging the expression for the ratio of the two taste parameters we obtain:

$$\frac{\alpha_1(\gamma)}{\alpha_2(\gamma)} = \frac{e^{\gamma\bar{w}} - E[e^{\gamma w^o(x)}]}{e^{\gamma\bar{w}} - \{E[e^{-\gamma w^o(x)}]\}^{-1}} \quad (14)$$

Since the inverse function is convex, Jensen's inequality implies  $E[e^{-\gamma w^o(x)}] > E[e^{\gamma w^o(x)}]^{-1}$  or  $\{E[e^{-\gamma w^o(x)}]\}^{-1} < E[e^{\gamma w^o(x)}]$ , and consequently  $\alpha_1(\gamma) < \alpha_2(\gamma)$  for all positive  $\gamma$ .

To summarize, this discussion shows that, given a probability density  $f(x)$  for  $x$  and a compensation schedule  $w^o(x)$  satisfying  $w^o(x) \rightarrow \bar{w}$  as  $x \rightarrow \infty$ , identified and estimated from observations  $(x_n, w_n)$ , we can construct, for any positive  $\gamma$ , a likelihood ratio  $g(x, \gamma)$  and the taste parameters  $\alpha_1(\gamma)$  and  $\alpha_2(\gamma)$  that serve as primitives for a principal agent model of the type studied in the previous subsection, where the principal minimizes expected costs to elicit participation and working from the agent. Theorem 2.1 is a stronger result: if the risk parameter is known, then other primitives of the model are identified off data on compensation and returns using Equations (11), (12) and (13).

**Theorem 2.1** *Suppose the data on  $x_n$  and  $w_n$  is generated by a parameterization of a pure moral hazard model with risk aversion  $\gamma^*$ . Then:*

$$\begin{aligned} \alpha_1 &= \alpha_1(\gamma^*) \\ \alpha_2 &= \alpha_2(\gamma^*) \\ g(x) &= g(x, \gamma^*) \end{aligned}$$

The basic ideas for the proof of this theorem are straightforward. Making  $g(x)$  the subject of the compensation equation (6) and differentiating with respect to  $x$  yields:

$$g'(x) = -(\theta\alpha_2)^{-1} \gamma e^{\gamma w(x)} \partial w(x) / \partial x$$

From this equation it is evident that the slope is defined up to one normalization; a second normalization determines the level of  $g(x)$ . In our setup the regularity condition (1) provides one normalization; the fact that  $E[g(x)] = 1$  provides another. The proof to Lemma 2.1 shows that the participation constraint (4) is met with equality, and that gives the formula for  $\alpha_2(\gamma^*)$ . The incentive compatibility constraint is also met with equality, so:

$$\alpha_1 E[e^{-\gamma w^o(x)} g(x)] = \alpha_2 E[e^{-\gamma w^o(x)}] = 1 \quad (15)$$

Substituting for  $g(x)$  from (11) on the left side and rearranging to make  $\alpha_1$  the subject of the equation yields (12) evaluated at  $\gamma^*$ .

### 2.2.2 Restrictions from the firm's choice of contract

The restrictions from cost minimization tie down all the parameters up to preferences for risk, but place no restrictions at all on  $\gamma$ . Imposing profit maximization, as opposed to cost minimization only, does limit the set of admissible  $\gamma$ . Since expected profits from paying the agent  $w^o(x)$  are higher than paying him  $\gamma^{-1} \log(\alpha_1)$ , it follows from (8) that:

$$E[x] - E[w^o(x)] - E[xg(x)] + \gamma^{-1} \ln(\alpha_1) \geq 0 \quad (16)$$

Substituting for  $g(x)$  and  $\alpha_1$  from (11) and (12) in (16) define:

$$Q_0(\gamma) \equiv E[x] - E[w^o(x)] - E\left[x \frac{e^{\gamma \bar{w}} - e^{\gamma w^o(x)}}{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]}\right] + \gamma^{-1} \log\left(\frac{1 - E[e^{\gamma w^o(x) - \gamma \bar{w}}]}{E[e^{-\gamma w^o(x)}] - e^{-\gamma \bar{w}}}\right) \quad (17)$$

From Theorem 2.1  $Q_0(\gamma^*) \geq 0$ . This inequality restricts the set of  $\gamma$  that are admissible for the data generating process.<sup>3</sup>

### 2.2.3 Sharp and tight bounds

Theorem 2.1 only exploits the first order conditions, equalities for the participation and incentive compatibility conditions which we establish in the proof of the first lemma, and an inequality of the optimization problem. The second order conditions of the cost minimization problem are satisfied for all  $\gamma > 0$ . Are there any other restrictions? The short answer is no. We now establish that, given the underlying data generating process, every positive  $\gamma$  satisfying (17) is admissible. Thus  $\Gamma$ , a Borel set of risk aversion parameters, defined as:

$$\Gamma \equiv \{\gamma > 0 : Q_0(\gamma) \geq 0\} \quad (18)$$

indexes all parameterizations that are observationally equivalent to the true model. Theorem 2.2 provides sharp and tight bounds for the set of identified parameters of the pure moral hazard model.

---

<sup>3</sup>The principal also prefers a contract that induces diligent work to none at all. That is  $E[x] > E[w(x)]$ . However, as Theorem 2.2 below demonstrates, this inequality does not impose any further restrictions over those already implied in Theorem 2.1.

**Theorem 2.2** Consider any data generating process  $x_n$  and compensation schedule  $w(x)$  satisfying  $w_n = w(x_n)$  for all  $n$ . Define  $\Gamma$  from the  $x_n$  process using (18). If  $\Gamma$  is not empty, then  $(x_n, w_n)$  is observationally equivalent to every data process generated by the pure moral hazard model parameterized by each  $\gamma \in \Gamma$ . If  $\Gamma$  is empty, then  $(x_n, w_n)$  is not generated by such a pure moral hazard model.

Does there always exist for each joint probability distribution of  $(x, w)$  a positive  $\gamma$  that rationalizes this model of pure moral hazard? No, the next lemma shows the model has empirical content and would be rejected by some data generating processes. It proves the profit inequality embodies overidentifying restrictions that imply some joint distributions of profit and compensation are incompatible with any positive risk aversion parameter.

**Lemma 2.2** There exist joint distributions of  $(x, w)$  such that  $Q_0(\gamma) < 0$  for all  $\gamma > 0$ .

Intuitively, our results on identification can be summarized thus far as a four step process. Suppose that both the probability density of gross profits conditional on working, and the compensation schedule as a function of gross profits, are both known. Up to a normalization reflecting the outside option, the certainty equivalent utility of the contract, and hence the taste or ability parameter for working, is recovered through the participation constraint from the overall level of compensation, by adjusting for its variation using a given risk aversion parameter. Given a regularity condition and a risk aversion parameter, variation in compensation as a mapping of profits yields, through the first order condition, the likelihood ratio of the densities for shirking versus working. Since cost minimization also implies the incentive compatibility constraint is also met with equality, and the distribution of output conditional on shirking can be recovered using the previous step, the preference for shirking is also recovered as another level effect from that constraint for a known risk aversion parameter that adjusts for variation in compensation that arises when the gross profit distribution is conditioned on shirking. Only the profit maximization is left to partially identify the risk aversion parameter: there is just one inequality which reflects the principal's preference for a working contract, that yields the expected net profits calculated directly from the distribution of gross profits and expected compensation, over a fixed wage shirking contract, calculated using the second and third steps for any given risk aversion parameter.

## 2.3 Empirical Approach

The identified set of risk parameters defined  $\Gamma$  has a simple empirical analogue. Suppose we have  $N$  cross sectional observations on  $(x_n, w_n)$  on identical firms and their managers. To estimate  $Q_0(\gamma)$ , we replace  $\bar{w}$  with  $\bar{w}^{(N)} \equiv \max\{w_1, \dots, w_N\}$  in (17), and substitute sample moments for their population corresponding expectations, to obtain upon rearrangement:

$$Q_0^{(N)}(\gamma) \equiv \sum_{n=1}^N (x_n - w_n) / N - \sum_{n=1}^N x_n \left( e^{\gamma \bar{w}^{(N)}} - e^{\gamma w_n} \right) / \sum_{n=1}^N \left( e^{\gamma \bar{w}^{(N)}} - e^{\gamma w_n} \right) \\ + \gamma^{-1} \log \left[ \sum_{n=1}^N \left( e^{\gamma \bar{w}^{(N)}} - e^{\gamma w_n} \right) \right] + \gamma \log \left[ \sum_{n=1}^N e^{\gamma (\bar{w}^{(N)} - w_n)} - N \right] \quad (19)$$

Our tests are based on the fact that if  $\gamma \in \Gamma$  then sampling error is the only explanation for why  $Q_0^{(N)}(\gamma)$  might be negative. Clearly  $Q_0^{(N)}(\gamma)$  converges at the rate of its slowest

converging component. For simplicity suppose there exists some  $\bar{x} < \infty$  such that  $g(x) = 0$  for all  $x > \bar{x}$ . In words, there is a revenue threshold that shirking cannot achieve. Thus compensation is flat at  $\bar{w}$  for all profits levels above  $\bar{x}$ , and  $\bar{w}^{(N)}$  converges to  $\bar{w}$  at a faster rate than  $\sqrt{N}$ . Since all the other components of  $Q_0^{(N)}(\gamma)$  are sample moments, we conclude  $Q_0^{(N)}(\gamma)$  converges at rate  $\sqrt{N}$ .

Denote by  $\Gamma_\delta^{(N)}$  the set of risk aversion parameters that asymptotically cover the observationally equivalent set of  $\gamma > 0$  with probability  $1 - \delta$ . For the critical value  $c_\delta$  associated with test size  $\delta$ , this set is defined as:

$$\Gamma_\delta^{(N)} \equiv \left\{ \gamma > 0 : \min \left\{ 0, \sqrt{N} Q_0^{(N)}(\gamma) \right\}^2 \leq c_\delta \right\}$$

A consistent estimate of  $c_\delta$  can be determined numerically by following the subsampling procedures used in our empirical application and described in the appendix. Intuitively, if  $\sqrt{N} Q_0^{(N)}(\gamma)$  is negative and large in absolute value for all  $\gamma > 0$  we reject the null hypothesis that the pure moral hazard model generated the data. On the other hand if  $\sqrt{N} Q_0^{(N)}(\gamma^{**})$  is small in absolute value, or positive, we do not reject the null hypothesis that  $\gamma^{**}$  belongs to the identified set.

## 2.4 Multiple States and Exclusion Restrictions

When there is heterogeneity in probability distribution for revenue, additional restrictions on the set of observationally equivalent parameters can be imposed if preferences are invariant across states. Suppose there are two states denoted by  $s \in \{1, 2\}$ . We denote the probability density function of revenue from working diligently in state  $s$  by  $f_s(x)$ , and similarly express the corresponding likelihood ratio in  $s$  as  $g_s(x)$ . We assume  $f_1(x) \neq f_2(x)$  and  $g_1(x) \neq g_2(x)$ . The optimal contract, state dependent, but solved the same way as the one state model, is denoted by  $w_s^o(x)$ . We also write  $\bar{w}_s$  for the limiting constant wage as  $x \rightarrow \infty$  in state  $s$ . If the heterogeneity is observed, the data records the state  $s_n \in \{1, 2\}$ , revenue  $x_n \in R$  and compensation  $w_n \in R$  for each observation  $n \in \{1, \dots, N\}$ .

Suppose the risk aversion parameter of the agent does not vary across states, for example because the same type of agent works in both states. The solution to the cost minimization problem of inducing diligence, now denoted by  $w_s^o(x)$  to reflect the state dependence, is derived the same way as (6). For each state  $s \in \{1, 2\}$  we define  $Q_s(\gamma)$  analogously to  $Q_0(\gamma)$  for  $s \in \{1, 2\}$ , by substituting  $w_s^o(x)$  and  $\bar{w}_s$  for  $w^o(x)$  and  $\bar{w}$  respectively, as well as conditioning the expectations operator on the state, substituting  $E_s[\cdot]$  for  $E[\cdot]$  in (17).<sup>4</sup> Following the same reasoning as the derivation of (17),  $Q_s(\gamma^*) \geq 0$  for  $s \in \{1, 2\}$ . More generally, increasing the states while maintaining the hypothesis that the risk aversion parameter is invariant across states, increases the number of inequalities from profit maximization by the same number.

Now suppose that in addition nonpecuniary benefits from diligently working,  $\alpha_2$ , do not vary by state. Although there might only be one participation constraint ensuring that the unconditional expected utility of the agent is at least as high as what the outside alternative

<sup>4</sup>We adopt the notation that  $E_s[\cdot] \equiv E[\cdot | s]$  throughout.

offers, it is straightforward to show that the participation condition (4) holds with equality for each state  $s \in \{1, 2\}$  in the optimal contract, implying  $\alpha_2 E_s [e^{-\gamma w_s^o(x)}] = 1$ . Defining:

$$\Upsilon_2(\gamma) = E_1 [e^{-\gamma w_1^o(x)}] - E_2 [e^{-\gamma w_2^o(x)}] \quad (20)$$

it follows that  $\Upsilon_2(\gamma^*) = 0$ .<sup>5</sup> Intuitively, a person's risk preferences cannot be identified from playing a single lottery if there are unobserved components to the reward from entering the lottery. When offered the chance to play two lotteries with different risk characteristics but the same unobserved nonpecuniary components, his risk preferences are partially revealed by the pecuniary compensating differential between them, which equalizes his expected utility from playing one versus the other.

Another potential restriction is that the nonpecuniary benefits from shirking,  $\alpha_1$ , do not vary by state. Since the incentive compatibility constraint (5) also holds with equality in each state, Theorem 2.1 implies:

$$\alpha_1^{-1} = \frac{1 - E_1 [e^{\gamma w_1^o(x) - \gamma \bar{w}_1}]}{E_1 [e^{-\gamma w_1^o(x)}] - e^{-\gamma \bar{w}_1}} = \frac{1 - E_2 [e^{\gamma w_2^o(x) - \gamma \bar{w}_2}]}{E_2 [e^{-\gamma w_2^o(x)}] - e^{-\gamma \bar{w}_2}} \quad (21)$$

In this case the restriction is based on two hypothetical lotteries, compensation from shirking in the different states. To incorporate these restrictions into the testing and estimation framework we define:

$$\Upsilon_1(\gamma) = \frac{1 - E_1 [e^{\gamma w_1^o(x) - \gamma \bar{w}_1}]}{E_1 [e^{-\gamma w_1^o(x)}] - e^{-\gamma \bar{w}_1}} - \frac{1 - E_2 [e^{\gamma w_2^o(x) - \gamma \bar{w}_2}]}{E_2 [e^{-\gamma w_2^o(x)}] - e^{-\gamma \bar{w}_2}} \quad (22)$$

From (20) and (21), Theorem 2.1 implies  $\Upsilon_1(\gamma^*) = 0$ , if  $\alpha_1$  does not vary across states. A joint test of these restrictions can be based on the criterion function:

$$\min \{0, Q_1(\gamma)\}^2 + \min \{0, Q_2(\gamma)\}^2 + \Upsilon_1(\gamma)^2 + \Upsilon_2(\gamma)^2$$

which attains a minimum of zero at all risk aversion parameter values that are observationally equivalent to  $\gamma^*$ .

To summarize, we provide an intuitive explanation of how the extension of the static pure moral hazard model to two states affects identification. Consider as a baseline a framework with maximal heterogeneity, where the taste parameters for working and shirking, as well as the risk parameter, vary by state. With maximal heterogeneity, we obtain just two inequalities from the profit maximization condition. The two risk parameter sets are separately determined state by state. If a single risk parameter satisfies both profit inequalities, then it must belong to the intersection of the individually determined sets. In this way we can derive the set of risk parameters that are common across states, without imposing homogeneity on the other preference parameters. To impose homogeneity on the taste parameter for work as well, we would extract from the intersection derived above those risk parameters that induced the same taste parameter for work in both states. Alternatively, imagine permitting heterogeneity in the risk parameter across states, but imposing instead homogeneity on the taste parameter for work. We would seek to equalize the taste parameter for work across states using two different risk parameters that individually satisfy the profit inequalities for their respective states.

---

<sup>5</sup>If the equation was linear in  $\gamma$ , point identification could be easily determined, but it is not.

### 3 Hybrid Moral Hazard

When multiple states in the principal agent model arise from the resolution of uncertainty about the revenue generation process, it is reasonable to entertain the possibility that the agent might be more informed than the principal, especially in applications of the model to principals and agents. In hybrid models of moral hazard the agent is subject to moral hazard, but also has more information than the principal about the probability distribution from which revenues are drawn. This section presents a hybrid model that differs from the two state pure moral hazard model discussed in the previous section only because the agent is more informed. We explain the differences between the hybrid model and the pure model of moral hazard. Then we show how these differences translate to identification.

#### 3.1 Modifying the Pure Moral Hazard Model

In contrast to its pure moral hazard counterpart, in the two state hybrid model the agent has full information about the state but the principal does not. We model this below, imposing throughout the restrictions made in Section 2.4 that none of the preference parameters depend on the state. Then we explain how the information asymmetry is captured by inequality constraints that restrict the range of feasible contracts (implicit agreements). This leads to a statement of the optimization problem the principal solves, and its associated first order conditions. The identification analysis is based on the feasibility constraints of the contract and the first order conditions. Throughout we emphasize how and why the pure moral hazard analysis must be modified to account for the information asymmetry.

##### 3.1.1 Hidden information about the firm

As before, at the beginning of the period, the principal proposes a contract. In the hybrid model however, the agent's compensation is determined by what he discloses about the probability distribution of gross revenue, denoted by  $r \in \{1, 2\}$ , and its subsequent performance,  $x$ , revealed to both parties at the end of the period. We denote this mapping by  $w_r(x)$ . If the agent accepts employment with the principal over his outside option, the probability distribution of revenue is then fully revealed to the agent but remain partially hidden to the principal. There are two states  $s \in \{1, 2\}$ , and the probability state  $s$  occurs is identically and independently distributed with probability  $\varphi_s \in (0, 1)$ . As in the multistate pure moral hazard model, if the state is  $s$ , revenue is drawn from the probability density function  $f_s(x)$  if the agent is diligent and from  $g_s(x)$  if the agent shirks. We assume that the agent privately observes the true state  $s$ , reports the state  $r \in \{1, 2\}$  to the principal, and makes his effort choice. If the agent reports the second state, meaning  $r = 2$ , then the principal can independently confirm or refute it. (For example imagine principals can review geological surveys of new oil fields, but that agents exercise some discretion about when to disclose them.) This prevents the agent from lying when the first state occurs, and models the idea that legal considerations induce the agents not to overstate revenue prospects, but that incentives must be provided to dissuade agents from understating them.<sup>6</sup> If  $s = 2$  the

---

<sup>6</sup>Thus the managers at Enron, for example, were prosecuted as criminals, not penalized internally by shareholders, for overstating the firm's prospects. Dye and Sridhar (2005) make a similar assumption about

agent then truthfully declares or lies about the firm's prospects by announcing  $r \in \{1, 2\}$ , effectively selecting one of two schedules,  $w_1(x)$  or  $w_2(x)$  in that case, but if  $s = 1$  he reports  $r = 1$ . By way of contrast, in the analogous pure moral hazard model, the true state  $s \in \{1, 2\}$  is observed by the principal, thereby eliminating any role for reporting. This is the only difference between the pure and hybrid frameworks.<sup>7</sup>

If  $h(x) \equiv \varphi_2 f_2(x) / \varphi_1 f_1(x)$ , the weighted likelihood ratio of the second state occurring relative to the first given any observed value of excess returns  $x \in R$ , is unbounded for some value of  $x'$  (because  $f_2(x') > 0$  and  $f_1(x') = 0$ ), then truth telling about the second state can be enforced without cost.<sup>8</sup> To rule out this possibility, we assume  $h(x)$  is bounded throughout. The assumption implies the principal cannot be sure that the second state occurred simply by observing profits. To capture the idea that the second state is weakly more desirable than the first, we also assume that the second state is most likely to have occurred relative to the first state at the limit  $x \rightarrow \infty$ .<sup>9</sup> Summarizing our assumptions about  $h(x)$  mathematically:

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

### 3.1.2 Truth telling and sincerity constraints

Contracts between the principal and the agent that induce honest reporting in the second state and working in both states must satisfy a participation constraint plus two incentive compatibility constraints (one for each state) identical to those derived for the pure moral hazard model with two states, and two additional conditions that distinguish the hybrid from the pure moral hazard model, inducing the agent to truthfully reveal his private information. Define  $v_s(x) \equiv \exp[-\gamma w_s(x)]$  as the multiplicative utility value from the payoff  $w_s(x)$ . We rewrite the incentive compatibility constraint for each state as:

$$\int [\alpha_2 - \alpha_1 g_s(x)] v_s(x) f_s(x) dx \equiv E_s \{[\alpha_2 - \alpha_1 g_s(x)] v_s(x)\} \leq 0 \quad (24)$$

and the participation constraint for working as:

$$\alpha_2 \sum_{s=1}^2 \varphi_s \int [v_s(x)] f_s(x) dx \equiv \alpha_2 E[v_s(x)] \leq 1 \quad (25)$$

These three constraints correspond exactly to the three constraints in the two state pure moral hazard model. In the hybrid model we append them with two further constraints.

Comparing the expected value from lying about the second state and working with the expected utility from reporting honestly in the second state and working, the principal can

---

information disclosure in their theoretical analysis of the severity of moral hazard.

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

<sup>8</sup>The principal could promise to severely punish the agent if the first state is reported but  $x'$  is subsequently drawn as the revenue outcome.

<sup>9</sup>This assumption is implied, for example, by first order dominance.

induce the agent to tell the truth telling by restricting himself to contracts that satisfy:

$$\int [v_2(x) - v_1(x)] f_2(x) dx \equiv E_2 [v_2(x) - v_1(x)] \leq 0 \quad (26)$$

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

$$\int [\alpha_2 v_2(x) - \alpha_1 v_1(x) g_2(x)] f_2(x) dx \equiv E_2 [\alpha_2 v_2(x) - \alpha_1 v_1(x) g_2(x)] \leq 0 \quad (27)$$

where  $-\alpha_1 v_1(x)$  is 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.

### 3.1.3 Optimal contracting in the hybrid model

In the pure moral hazard framework the cost minimization problem is additively separable across states. In the hybrid model the information structure is not separable and this feature complicates the solution. Since  $v_s(x)$  is monotone decreasing in  $w_s(x)$ , deriving  $w_s(x)$  to minimize expected compensation of inducing work in both states subject to the five constraints is tantamount to choosing  $v_s(x)$  for each  $(s, x)$  to maximize:

$$\sum_{s=1}^2 \int \varphi_s \log [v_s(x)] f_s(x) dx \equiv E [\log v_s(x)] \quad (28)$$

subject to the same five constraints. To induce work and truth telling in both states, the principal maximizes the Lagrangian:

$$\begin{aligned} & \sum_{s=1}^2 \varphi_s \int \{\log [v_s(x)] + \eta_0 [1 - \alpha_2 v_s(x)]\} f_s(x) dx \\ & + \sum_{s=1}^2 \varphi_s \eta_s \int_{\underline{x}}^{\infty} v_s(x) [(\alpha_1 g_s(x) - \alpha_2)] f_s(x) dx \\ & + \varphi_2 \int \{\eta_3 [v_1(x) - v_2(x)] + \eta_4 [\alpha_1 v_1(x) g_2(x) - \alpha_2 v_2(x)]\} f_2(x) dx \end{aligned} \quad (29)$$

with respect to  $v_s(x)$ , where  $\eta_0$  through  $\eta_4$  are the shadow values assigned to the linear constraints. Since each constraint is a convex set, their intersection is too. Also  $\log v$  is concave increasing in  $v$ , the expectations operator preserves concavity, so the objective function is concave in  $v_s(x)$  for each  $x$ . Hence the Kuhn Tucker theorem guarantees there is a unique positive solution to the equation system formed from the first order conditions augmented by the complementary slackness conditions.

The differences between the cost minimization problems for the pure and hybrid moral hazard models are evident from (29). In the pure moral hazard model  $\eta_3 \equiv \eta_4 \equiv 0$  because

the truth telling and sincerity constraints do not figure into the formulation of the problem. The first order conditions for this problem are:

$$\begin{aligned} v_1(x)^{-1} &= \eta_0 + \eta_1 [(\alpha_2/\alpha_1) - g_1(x)] - \eta_3 h(x) - \eta_4 (\alpha_1/\alpha_2) g_2(x) h(x) \\ v_2(x)^{-1} &= \eta_0 + \eta_2 [(\alpha_2/\alpha_1) - g_2(x)] + \eta_3 + \eta_4 \end{aligned} \quad (30)$$

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

**Lemma 3.1** *The Lagrange multipliers satisfy:*

1.  $\eta_0 = \alpha_2$
2.  $\eta_3 + \eta_4 = E_2[v_2(x)]^{-1} - E[v_s(x)]^{-1}$

From the second equality in Lemma 3.1 we infer that if, as in the pure moral hazard model,  $\eta_3 = \eta_4 = 0$ , then:

$$E_2[v_2(x)] = E[v_s(x)] = E_1[v_1(x)]$$

In words, if neither the truth telling nor the sincerity constraints bind, or if the state is directly observed by the principal, then the pure moral hazard case applies, and expected utility is equalized across states. Otherwise  $(\eta_3 + \eta_4)$  is strictly positive implying expected utility from the pure moral hazard case straddles the expected utility attained in the two states of the hybrid model:

$$E_2[v_2(x)] < E[v_s(x)] < E_1[v_1(x)]$$

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

As in the two state pure moral hazard model, there are three other contracts the principal might design, all of which involve the agent shirking in at least one state. If the principal requires only participation but not diligence in both states, there is no reason to distinguish between the two. Thus the cost minimizing contract for achieving  $(l_1, l_2) = (0, 0)$  is found by setting  $\eta_1 = \eta_2 = 0$ , yielding an identical solution to the pure moral hazard model. In both states the agent is paid  $\gamma^{-1} \ln(\alpha_1)$ , equalizing expected utilities across both states. However both the remaining two cases yield counterintuitive results to the pure moral hazard problem. Suppose the principal opts for  $(l_1, l_2) = (1, 0)$ , diligence in the first state but shirking in the second. Then  $\eta_2 = 0$ , but from the second part of Lemma 3.1, the expected utility in the second state where there is shirking, is higher than in the first state, where diligence is called for! This is because the agent must be paid to reveal the second state, so that the principal can identify the first state to install incentives. Finally consider the cost minimizing way of achieving  $(l_1, l_2) = (0, 1)$ . In this case at least one of the multipliers,  $\eta_3$  or  $\eta_4$ , is strictly positive, and from the first order condition for compensation in the first state, this immediately implies  $v_1(x)$  and therefore compensation in the first state where the agent shirks, depends on profits through  $h(x)$  and possibly  $g_2(x)$  too. Rather than load all the risk premium into the second state, compensation to the agent optimally also depends on

profits in the first state, not to induce work in that state, but only to induce sincerity and truth telling that indirectly encourages diligence in the other (second) state.

Finally the type of contract chosen by the principal in a two state hybrid model of moral hazard is determined the same way as in the two state pure moral hazard model, by comparing net profits from the solutions to the four cost minimization problems, and offering the agent the most profitable one for the principal.

### 3.2 Identification in the Hybrid Model

The parameters of the hybrid model are also the same as for the two state pure moral hazard model, namely  $f_s(x)$  and  $g_s(x)$  for each state  $s \in \{1, 2\}$ , which together define the probability density functions of abnormal returns from working and shirking in the states, the probability of each state occurring  $\varphi_s$ , the probability distribution for the states,  $(\alpha_1, \alpha_2)$ , the preference parameters for leaving the firm, versus shirking and working within the firm, and the risk aversion parameter  $\gamma$ .

Instead of receiving data on  $(s, x, w)$ , we assume repeated cross sectional data is available on  $(r, x, w)$ , where  $r \in \{1, 2\}$  is a report by the agent on the firm's financial state. This difference reflects the fact that the data only records what the agent reports, rather than independently verifying the actual state. However we assume that the data is generated by our model, where agents truthfully reveals the state in equilibrium. We directly apply the implication of the theory, that  $s = r(s)$ , by setting  $r = s$ . Hence, as in the two state pure moral hazard model,  $f_s(x)$  is identified from observations on abnormal returns. This also implies that the probability  $\varphi_1$ , are identified along with  $h(x)$  (including  $\bar{h}$ ). Finally  $w_s(x) = w_r(x)$  is identified from observations on  $(r, x, w)$ . As in the pure moral hazard model we assume the regularity condition on the schedule  $w_r(x)$  that for all  $r$ , compensation is bounded, and the bound is approached at extraordinarily high returns, and we also apply the regularity condition on  $\bar{h}$  as well.

Although the methods for establishing the identified set is the similar to the pure moral hazard model, there are some notable differences in what can be identified from the different information structures embodied in the hybrid model. In the pure moral hazard model the participation constraint is binding for each individual state, whereas in the hybrid model, the expected utility determining participation is integrated over both states. This reduces the number of restrictions in a two state model by one. Instead the hybrid model impose two other constraints, truth telling and sincerity constraint. Since these constraints shape the optimal contract as a function of the parameters, they provide several restriction on the population moments that do not hold in the pure moral hazard model. From an econometrics standpoint the models are nonnested.

Again we focus on the empirical content of the equilibrium where diligence is induced in both states. First we represent the model parameters as functions of  $\gamma$  and quantities directly computed from the data generating process. Then we display another overidentifying restriction that arises if the tastes for shirking do not vary by state because the incentive compatibility constraint is met with equality in both states. This restriction corresponds to Equation (21) in the pure moral hazard case. By Lemma 3.1 expected utilities are not equalized across states in the hybrid model, ruling out a potential restriction analogous to Equation (20). Instead we derive restrictions that arise from the truth telling and sincerity

constraints, a difference that implies the pure and hybrid models are nonnested. A final set of restrictions comes from the inequalities implied by contract selection, similar to those obtained in the pure moral hazard case.

### 3.2.1 Solving for preferences and the likelihood ratio

The likelihood ratio  $\hat{g}_2(x, \gamma)$  and the preference parameters  $\hat{\alpha}_1(\gamma)$  and  $\hat{\alpha}_2(\gamma)$  are defined as in Equations (11) through (13), which apply to the one state pure moral hazard model. Thus:

$$\hat{\alpha}_2(\gamma) \equiv E[v_s(x, \gamma)]^{-1} \equiv \int \sum_{s=1}^2 \varphi_s v_s(x, \gamma) f_s(x) dx \quad (31)$$

and  $\alpha_2(\gamma) = \hat{\alpha}_2(\gamma)$  for all  $\gamma$ . However in the hybrid model we see from the first order conditions (30) that  $E_1[v_1(x, \gamma)] \neq E_2[v_2(x, \gamma)]$  even if tastes for work are not state dependent, whereas in the two state pure moral hazard model, finding  $E_1[v_1(x, \gamma)] \neq E_2[v_2(x, \gamma)]$  can only be reconciled with state dependent tastes for work. Similarly  $\hat{g}_2(x, \gamma)$  is defined as:

$$\hat{g}_2(x, \gamma) \equiv \frac{\bar{v}_2(\gamma)^{-1} - v_2(x, \gamma)^{-1}}{\bar{v}_2(\gamma)^{-1} - E_2[v_2(x, \gamma)^{-1}]} \quad (32)$$

The striking resemblance to  $g(x, \gamma)$  arises because, as in the pure moral hazard model, the principal can independently verify the agent's declaration that the second state has occurred, obviating the need to impose a truth telling constraint within the compensation contract. Turning now to the parameter for shirking we define the function:

$$\hat{\alpha}_1(\gamma) \equiv E[v_s(x, \gamma)]^{-1} \left\{ \frac{[\bar{v}_2(\gamma)]^{-1} - E_2[v_2(x, \gamma)]^{-1}}{[\bar{v}_2(\gamma)]^{-1} - E_2[v_2(x, \gamma)^{-1}]} \right\} \quad (33)$$

If the second state occurred with certainty this formula also would match to its analogue defined in (13) for the one state pure moral hazard model.

The essential differences arise in the two models arise from the role of the shadow prices on truth telling and sincerity in shaping (or distorting) the optimal contract in the first state, and hence statistical inferences from  $w_1(x)$  about the likelihood ratio  $g_1(x)$ . Recall that the principal cannot independently verify whether the first state has occurred if the agent declares it. Accordingly we define  $\hat{g}_1(x, \gamma)$  as:

$$\hat{g}_1(x, \gamma) \equiv \eta_1(\gamma)^{-1} \{ \bar{v}_1(\gamma)^{-1} - v_1(x, \gamma)^{-1} + \eta_3(\gamma) [\bar{h} - h(x)] - \eta_4(\gamma) [\hat{\alpha}_1(\gamma) / \hat{\alpha}_2(\gamma)] \hat{g}_2(x, \gamma) h(x) \} \quad (34)$$

where the Kuhn Tucker multiplier representations are defined as:

$$\eta_4(\gamma) \equiv \frac{E[v_s(x, \gamma)]^{-1} E_1[v_1(x, \gamma)] - 1}{\frac{[\hat{\alpha}_1(\gamma) / \hat{\alpha}_2(\gamma)] E_1[v_1(x, \gamma) \hat{g}_2(x, \gamma) h(x)] - E_1[v_1(x, \gamma) h(x)]}{E_1[v_1(x, \gamma) h(x)] \{ E_2[v_2(x, \gamma)]^{-1} - E[v_s(x, \gamma)]^{-1} \}} - \frac{[\hat{\alpha}_1(\gamma) / \hat{\alpha}_2(\gamma)] E_1[v_1(x, \gamma) \hat{g}_2(x, \gamma) h(x)] - E_1[v_1(x, \gamma) h(x)]}{E_1[v_1(x, \gamma) h(x)] \{ E_2[v_2(x, \gamma)]^{-1} - E[v_s(x, \gamma)]^{-1} \}}} \quad (35)$$

$$\eta_3(\gamma) \equiv E_2[v_2(x, \gamma)]^{-1} - \eta_4(\gamma) - E[v_s(x, \gamma)]^{-1} \quad (36)$$

$$\eta_1(\gamma) \equiv [\hat{\alpha}_1(\gamma) / \hat{\alpha}_2(\gamma)] \{ \bar{v}_1(\gamma)^{-1} - E[v_s(x, \gamma)]^{-1} + \eta_3(\gamma) \bar{h} \} \quad (37)$$

It is straightforward to check that if  $\eta_3(\gamma) = \eta_4(\gamma) = 0$  as in the pure moral hazard case, then the expression for  $\widehat{g}_1(x, \gamma)$  simplifies to Equation in Section.

We are now ready to prove that if  $\gamma^*$  is known, then both parameters,  $\alpha_1^*$  and  $\alpha_2^*$ , plus the likelihoods,  $g_1^*(x)$  and  $g_2^*(x)$ , are identified from data on abnormal returns, agent reports and compensation  $(x_n, r_n, w_n)$ .

**Theorem 3.1** *If the data is generated by a hybrid model of moral hazard with positive risk aversion parameter  $\gamma^*$  then:*

$$\begin{aligned}\alpha_1 &= \widehat{\alpha}_1(\gamma^*) \\ \alpha_2 &= \widehat{\alpha}_2(\gamma^*) \\ g_1(x) &= \widehat{g}_1(x, \gamma^*) \\ g_2(x) &= \widehat{g}_2(x, \gamma^*)\end{aligned}$$

### 3.2.2 The identified set of parameters

The (restricted) hybrid model also yields a restriction from the same value of  $\alpha_1$  appearing in the incentive compatibility conditions for both states. Defining:

$$\begin{aligned}\Psi_1(\gamma) \equiv & [\bar{v}_1(\gamma)]^{-1} - \eta_3(\gamma)\bar{h} + E_1[v_{1t}(x, \gamma)^{-1}] - \eta_1(\gamma) \\ & - \eta_3(\gamma)\varphi_2/\varphi_1 - \eta_4(\gamma)(\widehat{\alpha}_1(\gamma)/\widehat{\alpha}_2(\gamma))E_1[\widehat{g}_2(x, \gamma)h(x)]\end{aligned}\quad (38)$$

we prove in the appendix that this restriction can be stated as  $\Psi_1(\gamma^*) = 0$  in the hybrid model. Setting  $\eta_3 = \eta_4 = 0$ , the functional form of  $\eta_1(\gamma)$  simplifies to yield the identical restriction embodied in Equation (21) of the two state pure moral hazard case where tastes for shirking are not state dependent.

Appealing to Lemma 3.1 the participation equations in pure moral hazard models do not hold in hybrid models state by state, because the sum of the shadow values for truth telling and sincerity are strictly positive, inducing an increase (offsetting decrease) between the expected utility received in the second (first) state relative to the expected utility when the participation decision is made. This removes a restriction. Offsetting this one restriction are two extra equalities plus seven extra inequalities in the hybrid model, that either do not apply or are automatically satisfied in the pure moral hazard model.

Define  $\Psi_2(\gamma)$  through  $\Psi_4(\gamma)$  as :

$$\begin{aligned}\Psi_2(\gamma) &\equiv E_1[1\{\widehat{g}_1(x, \gamma)\} - 1] \\ \Psi_3(\gamma) &\equiv E_2[v_2(x, \gamma) - v_1(x, \gamma)] \\ \Psi_4(\gamma) &\equiv E_2[\widehat{\alpha}_1(\gamma)v_1(x, \gamma)\widehat{g}_2(x, \gamma^*) - \widehat{\alpha}_2(\gamma)v_2(x, \gamma)]\end{aligned}\quad (39)$$

By Theorem 2.2, the truth telling constraint (26) implies  $\Psi_3(\gamma^*) \geq 0$ , while the sincerity constraint (27) implies  $\Psi_4(\gamma^*) \geq 0$ . The equality  $\Psi_3(\gamma^*)\Psi_4(\gamma^*) = 0$  guarantees at least one of the constraints holds strictly. Also both sets of complementary slackness conditions for truth telling and sincerity must be satisfied, meaning  $\Psi_3(\gamma^*)\eta_3(\gamma^*) = 0$  and  $\Psi_4(\gamma^*)\eta_4(\gamma^*) = 0$ . Since  $g_1(x)$  is a likelihood ratio in the hybrid model we ensure  $\widehat{g}_1(x, \gamma^*) \geq 0$  with unit mass by imposing the restriction that  $\Psi_2(\gamma^*) \geq 0$ . Finally three inequalities that ensure

the Kuhn Tucker multipliers,  $\eta_1(\gamma^*)$ ,  $\eta_3(\gamma^*)$  and  $\eta_4(\gamma^*)$  are positive. Although there are a greater the number of inequality and equality restrictions imposed by the hybrid model than the pure moral hazard model, we are not asserting that hybrid moral hazard is more restrictive: the two models are nonnested.

Turning now to the effort level induced by the principal in the hybrid model, we first remark that if shirking is demanded in both states, that is  $(l_1, l_2) = (0, 0)$ , then compensation is determined in Lemma 2.1 for the one state pure moral hazard model. Since this is suboptimal:

$$\Lambda_1(\gamma) \equiv E[x - w] - E[xg_s(x, \gamma)] + \gamma^{-1} \log[\hat{\alpha}_1(\gamma)] \quad (40)$$

is positive at  $\gamma^*$ .

Since  $(l_1^o, l_2^o) = (1, 1)$  the expected value to the firm would have been lower if either  $(l_1, l_2) = (1, 0)$  or  $(l_1, l_2) = (0, 1)$  had been chosen, and this observation yields two extra restrictions on  $\gamma^*$  to be utilized in identification. For any  $\gamma \in R^+$ , we denote by  $w_s^{(l_1, l_2)}(x, \gamma)$  the cost minimizing contracts to induce  $(l_1, l_2)$ . Thus  $w_s(x) \equiv w_s^{(1, 1)}(x, \gamma^*)$  and from the notation defined in the previous section,  $w_s^{(l_1, l_2)}(x) \equiv w_s^{(l_1, l_2)}(x, \gamma^*)$ . Given the parameterization indexed by  $\gamma$ , the difference in value to the principal from selecting  $(l_1, l_2) = (1, 1)$  versus  $(l_1, l_2) = (1, 0)$  is:

$$\Lambda_2(\gamma) = \varphi_2 E[x] - E[w] - \varphi_2 E_2 \left[ xg_2(x, \gamma) - w_2^{(1, 0)}(x, \gamma) \right] - \varphi_1 E_1 \left[ w_1^{(1, 0)}(x, \gamma) \right]$$

At  $\gamma = \gamma^*$  this expression is positive since  $(l_1^o, l_2^o) = (1, 1)$ . In similar fashion we define:

$$\Lambda_3(\gamma) = \varphi_1 E[x] - E[w] - \varphi_1 E_1 \left[ xg_1(x, \gamma) - w_2^{(0, 1)}(x, \gamma) \right] - \varphi_2 E_2 \left[ w_1^{(0, 1)}(x, \gamma) \right] \quad (41)$$

and note that  $\Lambda_3(\gamma^*) \geq 0$  because setting  $(l_1, l_2) = (1, 1)$  is more profitable than setting  $(l_1, l_2) = (0, 1)$ .

Consolidating the restrictions directly applied to the hybrid model, we define  $\hat{\Gamma}$ , a Borel set of risk aversion parameters, as:

$$\hat{\Gamma} \equiv \left\{ \gamma > 0 : \begin{array}{l} \Lambda_i(\gamma) \geq 0 \text{ for } i \in \{1, 2, 3\}. \\ \eta_j(\gamma) \geq 0 \text{ for } j \in \{1, 3, 4\}. \\ \Psi_1(\gamma) = 0 \text{ and } \Psi_k(\gamma) \geq 0 \text{ for } k \in \{3, 4\}. \\ \Psi_3(\gamma) \Psi_4(\gamma) = \Psi_3(\gamma) \eta_3(\gamma) = \Psi_4(\gamma) \eta_4(\gamma) = 0. \end{array} \right\} \quad (42)$$

Our last theorem establishes an analogous result to Theorem 2.2. Theorem 3.2 establishes that the bounds we have constructed for the hybrid model are sharp and tight.

**Theorem 3.2** *Consider any data generating process  $(r_n, x_n)$  and compensation schedule  $w_r(x)$  satisfying  $w_n = w_{r_n}(x_n)$  for all  $n$ . Upon setting  $r_n = s_n$ , define  $\hat{\Gamma}$  from the  $(r_n, x_n)$  process using (42). If  $\hat{\Gamma}$  is not empty, then  $(r_n, x_n, w_n)$  is observationally equivalent to every data process generated by the hybrid moral hazard model parameterized by each  $\gamma \in \hat{\Gamma}$ . If  $\hat{\Gamma}$  is empty, then  $(r_n, x_n, w_n)$  is not generated by such a hybrid moral hazard model.*

Summarizing the essential differences between the two state pure and hybrid moral hazard models from the perspective of identification: they are clearly non-nested. There are three profit inequalities in the hybrid, but only two in the pure, whereas the hybrid has only one participation constraint, while the pure has two. Finally, both have two incentive compatibility constraints, but the hybrid also has truthtelling and sincerity constraints.

## 4 An Empirical Application

As an empirical matter, managerial compensation varies significantly with abnormal financial return.<sup>10</sup> The theory of pure moral hazard postulates that risk averse managers should receive compensation that fluctuates with signals risk neutral shareholders observe about decisions their managers make, most notably abnormal returns, in order to align the incentives of the managers when their nonpecuniary goals differ from maximizing shareholder wealth and the actions and decision of management are not monitored. Although the dominant paradigm, this explanation for executive compensation has been challenged on several fronts. First, as we show below, managerial compensation not only depends on the financial returns of the firm, but also its accounting returns. In models of pure moral hazard, shareholders might use signals other than financial returns to determine optimal compensation, but the reporting of accounting income is subject to considerable discretion by the manager. Our hybrid model of moral hazard provides a framework for analyzing unverifiable claims by management that are credible because of financial incentives embedded in their compensation to be truthful and sincere. Second, managers are paid for luck, risk factors beyond executive control that increase the volatility of their income,<sup>11</sup> which is inconsistent with the notion of mitigating uncertainty in compensation to risk averse agents. The optimal contract in our hybrid model rewards managers for announcing good news they observe privately, so that they are incentivized to work in favorable states. Third, several empirical studies find that trading by corporate insiders appears profitable.<sup>12</sup> In models of pure moral hazard, managers do not have private information about the firm's future prospects; by way of contrast a prediction of the hybrid model is that managers take advantage of private information they have. In our empirical study we investigate alternative ways of rationalizing these three anomalies, by allowing for sufficient heterogeneity of preferences and abilities within a dynamic version of the pure moral hazard model, versus resorting to the richer information structure that characterizes a model of hybrid moral hazard.

Our approach to estimation and testing can be conducted by data pairing profits,  $x_n$ , with contractual arrangements,  $w_s(x)$ , or with actual payouts  $w_n$ . If as in Section 2 the data the contractual arrangements comprehensively define the payouts in every contingency  $s$ , and compensation is measured without error, the two approaches are equivalent because  $w_n = w(x_n)$ . Findings in Hayes and Schaefer (2000) and Gillen, Hartzell and Parrino (2009) suggest that implicit understandings between the board and management are an important factor in managerial compensation policy. The details of the institutional process determin-

---

<sup>10</sup>See Antle and Smith (1985, 1986), Hall and Liebman (1998) and Gayle and Miller (2008a), who find that about half the total variation in compensation can be explained by a nonlinear regression on excess returns, industry effects and bond prices.

<sup>11</sup>See Bertrand and Mullainathan (2001) and Bebchuk and Fried (2003).

<sup>12</sup>See Lorie and Niederhoffer (1968), Jaffe (1974), and Finnerty (1976) and Seyhun (1986), who finds that insiders tend to buy before an abnormal rise in stock prices and sell before an abnormal decline. Seyhun (1992a, 1992b) presents evidence showing that insiders earn over 5 percent abnormal returns on average, and determines that insider trades predict up to 60 percent of the total variation in one-year-ahead returns. Hayes and Schaefer (2000) presents evidence that the unexplained variation in current compensation predicts the future variation in firm performance. Gayle and Miller (2009a) construct a simple self-financing dynamic portfolio strategy based on changes in asset holdings by managers that significantly outperforms the market portfolio, realizing over 90 percent of the gains that could have been achieved with perfect foresight.

ing compensation only adds plausibility to their results: typically executive compensation committees convene several times a year, perhaps in conference calls, using spreadsheets, and benchmarking other comparably placed executives against firm performance, to determine a value for each component of compensation, that is how much cash and bonus to pay, how many stock options of a given type to grant, and so forth. Reliable information on how the manager would have been paid if firm profits had deviated from the actual outcome do not exist. For this reason all empirical work on CEO compensation is based on actual payouts, that is  $w_n$ , rather than an incomplete understanding of the contractual arrangement,  $w_s(x)$ .<sup>13</sup>

We follow Antle and Smith (1985, 1986), Hall and Liebman (1998), Margiotta and Miller (2000) and Gayle and Miller (2009a, 2009b) to measure total executive compensation. It is the sum of salary and bonus, the value of restricted stocks and options granted, the value of retirement and long term compensation schemes, plus changes in wealth from holding firm options, and changes in wealth from holding firm stock relative to a well diversified market portfolio instead.<sup>14</sup> However we do not assume that compensation is measured without error; on the one hand the data source, the Securities and Exchange Commission, imposes harsh penalties on fraudulent reporting; on the other hand some income managers receive might not be recognized as such and consequently not subject to the same reporting requirements.

The remainder of this section describes the longitudinal data set used in our empirical study. To account for cross sectional differences between firms in any given year we use firm level data on sector, employment, assets, and debt to equity ratio. Financial returns on the market portfolio and bond prices index the aggregate variation over time. Accounting return is treated as a signal the manager sends stockholders about the state of the firm and its future profitability prospects relative to firms with similar characteristics. In our empirical framework managerial compensation is explained by these variables.

## 4.1 Data

Our primary data source is Standard & Poor’s ExecuComp database. We extracted compensation data on the current chief executive officer (CEO) of 2,610 firms in the S&P 500, Midcap, and Smallcap indices spanning the years 1992 to 2005. We supplemented these data with firm level data obtained from the S&P COMPUSTAT North America database and monthly stock price data from the Center for Securities Research (CRSP) database. The sample was partitioned into three industrial sectors by GICS code. Sector 1, called primary, includes firms in energy (GICS:1010), materials (1510), industrials (2010,2020,2030), and utilities (5510). Sector 2, consumer goods, comprises firms from consumer discretionary

---

<sup>13</sup>Whether  $x_n$  should be paired with  $w(x)$  or  $w_n$  depends on the application. For example in Cohen and Einav (2007) empirical study of auto insurance contracts, information on  $w(x)$  is used, but information on  $w_n$  is not. Because the contracts are marketed to large numbers of people, most if not all contingencies are explicit within the contractual arrangement. With minimal modifications, our econometric approach can be applied to such settings by substituting  $w(x)$  instead of  $w_n$  where appropriate.

<sup>14</sup>Changes in wealth from holding firm stock and options reflect the costs a manager incurs from not being able to fully diversify his wealth portfolio because of restrictions on stock and option sales. When forming their portfolio of real and financial assets, managers recognize that part of the return from their firm denominated securities should be attributed to aggregate factors, so they reduce their holdings of other stocks to neutralize those factors.

(2510,2520,2530,2540,2550) and consumer staples (3010,3020,3030). Firms in health care (3510,3520), financial services (4010,4020,4030,4040), information technology and telecommunication services (410, 4520, 4030, 4040, 5010) comprise Sector 3, which we call services.

Table 1 summarizes the cross sectional features of our data. Almost twice as many firms in services, as in consumerables, with the primary sector accounting for about half the observations. Average firm size by total assets is highest in the services sector and lowest in the consumer sector. This ordering is reflected by the debt equity ratio, the sector with largest firms by asset also being the most highly leveraged, but reversed when employment is used to measure firm size instead. For this reason we used both total assets and employment as two measures of size, and included the debt equity ratio as a factor that might affect the distribution of abnormal returns, and hence managerial compensation.<sup>15</sup> In this study we assume that firm sector, the firm's total assets, the number of its employees, and its debt equity ratio, is public information.

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

Table 2 summarizes the longitudinal features of our data. There are roughly the same number of observations per year, apart from 2005, where we only include data on firms whose financial records for that financial year ended before December.<sup>16</sup> 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. Note that the actions of an individual manager are too inconsequential to appreciably affect the stock index. For this reason we take, as our measure of the component of profit that managers can affect through his actions, financial returns to the firm net of the share market index return. This latter variation in abnormal returns, rather than variability due to aggregate factors, is critical to explaining managerial compensation.

The collective signal managers send about business, average accounting returns, is highly correlated with financial returns, almost without exception rising and falling together. Note though that accounting returns have a considerably higher standard deviation, in part attributable to fixed effects across firms, but also to higher idiosyncratic variability over time.

The term structure of interest rates underlying the bond price series were constructed from data on Treasury bills of varying maturities, and the prices were derived using methods described in Gayle and Miller (2009b). 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.

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.

---

<sup>15</sup>Findings of several studies, including our own, show this is indeed the case.

<sup>16</sup>Paranthenetically 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.

The cross sectional distribution of employees is similarly skewed, but in contrast to assets, firm employment on average grows by less than a quarter. More remarkable than changes in annual average debt equity ratio, which ranges between 2.41 and 4.69, is its standard deviation, which varies between 5 and 105.

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

## 4.2 Firm Characteristics and Signals

In our empirical analysis we allow for heterogeneity between firms by classifying firms within each of the three sectors on the basis of three indicators, total assets at the beginning of the period (or the end of the previous period), total employment, and its debt to equity ratio. More specifically we classify each firm by whether its total assets were less than or greater than median total assets for firms in the sector, whether its total employment were less than or greater than median employment for firms in the sector, and whether its debt to equity ratio was less than or greater than the median debt to equity ratio for firms in the sector. To notate these size indicators, let  $(S, S, L)$  mean lower total assets and employment than the median firm in the sector, but a higher debt to equity ratio than the median debt equity ratio for firms in the sector. Similarly let  $(L, S, L)$  mean lower employment than the median firm in the sector but greater than the median in the other two size indicators.

Managers release information about the state of the firm through accounting statements, and exercise considerable discretion over the values which are reported. They have many ways of directly affect the firm's balance sheets, choosing for example among different valuation methods for credits and liabilities, and using discretionary timing when writing off nonperforming assets. Exercising such liberties provides a mechanism for managers to signal the state of the firm to shareholders.

A commonly used accounting measure of the manager's accomplishments and firm's success is the difference between the change in assets and the changes in liabilities plus dividends, called comprehensive income. Let  $A_{nt}$  denote total assets reported at the end of the  $t^{th}$  period and  $Debt_{nt}$  the level of debt reported at the end of the period. Thus  $(A_{nt} - Debt_{nt})$  denotes net assets as reported in the annual report of the  $n^{th}$  firm up to the end of period  $t$ .<sup>17</sup> Normalizing comprehensive income, we define the accounting return  $\pi_{nt}$  for the firm in period  $t$  as:

$$\pi_{nt} = (A_{nt} - Debt_{nt} + Dividend_{nt}) / (A_{n,t-1} - Debt_{n,t-1}) \quad (43)$$

---

<sup>17</sup>We measure financial returns annually on a calendar year basis, while the month in which annual reports are released varies by company.

These variables are used to form our measures of  $r_{nt}$ . For a given firm type let  $E[\pi_{nt}]$  denote the expected accounting return of  $\pi_{nt}$  for firm  $n$  at the beginning of period  $t$  before the manager announces total assets  $A_{nt}$ . We define the manager's report about the hidden state  $r \in \{1, 2\}$  as an indicator variable, telling whether the firm's accounting return is higher or lower than the expected value of accounting returns in that period  $t$ .

We note that shareholders receive  $Dividend_{n,t-1}$  during the course of period  $t-1$ , and  $(A_{n,t-1} - Debt_{n,t-1})$  has been tabled by the end of period  $t-1$ , leaving  $\pi_{nt}$  to be determined by the manager's current report of  $(A_{nt} - Debt_{nt})$ . This definition is internally consistent with the timing in our model, because the dual effects of a payoff relevant announcement by the manager during the current period are evident in a capital gain or loss through a change in the stock price immediately following, and also in the balance sheet, tabled at the end of the period. In this fashion our model captures the cumulative impact of announcements made throughout the period by their net effect on the balance sheet and shareholder equity value at the end of the period.

Table 3 displays the number of observation in each sector and size category, and the probability that the report is *good*. 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. The latter columns of Table 3 provide a cross sectional summary of the average abnormal returns conditional for each size category by sector and report. The sample means for returns and compensation are without exception higher when a favorable report indicating the good state is released.

### 4.3 Bond Prices and Dynamic Considerations

The managers in our data set are about 55 years old and on average typically last less than 10 years as CEO before retiring.<sup>18</sup> They spend the compensation earned over that period throughout the remainder of their lives, taking account of future accruals from compensation and returns on wealth. It follows that their consumption and savings decisions, and the value of their compensation packages, are complicated by interest rate fluctuations. Thus variation in economic conditions provides a source of identification in models of moral hazard and restrictions in estimation. To account for the fact that the value of compensation, and also the compensating differential of nonpecuniary benefits, partly depends on the interest rate, we allow  $(\alpha_1, \alpha_2, \gamma)$  to depend on bond prices, by setting:

$$\alpha_1 \equiv \tilde{\alpha}_1^{1/(b_t-1)}, \quad \alpha_2 \equiv \tilde{\alpha}_2^{1/(b_t-1)}, \quad \gamma \equiv \tilde{\gamma}/b_{t+1} \quad (44)$$

where  $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\gamma})$  become the primitive preference parameters to be estimated as functions of  $z$ , the characteristics of the firm. Equation (2), the agent's preferences, becomes:

$$-l_{0t} - l_{1t} \tilde{\alpha}_1^{1/(b_t-1)} E \left[ g(x) \exp \left( \frac{-\tilde{\gamma} w_{t+1}(x)}{b_{t+1}} \right) \right] + l_{2t} \tilde{\alpha}_2^{1/(b_t-1)} E \left[ \exp \left( \frac{-\tilde{\gamma} w_{t+1}(x)}{b_{t+1}} \right) \right] \quad (45)$$

Comparing (2) with (45), instead of managers with risk aversion parameter  $\gamma$  receiving  $w(x)$ , they have risk aversion parameter  $\tilde{\gamma}$  and receive only the interest on the bonds

---

<sup>18</sup>A detailed description of the demographic traits of the managers in this sample is given in Gayle, Golan and Miller (2010).

purchased with the compensation, namely  $w_{t+1}(x)/b_{t+1}$ . Similarly instead of receiving the cash certainty equivalent of nonpecuniary benefit  $\alpha_j$ , which is  $\gamma^{-1} \log \alpha_j$ , the manager receives the one period deferred cash certainty equivalent of nonpecuniary benefit  $\tilde{\alpha}_j$ , which is  $(b_{t+1}/b_t - 1) \log \tilde{\alpha}$ .<sup>19</sup> In other words in our empirical application we use the annuity value of compensation, and the annuity value of the nonpecuniary benefits, to reflect the notion that the managers in our sample spread expenditure from their income over their lifecycle to smooth their consumption.

This way of modeling bond prices yields a precise dynamic interpretation our model: managers sequentially choose their consumption and work choices each period, and the contract we derived for the static model is the long term optimal contract shareholders would offer. More precisely, suppose preferences take the form:

$$-E \left[ \sum_{t=0}^{\infty} \beta^t (l_{0t} + l_{1t} \tilde{\alpha}_1 + l_{2t} \tilde{\alpha}_2) e^{-\tilde{\gamma} c_t} \right] \quad (46)$$

where  $(l_{0t}, l_{1t}, l_{2t}, c_t)$  are the choice variables for each period  $t$ , and  $\beta \in (0, 1)$  is the manager's subjective discount factor. Mirroring the static model,  $l_{0t} \in \{0, 1\}$  is an indicator variable for participation in the firm or taking the outside option in the  $t^{\text{th}}$  period,  $l_{0t} \in \{0, 1\}$  indicates whether the manager shirks or not in that period,  $l_{2t} \in \{0, 1\}$  indicates whether the manager is diligent or not,  $c_t$  is his consumption in period  $t$  and  $l_{0t} + l_{1t} + l_{2t} = 1$  for each period. We now let  $x_t$  denote abnormal profits of the firm received at the end of period  $t$ ,  $f(x_t)$  denote the density of abnormal profits in period  $t$  under diligence, and  $f(x_t)g(x_t)$  the returns under shirking. Reinterpreted within this dynamic setting, the participation constraint in period  $t$ , (4), is Equation (16) of Margiotta and Miller (2000), the incentive constraint (5) is their Equation (18), and the optimal compensation plan (6) is their Equation (21). They also prove the long term contract for the pure moral hazard model in this dynamic framework decentralizes to a sequence of short term contracts that mimic the contract described in Sections 2. In Appendix B we show that the contract derived in Section 3 for the hybrid model has the same dynamic interpretation.

Regarding estimation, Equations (44) treat bond prices as observed variables entering preferences in a restrictive way, thus providing a further source of identification. In the pure moral hazard model, we can substitute from (44) for  $\alpha_1$ ,  $\alpha_2$ , and  $\gamma$  into (12) and (13), to prove that Theorem 2.1 implies:

$$\tilde{\alpha}_1^{1/(b_t-1)} \equiv \frac{1 - E [e^{\tilde{\gamma}/b_{t+1}(w-\bar{w})}]}{E [e^{-\tilde{\gamma}/b_{t+1}w} - e^{-\tilde{\gamma}/b_{t+1}\bar{w}}]} \quad (47)$$

$$\tilde{\alpha}_2^{1/(b_t-1)} \equiv \{E [e^{-\tilde{\gamma}/b_{t+1}w}]\}^{-1} \quad (48)$$

for each period  $t$ . Raising both sides of both equations to the power of  $b_t - 1$  to make  $\tilde{\alpha}_1$  and  $\tilde{\alpha}_2$  the subject of (47) and (48), and then first differencing over  $t$  yields  $T - 1$  further restrictions that aid identification of  $\tilde{\gamma}$ , and hence the other structural parameters. An equivalent set of restrictions, derived in Appendix C, applies to the hybrid model. Intuitively these restrictions

<sup>19</sup>The reason for deferring the cash equivalent one period is to make it comparable to compensation, which is denominated in terms of cash next period.

mean that in the dynamic interpretation of our framework tastes for working, shirking and risk do not vary over time, implying that economic conditions supply the only reason in our models for contracts to vary over time. We view these restrictions as an appealing null hypothesis to test from.

#### 4.4 Measurement Error

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

$$x_{nt} \equiv \tilde{x}_{nt} + \frac{w_{nt}}{V_{n,t-1}} \quad (49)$$

In an optimal contract, compensation depends on  $x_{nt}$ , not  $\tilde{x}_{nt}$ . If  $w_{nt}$  was observed without error then we could get  $x_{nt}$  directly from  $(\tilde{x}_{nt}, w_{nt}, V_{n,t-1})$  and apply the estimator to obtain  $w_{nt}$  for each  $z_{nt}$ , and ignoring dynamic concerns, compute the test statistics described towards the end of Section 2.

However the series we construct on executive compensation,  $w_{nt}$ , is assumed to be measured with error, rendering inconsistent the estimator described in Section 2. Measured compensation, denoted  $\tilde{w}_{nt}$ , is the sum of true compensation  $w_{nt}$  plus an independently distributed disturbance term  $\varepsilon_t$ , assumed orthogonal to the other variables of interest:

$$\tilde{w}_{nt} = w_{nt} + \varepsilon_{nt} \quad (50)$$

Although  $(\tilde{w}_{nt}, \tilde{x}_{nt})$  rather than  $(w_{nt}, x_{nt})$  is observed for each  $(n, t)$ , we can nevertheless construct consistent estimates of  $(w_{nt}, x_{nt})$  from  $(\tilde{w}_{nt}, \tilde{x}_{nt})$  by exploiting a premise of the model that the manager is risk averse under a mild regularity condition, that net abnormal returns to shareholders increase with gross abnormal returns; in other words the manager does not appropriate all the increase in the firm value.

**Lemma 4.1** *For all  $(x_1, x_2) \in R^2$ :*

$$w_{nt} = E_t[\tilde{w}_{nt} | \tilde{x}_{nt}, r_{nt}, V_{n,t-1}] \quad (51)$$

This lemma implies that compensation schedule is the conditional expectation of measured compensation given net abnormal returns and lagged firm size. Pointwise consistent estimates of compensation  $w_{nt}$  can be obtained for each observation with Kernel estimators off successive cross sections. From our estimates of  $w_{nt}$  we then constructed a consistent estimator of the gross abnormal return, which we denote by:

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

We also require an estimate of  $\bar{w}_{st}$  to form estimates of  $\bar{v}_{st}(\gamma) \equiv \exp[-\tilde{\gamma}\bar{w}_{st}/b_{t+1}]$ . We use the fact that although  $\bar{w}_{st}$  is unknown,  $w_{st}(x)$  is a locally non-decreasing function in  $x$

in the limit as  $x \rightarrow \infty$ . Following Brunk (1958), given firm type, for each state  $s \in \{1, 2\}$  and period  $t \in \{1, \dots, T\}$ , we rank the observations on returns in decreasing order by  $x_{st}^{(1)}$ ,  $x_{st}^{(2)}$ ,  $\dots$  and so on, denoting by  $w_{st}^{(1)}$ ,  $w_{st}^{(2)}$ ,  $\dots$  the corresponding (estimated) compensations, and estimate  $\bar{w}_{st}$  with:

$$\bar{w}_{st}^{(N)} \equiv \max_q \sum_{z=1}^q q^{-1} w_{st}^{(z)} \quad (52)$$

Finally we require estimates of  $g_s(x)$ , which we denote by  $g_s^{(N)}(\gamma, x)$ . Note from Theorem 3.1 that  $g_2^{(N)}(\gamma, x)$  can be directly found from  $\bar{w}_{st}^{(N)}$  but that  $g_1^{(N)}(\gamma, x)$  also requires an estimate of  $\bar{h}$ . L'Hospital's rule yields:

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

Ranking excess returns realized in the first state achieved at the end of any period  $t \in \{1, \dots, T\}$ , we obtain the decreasing sequence  $x^{(1)}$ ,  $x^{(2)}$ ,  $\dots$ . Again, following Brunk (1958), we estimated  $\bar{h}$  with:

$$\bar{h}^{(N)} \equiv \max_q \left[ q^{-1} \sum_{t=1}^T \sum_{n=1}^N 1 \left\{ x_{nt}^{(N)} \geq x^{(q)} \right\} 1 \left\{ s_{nt}^{(N)} = 2 \right\} \right] \quad (53)$$

Table 4 provides a cross sectional summary of CEO compensation conditional on the accounting report  $r_{nt}$  based on the manager's hidden information for each firm type. On average compensation is higher when the good state is announced. There is a great deal of dispersion about the sample means. From the numbers of observations in each accounting state  $r_{nt}$  provided in Table 3, we infer that many their differences are significant. By way of contrast, there are no systematic differences between sample mean returns that depend on the publicly observed states. Compensation tends to be higher in companies that are larger on any of the three dimensions we have measured, and also higher in the service sector.

Figure 1 depicts the estimated probability density functions for abnormal returns, and compensation schedules<sup>20</sup>, in each sector for two of the eight observed states,  $(A, W, D) = (S, S, S)$  and  $(L, L, L)$ , and both unobserved states. Referring to Table 3 between 1,686 and 3483 observations are used to construct each graph. The probability density functions for the good state exhibit first order stochastic dominance over the bad. This suggests that accounting measures do anticipate financial performance. Hence a manager conditions on these measures when making her effort choice. It immediately follows that these accounting variables are relevant for analyzing empirical models of moral hazard.

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

<sup>20</sup>A kernel regression of measured compensation,  $\tilde{w}_{nt}$ , on gross abnormal returns,  $x_{nt}$ , accounting returns, our measures of observed heterogeneity, and bond price, shows that true compensation,  $w_{nt}$ , explains about 75% of the variation in measured compensation.

## 5 Estimation and Testing

The equilibrium restrictions imply there is only one primitive parameter in the econometric formulation of the pure and hybrid models, the constant coefficient of absolute risk aversion, denoted by  $\tilde{\gamma}$  in the dynamic version of the model. Our estimation and testing procedures directly exploit restrictions on the risk aversion parameter embodied in the equilibrium conditions modified to account for heterogeneity in firm type, manager type, and interest rate fluctuations that affect the optimal contract, and extended to allow for a panel of firms and managers containing several time periods. As foreshadowed in Section 4.2, we allow  $\tilde{\gamma}$  to vary by the three industrial sectors, the two measures of firm size, and the leverage indicator for the debt to equity ratio, as defined in the text. The basis for allowing  $\tilde{\gamma}$  to vary by type of firm is that heterogeneous firms might be matched with heterogeneous managers with respect to their risk attitudes. We do not impose restrictions across the 24 sector/size/leverage combinations.

This section adapts Section 2.3 to allow for measurement error in compensation and taste parameters that vary over time (for reasons explained in Section 4.3), and reports our empirical findings. We describe the procedures used to derive our econometric estimators and analyze their properties. Then we discuss the confidence regions we found for several variations of the moral hazard model and the most restrictive version of our hybrid moral hazard. Finally we infer the economic implications of our estimates, in terms of potential losses that optimal contracting mitigates, along with the costs of private information and moral hazard incurred to avoid those losses.

### 5.1 Constructing a confidence region

In the text we explain the estimation and testing of the unrestricted pure moral hazard model that only imposes profit maximization. First we characterize the equilibrium restrictions on  $\tilde{\gamma}$  and define a population analogue to the criterion function used in estimation and testing. Then we derive the critical values that determine the confidence region, accounting for the pre-estimation of several parameters that are determined by the probability density functions characterizing excess returns under shirking and working in the two states. Finally we review the subsampling procedures used to compute consistent estimates of the confidence regions. The more restricted versions of the pure moral hazard and hybrid moral hazard models are complicated only by additional or different inequalities and equalities that must be accounted for in estimation. However since they are treated in exactly the same way, we have relegated to Appendix C the extra detail which describes the estimation and testing of those models.

For each state  $s \in \{1, 2\}$  and time  $t \in \{1, 2, \dots, T\}$ , appealing to (44) we define  $Q_{st}(\tilde{\gamma})$  analogously to  $Q_0(\gamma)$ , given by (17) as:

$$\begin{aligned}
 Q_{st}(\tilde{\gamma}) \equiv & E_{st}[x - w^o(x)] - E_{st} \left[ x \frac{e^{\tilde{\gamma}\bar{w}/b_{t+1}} - e^{\tilde{\gamma}w^o(x)/b_{t+1}}}{e^{\tilde{\gamma}\bar{w}/b_{t+1}} - E_{st}[e^{\tilde{\gamma}w^o(x)/b_{t+1}}]} \right] \\
 & + \frac{b_{t+1}}{\tilde{\gamma}(b_t - 1)} \log \left( \frac{1 - E_{st}[e^{\tilde{\gamma}(w^o(x) - \bar{w})/b_{t+1}}]}{E_{st}[e^{-\tilde{\gamma}w^o(x)/b_{t+1}}] - e^{-\tilde{\gamma}\bar{w}/b_{t+1}}} \right) \quad (54)
 \end{aligned}$$

Interpreted either as a static model with time varying risk preferences, or as a dynamic model of the sort described in Appendix B, expected value maximization by the firm implies  $Q_{st}(\tilde{\gamma}) \geq 0$  for  $s \in \{1, 2\}$  and  $t \in \{1, \dots, T\}$ . Extending our definition of  $\Gamma$  given in (18) yields:

$$\tilde{\Gamma} \equiv \left\{ \tilde{\gamma} > 0 : \sum_{t=1}^T \sum_{s=1}^2 \min \{0, Q_{st}(\tilde{\gamma})\}^2 = 0 \right\} \quad (55)$$

Our empirical analysis is based on observing  $N$  firms over  $T$  periods, and the asymptotic properties described here are for large  $N$ .

To test the null hypothesis for the least restricted pure moral hazard model, we form a sample analogue to  $Q_{st}(\tilde{\gamma})$ , denoted by  $Q_{st}^{(N)}(\tilde{\gamma})$ , by substituting in  $w_{st}^{(N)}(x)$  and  $\bar{w}_{st}^{(N)}$  in (54) where appropriate, and replacing the expectations operators in (54) with cross sectional summations that condition on the reported state. Then we define:

$$Q^{(N)}(\tilde{\gamma}) \equiv \sum_{t=1}^T \sum_{s=1}^2 \min \left\{ 0, Q_{st}^{(N)}(\tilde{\gamma}) \right\}^2 \quad (56)$$

and estimate a confidence region for  $\tilde{\Gamma}$  with:

$$\tilde{\Gamma}^{(N)} \equiv \{ \tilde{\gamma} > 0 : N^a Q^{(N)}(\tilde{\gamma}) \leq c_\delta \} \quad (57)$$

where  $N^a$  is the asymptotic rate of convergence of  $Q^{(N)}(\tilde{\gamma})$ , and  $c_\delta$  is the  $\delta$  critical value of the test statistic. We reject the pure moral hazard model at level  $\delta$  if  $\tilde{\Gamma}^{(N)}$  is empty. Under standard regularity conditions  $Q_{st}^{(N)}(\tilde{\gamma})$  converges in probability to  $Q_{st}(\tilde{\gamma})$ . In the paragraphs below we explain how the rate of convergence,  $N^a$ , is determined in our application, and then describe the subsampling procedure used to obtain a consistent estimate of  $c_\delta$ .

The regularity condition about the upper bound  $\bar{x}_{st}$  plays a role in determining the rate of convergence of  $\bar{w}_{st}^{(N)}$  to  $\bar{w}_{st}$ , which in turn determines the rate of convergence of  $Q_{st}^{(N)}(\tilde{\gamma})$ . Suppose there exists a finite  $\bar{x}_{st}$  such that  $F_s(\bar{x}_{st}, z) < 1$  and if  $x > \bar{x}_{st}$ , then  $g_s(x) = 0$ . In that case the derivative of  $w_{st}(x)$  at  $\bar{x}_{st}$  is zero, and following Parsons (1978),  $\bar{w}_{st}^{(N)}$  converges to  $\bar{w}_{st}$  at  $N^{1/2}$ . Although  $w_{st}^{(N)}(x)$  is estimated nonparametrically in the first stage, and converges pointwise at a slower rate than  $N^{1/2}$ , appealing to results in Newey and MacFadden (1994) establishes:

$$N^{1/2} \left[ Q_{st}^{(N)}(\tilde{\gamma}) - Q_{st}(\tilde{\gamma}) \right] \implies N(0, \Omega_{st}(\tilde{\gamma})) \quad (58)$$

for a given  $\tilde{\gamma} > 0$  for some covariance matrix  $\Omega_{st}(\tilde{\gamma})$ . Alternatively, we can relax the assumption about the existence of a finite  $\bar{x}_{st}$ , and assume less restrictively, that  $\lim_{x \rightarrow \infty} g_s(x) = 0$ .

Then as Wright (1981) shows,  $\bar{w}_{st}^{(N)}$  converges to  $\bar{w}_{st}$  at rate  $N^{1/3}$ , and  $N^{1/3}$  replaces  $N^{1/2}$  in (58). Although the regularity assumption about  $\bar{x}_{st}$  does not affect the estimation of the model or the identification results, it affects the convergence rate of the estimates, and the formula for  $\Omega_{st}(\tilde{\gamma})$ . In our model  $a = 1$  if there exists a finite  $\bar{x}_{st}$  such that  $F_s(\bar{x}_{st}) < 1$  with  $g_s(x) = 0$  for all  $x > \bar{x}_{st}$ , but  $a = 2/3$  under the weaker assumption that  $\lim_{x \rightarrow \infty} g_s(x) = 0$ . As a practical matter we adopt the weaker assumption in our empirical work.

To obtain  $c_\delta^{(N)}$ , a consistent estimator for  $c_\delta$ , we modify a subsampling procedure of Chernozhukov, Tamer and Hong (2007). Since several of the components to the test statistic are ill defined because  $v_{st}(x, 0) = 1$  for all  $x$ , the modification bounds the set of  $\gamma$  considered away from zero.

**Algorithm 5.1 (Subsampling)** Consider all subsets of the data with size  $N_b < N$ , where  $N_b \rightarrow \infty$  but  $N_b/N \rightarrow 0$ , and denote the number of subsets by  $B_N$ . Define  $c_0$  and  $\Gamma_0^{(N)}$  as:

$$c_0 \equiv \inf_{\tilde{\gamma} > \tilde{\gamma}_N} [N^a Q^{(N)}(\tilde{\gamma})] + \kappa_N$$

$$\Gamma_0^{(N)} \equiv \{\tilde{\gamma} \geq \tilde{\gamma}_N : N^a Q^{(N)}(\tilde{\gamma}) \leq c_0\}$$

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

$$C^{(j, N_b)} \equiv \sup_{\tilde{\gamma} \in \Gamma_0^{(N)}} [(N_b)^a Q^{(j, N_b)}(\tilde{\gamma})]$$

and denote by  $c_\delta^{(N)}$  the  $\delta$ -quantile of the sample  $\{C^{(1, N_b)}, \dots, C^{(B_N, N_b)}\}$ .

Substituting  $c_\delta^{(N)}$  for  $c_\delta$  in (57), the estimated confidence region for  $\tilde{\gamma}$  is determined by selecting those values for which  $N^a Q^{(N)}(\tilde{\gamma})$  is less than  $c_\delta^{(N)}$ , and we reject the null hypothesis of the unrestricted pure moral hazard model if no such values exist.

## 5.2 Risk Aversion Parameter

Given the joint probability distribution of the data, all the equilibrium restrictions from cost minimization and profit maximization in our structural models of moral hazard can be expressed as a mapping of  $\tilde{\gamma}$ , the coefficient of absolute risk aversion. We apply the estimation and testing methods described in Section 5.1 to the data described in Section 4. Table 4 shows that compensation varies across sector and firm type. Nonpecuniary benefits for both working and shirking might differ across firm type and sector; executives with different abilities, who therefore command different outside options, might select into different firm types and sectors. For these reasons all the models we estimate allow for a full set of interactions between firm and sector type. We also incorporate the aggregate effect of bond prices on the structural parameters through (44). Our empirical analysis investigates whether this degree of (observed) heterogeneity suffices to reconcile either or both models to the data. For example is it necessary to allow tastes for working and shirking to vary with the accounting state of the firm, or with calendar time? Is selection by executives into firm and sector type driven by their attitude towards risk? Tables 5 through 7 presents the 95 percent confidence interval for the risk aversion parameter conditional on the firm type and sector under several variations of the moral hazard models considered.

In the pure moral hazard specification model reported in Table 5 the risk aversion parameter is allowed to vary over firm type and sector, but not over time or by accounting return; the coefficients for working and shirking are permitted to vary with firm type and sector, by accounting return and also across periods in an unrestricted way. However imposing the

restriction that all the variation in the taste parameters for working and shirking over calendar time comes from changes in the bond price has no effect on the confidence regions for  $\tilde{\gamma}$ . We do not reject either version of the pure moral hazard model at the five percent level. While the lower bound of the estimated confidence interval is either 0.01 or 0.02, the upper bound of the confidence interval varies considerably across firm and sector type ranging from 0.21 to 20.1; a wide range of risk attitudes are compatible with the data given this amount of heterogeneity. The intersection of the estimated intervals is  $(0.02, 0.21)$ , nonempty; we cannot reject the hypothesis that a common risk aversion parameter of  $\tilde{\gamma}$  applies to all firm types within all sectors.

In Table 6 we report our results from imposing additional exclusion restrictions that eliminate dependence of the nonpecuniary costs and benefits of working and shirking on the two accounting return states (good and bad). The left panel of Table 6 shows what happens to the estimated identified set of the risk aversion parameters when we impose the restriction that nonpecuniary benefits from shirking are not affected by accounting returns. Providing the risk aversion parameter is permitted to vary with firm and sector type, the pure moral hazard model is not rejected. However there is no common region of overlap for the risk aversion parameter across all 24 firm and sector types, and the model is rejected when we impose the further restriction that the risk aversion parameter is common across firm and sector types.<sup>21</sup> The right panel presents our results from imposing the restriction that the nonpecuniary cost of diligent work is equal across the two accounting return states. Here again we cannot reject the model if the risk aversion parameter is allowed to vary across firm and sector type, but is rejected if we maintain the assumption that the risk aversion of managers does not depend on the type of firm or sector they select into.

Saturating the model with the risk aversion parameter, by allowing it to depend on each of the 24 firm and sector combinations, does not give the pure moral hazard model enough flexibility to accept the hypothesis that abilities and tastes are independent of accounting returns, because the estimated identified set of risk parameters is empty for three of the firm and sector specific cells.<sup>22</sup> Inspecting the columns for the primary and service sectors in the left panel, we reject the hypothesis that a sector specific risk preference parameter can reconcile the data when we impose the additional restriction that the common shirking parameter does not depend on accounting returns; similarly there is no overlap in each the right three columns; we reject the model a working parameter does not depend on accounting returns for a sector specific risk parameter. Performing a similar exercise on each row reveals that for a given firm type, the model rejects a common risk parameter across sectors if we impose the additional joint restrictions that the nonpecuniary benefits of working and shirking are common within a sector. In other words, imposing restrictions on the working and shirking parameters across sectors or across firm types is inconsistent with a common risk aversion parameter in the selected types. Summarizing Table 6, pure moral hazard models that do not permit tastes and abilities to vary with the accounting state are

---

<sup>21</sup>We reach this conclusion comparing the maximum of the lower bounds in each cell with the minimum of the upper bounds in the respective cells, while accounting for the three instances of cells where the set of admissible risk aversion parameters are disjoint.

<sup>22</sup>The three firm sector combinations in which the risk parameter regions do not intersect in their corresponding left and right panels are  $(L, S, L)$  in the Primary sector,  $(S, L, L)$  in the Consumer sector and  $(S, L, S)$  in the Service sector.

rejected.

Table 7 presents the 95 percent confidence interval of the risk aversion parameter set for a restricted hybrid model. Here we assume the preferences for working and shirking do not depend on the accounting return state or calendar time, and we also impose a common risk aversion parameter across firm and sector type. We cannot reject that model at the 5 percent confidence level in any sector. In both the primary and consumer goods sectors the confidence regions for the identified set of risk aversion parameters consists of two intervals, whereas in the services sector there is only one. The bands are quite wide, especially in the primary and consumer goods sectors, evidence of a wide range of risk aversion parameters that reconcile the restricted hybrid model with the data. Moreover since the regions overlap all three sectors for  $\tilde{\gamma} \in (0.37, 0.42)$ , there is no evidence from applying this model to the data that managers with different attitudes towards risk sort into different sectors.

Summarizing the results of Tables 5 and 6 for the pure moral hazard model, we conclude that only if risk preferences are permitted to vary across firm and sector type, preferences for working and shirking depend on the accounting state, can a common risk aversion parameter generate the observations. After adjusting for bond prices to reflect the structure of the dynamics and accommodate aggregate shocks in a parsimonious way, imposing restrictions that preferences do not shift over calendar time is completely innocuous. However imposing restrictions that preferences for working and shirking are not affected by the accounting state shrinks the confidence regions for the risk parameter set; under such restrictions the pure moral hazard model is rejected unless we allows the risk parameter to vary with firm and sector type. These results contrast vividly with those given in Table 7. The hybrid model does not reject the joint restrictions of homogeneous risk preferences and the independence of working and shirking parameters with respect to the accounting state.

Neither the fully unrestricted pure moral hazard model nor the fully restricted hybrid moral hazard model is rejected, which implies they are observationally equivalent. For comparison purposes, suppose managers had a common risk aversion parameter. The upper bound of  $(0.02, 0.21)$ , the intersection of the risk parameter sets for the firm and sector types pure moral hazard model given in Table 5, is less than the lower bound of  $(0.37, 0.42)$ , the intersection of the sector types for the hybrid model in Table 7. Our measure of compensation units is in millions of dollars. Thus a manager with risk-aversion parameter between 0.02 and 0.21 would be willing to pay between \$8,849 and \$92,390 to avoid a gamble that has an equal probability of losing or winning one million dollars and a manager with risk aversion parameter between 0.37 and 0.42 would be willing to pay between \$160,870 and \$181,710 to avoid the same gamble.

### 5.3 Certainty Equivalent Wages

For the purposes of identification and estimation, the parameters of both models can be concentrated to just one primitive,  $\tilde{\gamma}$ , but from an economics standpoint, the estimates of the preferences for working and shirking,  $\tilde{\alpha}_2$  and  $\tilde{\alpha}_1$  (induced by  $\tilde{\gamma}$  in estimation) are also informative about the plausibility of the models. Along with  $\tilde{\gamma}$  and the bond price,  $b_t$ , these parameters determine the nonpecuniary costs of effort that offset pecuniary compensation. In particular, the manager's reservation wage to shirk (a certainty equivalent) is given by the expression  $b_{t+1} \ln(\tilde{\alpha}_1) / \tilde{\gamma} (b_t - 1)$ , which is the dynamic extension of the formula for

the shirking wage given in Lemma 2.1. Similarly  $b_{t+1} \ln(\tilde{\alpha}_2) / \tilde{\gamma} (b_t - 1)$  is the manager's reservation wage to work (derived in an analogous manner); this is the equilibrium wage the firm would pay the manager to work if his actions were observed. In both cases a negative reservation wage means that in equilibrium the manager would pay shareholders for the privilege of holding the job, presumably because they enjoy wielding power and the other perquisites of executive life. Finally the difference between the two is the manager's compensating differential from shirking versus working.

We computed 95 percent confidence regions of the identified sets for the unrestricted pure moral hazard model with a common risk aversion parameter and for the restricted hybrid model with a common risk aversion parameter. These regions are determined by substituting corresponding regions for the risk aversion parameters into the formulae for  $\tilde{\alpha}_1$  and  $\tilde{\alpha}_2$ . For example using (13) and (44) we calculate estimates of:

$$b_{t+1} \ln(\tilde{\alpha}_2) / \tilde{\gamma} (b_t - 1) = -b_{t+1} \ln E \left[ e^{-\tilde{\gamma} w^o(x) / b_{t+1}} \right] / \tilde{\gamma}$$

for all  $\tilde{\gamma} \in (0.02, 0.21)$  by firm and sector type to determine the estimated the reservation wage to work in the pure moral hazard model. Table 8 presents the estimated identified set of the manager's reservation wages, to shirk and work, for both the unrestricted pure and the restricted hybrid models at the median bond price. Because the unrestricted model has a reservation wage for each accounting state, but nonpecuniary benefits do not vary by state in restricted models, there are twice as many regions for the pure moral hazard model to report as for the hybrid.

The top panel of Table 8 presents the estimated identified set of the manager's reservation wage to shirk for both models. The confidence interval is quite small in all the cells; the difference between the upper and lower bound is usually less than \$20,000. This is because, given the probability distribution of our data, the formula for the shirking wage is not very sensitive to the risk aversion parameter. In the pure moral hazard model the shirking wage is always higher in the good state than in the bad. The differences between the states are invariably more than a million dollars, another reminder that the only way to reconcile our data with a pure moral hazard model sporting a common risk aversion parameter is to assume reservation wages differ by accounting state. In 18 out of 24 firm and sector types the hybrid shirking wage lies between the two for the pure moral hazard shirking wages, while in the remaining 6 the hybrid shirking wage is less than the region for the shirking wage in the bad state of the pure moral hazard.<sup>23</sup> In the pure moral hazard models the shirking wage is negative for more than half of the firm types in the bad state, but in the good state, the manager would demand positive compensation to shirk in 22 out of 24 firm and sector types. About half the shirking reservation wages in the hybrid model are positive, and because the estimated regions typically lie between the pure moral hazard shirking wages for the two states the magnitudes for the hybrid are lower, and in our opinion, more plausible.

The bottom panel of Table 8 presents the identified set of the (certainty equivalent) reservation wages for working. In the pure moral hazard model the certainty equivalent compensation is negative for 9 out of 24 firm type in the bad state. This striking finding is not surprising. As shown in Table 4 the same 9 firm and sector types have negative average

---

<sup>23</sup>Three of the six exceptions occur in the  $(S, S, S)$  firm type, that is for all three sectors.

compensation. Since the difference between expected compensation and its certainty equivalent is the risk premium, the former must be higher than the latter for risk averse agents. One way of rationalizing why a manager is willing to pay the firm to work in bad accounting states is to argue that in such states firms diversify from their core competencies to make themselves attractive to managers as a revenue source. We are skeptical this explanation applies to our firm population, and prefer to interpret this finding as evidence against the pure moral hazard model. There is no reason to resort to this implausible explanation to rationalize the hybrid model. Although 4 out of 24 of the firm types have a negative lower bound, the confidence region under the hybrid model always has an interval containing only positive numbers.

## 5.4 Agency Costs

To gauge the importance of agency in our data we use two measures. The first is the expected gross loss shareholders would incur from the manager shirking. The second is the risk premium, which measures the expected extra compensation paid to managers because of the agency problem.

Table 9 depicts estimates of both measures for the pure and hybrid moral hazard models. The first measure, denoted by  $\Delta_1$ , is the expected gross output loss to the firm switching from the distribution of abnormal returns for diligent work to the distribution for shirking, that is the difference between the expected output to the plant from the manager pursuing the firm's goals versus his or her own, before netting out expected managerial compensation given the state (note that in hybrid it is given the state is reported truthfully). It is a function of the likelihood ratio of abnormal returns from shirking versus working. In symbols:

$$\Delta_1 \equiv E \{x [1 - \tilde{g}_s(x)]\}$$

where the expectation is over  $(x, s)$ . This was computed by numerically integrating over  $x$  where appropriate after substituting the 95 percent confidence region for the risk aversion parameters into the appropriate formulae for  $\tilde{g}_s(x)$ .

The top panel of Table 9 presents our estimated set of gross losses for both the pure and hybrid model hazard model. The estimated confidence intervals of gross losses are very tight irrespective of the model specifications; for example the average length of the confidence interval range from 0.86 percent to 1.73 percent for the pure moral hazard model and 0.68 percent to 1.04 percent for the hybrid model. The differences between the two model specifications are small relatively to relative the variation over firm type using any reasonable of distance. For example the median minimum distance between the confidence intervals for the pure and hybrid moral hazard model is 0.32 percent in the primary sector, 1.74 percent in the service sector, and 1.49 percent in the service sector. Meanwhile the variation of the confidence interval across firm type is of several order of magnitude larger; for the pure moral hazard it range between 7.84 percent and 14.89 percent and for the hybrid model between 17.35 percent and 24.57 percent depending on sector. Therefore heterogeneity over firm type is much more important for the estimates of the gross losses than which model specification is used. It is worth noting however that the variation across firm type is higher hybrid specification than under the pure moral hazard specification. This is further illustrated by

applying our measures of market value of the firms to our estimates of  $\Delta_1$  after integrating out firm type; this is done by taking the average over firm type of the bounds of the confidence intervals and then multiplying these average bounds by the average market value in Table 2. For the pure moral hazard model the average loss to the firm varies from \$545 million to \$601 million in the primary sector, \$918 million to \$1.00 billion in the consumer goods sector, and \$1.46 billion to \$1.66 billion in the services sector per year. For the hybrid model it varies from \$580 million to \$648 million in the primary sector, \$1.00 billion to \$1.07 billion in the consumer goods sector, and \$1.42 billion to \$1.49 billion in the services sector per year. Clearly the estimated dollar value of gross losses between the two specifications is minor compared to their overall magnitudes. Another way of making this point is to note that the average stock market return over this period was roughly 10 percent per annum, so that for more than half the firm and sector types in both specifications, the expected gross return would have been negative if shareholders had ignored the moral hazard.

The second measure, the risk premium, is denoted by  $\Delta_2$ . It shows how much the firm would be willing to pay to eliminate the moral hazard problem. It is a function of the risk aversion parameter, the nonpecuniary utility loss from working, and bond prices. Under a perfect monitoring scheme shareholders would pay the manager the fixed wage of  $b_{t+1} \ln(\tilde{\alpha}_2) / \tilde{\gamma}(b_t - 1)$ . Hence the expected value of a perfect monitor to shareholders is the difference between expected compensation under the current optimal scheme and its certainty equivalent, which we average over the time periods in the sample:

$$\Delta_2 = \frac{1}{T} \sum_{t=1}^T \{E[w_{st}(x) - b_{t+1} \ln(\tilde{\alpha}_{2s}) / \tilde{\gamma}(b_t - 1)]\} \quad (59)$$

The bottom panel of Table 9 presents our estimates of the identified set for the risk premium. We find the risk premium is higher in the hybrid than in the pure moral hazard model for every firm type, by several hundred thousand dollars. Evidently the direct effect of the higher estimates of risk aversion for the hybrid model on (59) outweigh its indirect effects transmitted through work preferences in (13) and (31).

Despite the quantitative differences between the pure and hybrid models, most of the qualitative comparisons between firm types and industry sectors match up. For both the hybrid and pure moral hazard specifications, after conditioning on firm type, the risk premium is lower in the primary than the consumer sector with just one exception ( $L, S, L$ ). In both specifications the risk premium for consumer sector is generally lower than for the service sectors.<sup>24</sup> Controlling for assets and employment, all firm types with a higher debt to equity have a lower risk premium than their counterparts with a lower debt equity ratio in both the pure and hybrid models. Thus managers are more uncertain about their compensation, attributable in our framework to moral hazard and hidden information, when the population distribution of stakeholder claims to the firm's assets is tilted towards those who are most affected by firm performance. As a rule the CEO of a firm employing more workers is usually paid a higher risk premium, given total asset and the debt equity category. Now only does this hold for both the pure and hybrid specification; the two exceptions to this rule, which occur in primary sector, ( $L, L, L$ ) versus ( $L, S, L$ ) and ( $L, L, S$ ) versus ( $L, S, S$ ) occur for both specifications as well. The relationship between firms assets and the risk premium is

---

<sup>24</sup>Firm type ( $S, L, L$ ) is an exception for both specifications, and ( $S, S, L$ ) is a second exception for the pure moral hazard model.

somewhat weaker, but generally speaking, higher firm assets are associated with a higher risk premium.

Finally our findings for the risk aversion parameter set for the hybrid model, the risk premium, and the losses that would be incurred by the firms if they ignored moral hazard, are quite close to those found for pure moral hazard model by Margiotta and Miller (2000), Gayle and Miller (2009) and Gayle, Golan, and Miller (2011), although these papers uses different estimation methods and data from industrial sectors, executive ranks, and time periods.<sup>25</sup> Thus if accounting returns are treated as hidden information in a hybrid model of moral hazard, or if accounting information is ignored by integrating out those states, estimates obtained from structural models of moral hazard applied to executive compensation seems robust to a variety of econometric techniques and data sources. But if accounting data is used in the estimation, the estimates of the social cost of moral hazard are quite sensitive to assumptions about hidden information. Our results favor the hybrid model, which treats accounting data as unverifiable information that shareholders value because the manager has incentivizes to reveal his knowledge in the optimal contract.

## 6 Conclusion

If every piece of information a manager knows about his or her firm is codified and independently verifiable in a court of law, managers can be compelled to reveal all their privy information through the firm's accounting records. In that case a multistate pure moral hazard model would apply, different states being distinguished by distinct records. Within the current legal system, however, managers exercise considerable discretion about how much information they release describing the state of their own firms. If the penal code for accounting protocol was augmented by incentives embedded in managerial compensation designed to elicit truthful revelation, a hybrid model of moral hazard would apply.

Our empirical investigation is based on a large panel data set measuring compensation of chief executive officers, financial and accounting returns, as well as size and sector background characteristics of the publicly trade firms they manage. In the pure moral hazard models we estimate and test, managers do not have discretion about how they report accounting returns. In the hybrid model, we interpret data on accounting returns as information reported by the CEO that cannot be fully corroborated by shareholders. Thus our empirical study compares and contrasts the role of these alternative information assumptions about accounting returns within competing models of moral hazard.

We derive the equilibrium restrictions from optimal contracting to predict the shape of the compensation schedule when there are only hidden actions (pure moral hazard), and when there is hidden information as well (hybrid). These restrictions fully characterize the empirical content of our models. We establish sharp and tight bounds for the risk aversion

---

<sup>25</sup>Margiotta and Miller (2000) use data for from a subset of the primary sector for the period 1944-1979; Gayle and Miller (2009) compared results from this period of time with later data for the period 1993-2004. Both studies use a fully parametric model without and deploy a nested, fixed-point, full-solution estimation technique to identify and estimate the risk attitude. The confidence interval for the risk aversion parameter from Gayle, Golan, and Miller (2011) covers the estimated identified set for the hybrid model, yet they estimate a dynamic moral hazard model with human capital accumulation and sorting over firm and ranks.

parameter, and show that all the other parameters can be expressed as mappings of the risk aversion parameter and probability distribution of the data generating process, for which we have sample analogues. Our estimation and testing procedures are based on inferring the bounds of the risk aversion parameter. The benchmark static model of moral hazard is only partially identified, because every risk parameter that satisfies just one inequality derived from profit maximization generates an observationally equivalent model. The identified set of risk aversion parameters shrinks as we add constraints that impose exclusion restrictions to limit the scope of heterogeneity in multistate models of moral hazard. However point identification is by no means assured, either in theory, or in the confidence intervals for the identified risk aversion parameter sets we estimate.

The pure moral hazard model with homogenous preferences is rejected, but the hybrid moral hazard model with homogenous preferences is not. It is observationally equivalent to a pure moral hazard model with heterogeneous preferences. Yet the hybrid model provides a more satisfying economics explanation of the data than the heterogenous pure moral hazard model.

The data show that expected compensation for next period increases with current accounting returns, and also that the gradient of compensation in financial returns is higher when the accounting return has been greater. The hybrid model predicts that the expected utility of the agent is higher in the firm's good state than its bad state. Moreover to induce truth telling and report higher earnings when the firm's prospects are good, the principal lowers and flattens the schedule when the agent reports the bad state, reducing expected compensation and making realized compensation less dependent on the outcome. In our application this permits financial and accounting returns data to play bigger roles in explaining compensation. Relatively high estimated values of the risk parameter, which are consistent with previous work on pure moral hazard models that do not exploit the accounting data, reduce the certainty equivalent of compensation in the good accounting state. These features reconcile the hybrid model to the data even when tastes for working and risk attitudes are not allowed to vary with the firm's accounting state.

In contrast to the hybrid model, the pure moral hazard model equalizes expected utility across states. The heterogeneous pure moral hazard model mitigates the effects of curvature differences in compensation schedules across states, by making the managers appear almost risk neutral, and simultaneously attributing to nonpecuniary benefits the differences in expected compensation across accounting states. The risk parameter in the heterogenous pure moral hazard model is considerably lower than previous findings for pure moral hazard models that do not exploit differences in accounting states. The nonpecuniary benefits from working for the firm in the bad accounting state are so high that the estimated certainty equivalent compensation is negative. But unless work preferences or risk attitudes differ across accounting states, the pure moral hazard framework lacks the degrees of freedom necessary to fit the differently shaped compensation schedules.

## A Proofs of Theorems and Lemmas

### Proof of Lemma 2.1.

We define  $v(x) \equiv \exp[-\gamma w(x)]$  and note that the participation constraint can be ex-

pressed as:

$$\alpha_2 E[v(x)] \leq 1 \quad (60)$$

Similarly the incentive compatibility constraint for work can be expressed as:

$$\alpha_2 E[v(x)] \leq \alpha_1 E[v(x)g(x)] \quad (61)$$

To minimize expected compensation subject to (60) and (61), we choose  $v(x)$  to maximize:

$$E\{\log[v(x)]\} + \theta_0 E[1 - \alpha_2 v(x)] + \theta_1 E[\alpha_1 g(x)v(x) - \alpha_2 v(x)] \quad (62)$$

The first order condition is given by:

$$v(x)^{-1} = \theta_0 \alpha_2 + \theta_1 \alpha_2 - \theta_1 \alpha_1 g(x) \quad (63)$$

Multiplying through by  $v(x)$  and taking expectations yields:

$$1 = \theta_0 \alpha_2 E[v(x)]$$

since the complementary slackness condition for incentive compatibility implies:

$$\theta_1 E[\alpha_1 g(x)v(x) - \alpha_2 v(x)] = 0$$

Substituting 1 for  $\theta_0 \alpha_2 E[v(x)]$  in the complementary slackness condition for participation proves that  $\theta_0 = 1$  and consequently:

$$\alpha_2 E[v(x)] = 1 \quad (64)$$

Thus the first order condition simplifies to:

$$\begin{aligned} v(x)^{-1} &= \alpha_2 + \theta_1 \alpha_2 - \theta_1 \alpha_1 g(x) \\ &= \alpha_2 [1 + \theta (\alpha_2 / \alpha_1) - \theta g(x)] \end{aligned} \quad (65)$$

where  $\theta \equiv \theta_1 \alpha_1 / \alpha_2$ . Substituting for  $v(x) \equiv \exp[-\gamma w(x)]$  and taking logarithms then yields (6), the optimal compensation equation for work. A contradiction argument establishes the incentive compatibility constraint holds with equality too. Substituting Equation (65) into the incentive compatibility condition and imposing equality gives the solution to  $\theta$ , namely (7). Finally the optimal contract for shirking is found by setting  $\theta_1 = 0$  and substituting  $\alpha_1$  for  $\alpha_2$  in (62) and solving for  $\gamma$  from the first order condition to obtain  $\gamma^{-1} \log(\alpha_1)$ . ■

**Proof of Theorem 2.1.** Upon substituting  $\gamma^*$  for  $\gamma$ , Equation (13), the expression for  $\alpha_2(\gamma^*)$  follows directly from (64). Rearranging Equation (10) yields:

$$e^{\gamma^* \bar{w}} = \alpha_2 [1 + \theta (\alpha_2 / \alpha_1)] \quad (66)$$

Subtracting Equation (65) from (66) we obtain:

$$\alpha_2 \theta g(x) = e^{\gamma^* \bar{w}} - e^{\gamma^* w^o(x)} \quad (67)$$

Appealing to Equation (65):

$$E [e^{\gamma^* w^o(x)}] = \alpha_2 [1 + \theta (\alpha_2 / \alpha_1) - \theta] \quad (68)$$

Subtracting Equation (68) from (66) we obtain:

$$\alpha_2 \theta = e^{\gamma^* \bar{w}} - E [e^{\gamma^* w^o(x)}] \quad (69)$$

Substituting for  $\alpha_2 \theta$  using (69) in (67) and making  $g(x)$  the subject of equation yields the expression for  $g(x, \gamma^*)$  given in (11). Substituting for  $\alpha_2 \theta$  using (69), and also for  $\alpha_2$  using (64), in Equation (68), yields upon rearrangement the expression for  $\alpha_1(\gamma^*)$  given in (12). ■

**Proof of Theorem 2.2.** Suppose there exists  $\gamma > 0$  such that  $Q_0(\gamma) \geq 0$ . Let  $v(x_n, \gamma) \equiv \exp[-\gamma w_n]$  and set  $\bar{v}(\gamma) \equiv \exp[-\gamma \bar{w}]$ . Since the objective function in (62) is strictly concave, and the constraints are linear, the first order and complementary slackness conditions in this Kuhn Tucker formulation uniquely determine the solution to the optimal contract. We prove the theorem by showing that  $v(x, \gamma)$  satisfies the first order conditions for the Lagrangian (62), and that the complementary slackness conditions are satisfied when the Kuhn Tucker multipliers, denoted by  $\theta_0(\gamma)$  and  $\theta_1(\gamma)$ , are defined as:

$$\theta_0(\gamma) \equiv E [v(x, \gamma)]^{-1} \quad (70)$$

and

$$\theta_1(\gamma) \equiv \bar{v}(\gamma)^{-1} - E [v(x, \gamma)^{-1}] \quad (71)$$

From their respective definitions both  $\theta_0(\gamma)$  and  $\theta_1(\gamma)$  are strictly positive.

1. Appealing to the definitions of  $\alpha_1(\gamma)$ ,  $\alpha_2(\gamma)$  and  $g(x, \gamma)$  given in (11) through (13):

$$g(x, \gamma) - \frac{\alpha_2(\gamma)}{\alpha_1(\gamma)} = \frac{\bar{v}(\gamma)^{-1} - v(x, \gamma)^{-1}}{\bar{v}(\gamma)^{-1} - E [v(x, \gamma)^{-1}]} - \left\{ \frac{\bar{v}(\gamma)^{-1} - E [v(x, \gamma)]^{-1}}{\bar{v}(\gamma)^{-1} - E [v(x, \gamma)^{-1}]} \right\}$$

Multiplying both sides by  $\theta_1(\gamma)$  defined in (71) we obtain:

$$\{\bar{v}(\gamma)^{-1} - E [v(x, \gamma)^{-1}]\} \left\{ g(x, \gamma) - \frac{\alpha_2(\gamma)}{\alpha_1(\gamma)} \right\} = E [v(x, \gamma)]^{-1} - v(x, \gamma)^{-1} \quad (72)$$

Substituting  $\theta_0(\gamma)$  for  $E [v(x, \gamma)]^{-1}$  from (70) yields the first order condition for  $v(x, \gamma)$  given in (63).

2. From the definition of  $\alpha_2(\gamma)$  in (13) the participation constraint in (60) is met with equality. From (70),  $\theta_0(\gamma)$  is positive. Therefore the complementary slackness condition for participation is satisfied. Noting from (71) that  $\theta_1(\gamma)$  is positive, it follows

from the definitions of  $\alpha_1(\gamma)$ ,  $\alpha_2(\gamma)$  and  $g(x, \gamma)$  given in (11) through (13) that:

$$\begin{aligned} \theta_1(\gamma) E \left\{ \left[ g(x, \gamma) - \frac{\alpha_2(\gamma)}{\alpha_1(\gamma)} \right] v(x, \gamma) \right\} &= E \{ [E[v(x, \gamma)]^{-1} - v(x, \gamma)^{-1}] v(x, \gamma) \} \\ &= 0 \end{aligned}$$

Therefore the complementary slackness condition for incentive compatibility is satisfied.

■

**Proof of Lemma 2.2.** There are three steps. First we show that for all  $\gamma > 0$ :

$$E[w^o(x)] > \gamma^{-1} \log \left\{ \frac{1 - E[e^{\gamma w^o(x) - \gamma \bar{w}}]}{E[e^{-\gamma w^o(x)}] - e^{-\gamma \bar{w}}} \right\} \quad (73)$$

Then we show that if  $\text{cov}\{x, e^{\gamma w^o(x)}\} < 0$ , then:

$$E[x] < E \left[ x \frac{e^{\gamma \bar{w}} - e^{\gamma w^o(x)}}{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]} \right] \quad (74)$$

Finally we construct a joint distribution for  $(x, w)$  in which the covariance is negative. Upon combining the inequalities the lemma now follows from the definition of  $Q_0(\gamma)$  given in (17).

1. Since  $e^w$  is convex in  $w$ , Jensen's inequality implies

$$E[e^{-\gamma w^o(x)}] > e^{-E[\gamma w^o(x)]}$$

Taking logarithms of both sides, dividing through by  $\gamma$  and rearranging yields:

$$E[w^o(x)] + \gamma^{-1} \ln E[e^{-\gamma w^o(x)}] > 0 \quad (75)$$

But from (13) and the discussion following (14):

$$\begin{aligned} E[w^o(x)] + \gamma^{-1} \ln E[e^{-\gamma w^o(x)}] &= E[w^o(x)] - \gamma^{-1} \ln \alpha_2(\gamma) \\ &< E[w^o(x)] - \gamma^{-1} \ln \alpha_1(\gamma) \end{aligned} \quad (76)$$

Combining Inequalities (75) and (76) we obtain (73) upon substituting in the expression for  $\alpha_1(\gamma)$  given by (12).

2. Suppose  $\text{cov}\{x, e^{\gamma w^o(x)}\} < 0$ . Since  $e^{\gamma \bar{w}} > E[e^{\gamma w^o(x)}]$  for all positive  $\gamma$ , it now follows that:

$$E[x] - E \left[ x \frac{e^{\gamma \bar{w}} - e^{\gamma w^o(x)}}{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]} \right] = \frac{\text{cov}\{x, e^{\gamma w^o(x)}\}}{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]} < 0$$

3. Suppose the probability density function for  $x$  is symmetric, let  $E[x] = 0$  and  $w^o(x) =$

– $x$ . Then:

$$\begin{aligned} \text{cov}(x, \exp[\gamma w^o(x)]) &= E[xe^{-\gamma x}] \\ &= E[x(e^{-\gamma x} - e^{\gamma x}) | x > 0] / 2 \\ &< 0 \end{aligned}$$

■

**Proof of Lemma 3.1.** Multiplying each first order equation in the text by  $\varphi_s v_s(x) f_s(x)$ , then summing and integrating over  $x$  yields:

$$1 = \eta_0 \left[ \sum_{s=1}^2 \int_x^\infty \varphi_s v_s(x) f_s(x) dx \right] \equiv \eta_0 E[v_s(x)]$$

where we make use of the complementary slackness conditions. Substituting for  $\eta_0 = E[v_s(x)]^{-1}$  into the complementary slackness condition for participation then gives the first numbered item in the lemma. Multiplying the first order conditions for the second state by  $v_2(x)$ , after solving for  $\eta_0$  we obtain:

$$1 = E[v_s(x)]^{-1} v_2(x) + \eta_3 v_2(x) + \eta_2 v_2(x) [(\alpha_2/\alpha_1) - g_2(x)] + \eta_4 v_2(x)$$

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

$$\begin{aligned} 1 &= E[v_s(x)]^{-1} E_2[v_2(x)] + \eta_3 E_2[v_2(x)] + \eta_4 E_2[v_2(x)] \\ &= E_2[v_2(x)] (E[v_s(x)]^{-1} + \eta_3 + \eta_4) \end{aligned}$$

Dividing through by  $E_2[v_2(x)]$  proves the second numbered item in the lemma. ■

**Proof of Theorem 3.1.** Let  $v_s(x) \equiv \exp[-\gamma^* w_s(x)]$  and  $\bar{v}_s \equiv \exp[-\gamma^* \bar{w}_s]$ . We prove the theorem by treating each component successively. Upon substituting  $\gamma^*$  for  $\gamma$ :

1. Since the participation constraint is met with equality in the optimal contract:

$$\alpha_2 = E[v_s(x)] = \hat{\alpha}_2(\gamma^*)$$

2. Substituting the solution for  $\eta_0$  into the first order condition for the second state yields:

$$v_2(x)^{-1} = E[v_s(x)]^{-1} + \eta_2 [(\alpha_2/\alpha_1) - g_2(x)] + \eta_3 + \eta_4$$

Taking expectations we obtain:

$$E_2[v_2(x)^{-1}] = E[v_s(x)]^{-1} + \eta_2 [(\alpha_2/\alpha_1) - 1] + \eta_3 + \eta_4$$

Also:

$$\bar{v}_2^{-1} = E[v_s(x)]^{-1} + \eta_2(\alpha_2/\alpha_1) + \eta_3 + \eta_4$$

Differencing the second two equations:

$$\eta_2 = \bar{v}_2^{-1} - E_2 [v_2(x)^{-1}] = \eta_2(\gamma^*)$$

3. Proving  $g_2(x) = g_2(x, \gamma^*)$  comes from differencing:

$$v_2(x)^{-1} = E[v_s(x)]^{-1} + \eta_2[(\alpha_2/\alpha_1) - g_2(x)] + \eta_3 + \eta_4$$

from:

$$\bar{v}_2^{-1} = E[v_s(x)]^{-1} + \eta_2(\alpha_2/\alpha_1) + \eta_3 + \eta_4$$

to give:

$$\bar{v}_2^{-1} - v_2(x)^{-1} = \eta_2 g_2(x)$$

Upon rearrangement, we appeal to the result in Item 2, that  $\eta_2 = \eta_2(\gamma^*)$  to obtain:

$$g_2(x) = \eta_2^{-1} [\bar{v}_2^{-1} - v_2(x)^{-1}] = g_2(x, \gamma^*)$$

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

$$\bar{v}_2^{-1} = E[v_s(x)]^{-1} + \{\bar{v}_2^{-1} - E_2[v_2(x)^{-1}]\}(\alpha_2/\alpha_1) + \eta_3 + \eta_4$$

or, upon appealing to Lemma 3.1:

$$(\alpha_2/\alpha_1) = \frac{\bar{v}_2^{-1} - E[v_s(x)]^{-1} - \eta_3 - \eta_4}{\bar{v}_2^{-1} - E_2[v_2(x)^{-1}]} = \frac{\bar{v}_2^{-1} - E[v_2(x)]^{-1}}{\bar{v}_2^{-1} - E_2[v_2(x)^{-1}]}$$

Making  $\alpha_1$  the subject of the equation:

$$\alpha_1 = \alpha_2 \left[ \frac{\bar{v}_2^{-1} - E[v_2(x)]^{-1}}{\bar{v}_2^{-1} - E_2[v_2(x)^{-1}]} \right] = \alpha_1(\gamma^*)$$

5. To prove  $\eta_4 = \eta_4(\gamma^*)$  we first multiply the first order conditions for the first state by  $v_1(x)$ , after solving for  $\eta_0(\gamma) \equiv E[v_s(x, \gamma)]^{-1}$  to obtain:

$$\begin{aligned} & 1 - \eta_1 v_1(x) [(\alpha_2/\alpha_1) - g_1(x)] \\ = & E[v_s(x)]^{-1} v_1(x) - \eta_3 v_1(x) h(x) - \eta_4 (\alpha_1/\alpha_2) v_1(x) g_2(x) h(x) \end{aligned}$$

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

$$1 = \{E[v_s(x)]\}^{-1} E_1[v_1(x)] - \eta_3 E_1[v_1(x) h(x)] - \eta_4 (\alpha_1/\alpha_2) E_1[v_1(x) g_2(x) h(x)]$$

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

$$\eta_3 = \{E_2[v_2(x)]\}^{-1} - E[v_s(x)]^{-1} - \eta_4$$

we obtained from Lemma 3.1 reduces this expression to:

$$1 = E[v_s(x)]^{-1} E_1[v_1(x)] - \eta_4(\alpha_1/\alpha_2)E_1[v_1(x)g_2(x)h(x)] \\ - \{E_2[v_2(x)]^{-1} - E[v_s(x)]^{-1} - \eta_4\} E_1[v_1(x)h(x)]$$

Upon collecting terms:

$$\eta_4 \{(\alpha_1/\alpha_2)E_1[v_1(x)g_2(x)h(x)] - E_1[v_1(x)h(x)]\} \\ = \{E[v_s(x)]\}^{-1} E_1[v_1(x)] - E_1[v_1(x)h(x)] \{E_2[v_2(x)]^{-1} - E[v_s(x)]^{-1}\} - 1$$

so solving for  $\eta_4$  we now have:

$$\eta_4 = \frac{E[v_s(x)]^{-1} E_1[v_1(x)] - E_1[v_1(x)h(x)] \{E_2[v_2(x)]^{-1} - E[v_s(x)]^{-1}\} - 1}{(\alpha_1/\alpha_2)E_1[v_1(x)g_2(x)h(x)] - E_1[v_1(x)h(x)]} = \eta_4(\gamma^*)$$

6. Proving  $\eta_3 = \eta_3(\gamma^*)$  follows directly from Lemma 3.1, which implies:

$$\eta_3 \equiv E_2[v_2(x)]^{-1} - \eta_4(\gamma^*) - E[v_s(x)]^{-1}$$

7. To prove  $\eta_1 = \eta_1(\gamma^*)$ , rewrite the first order condition for the first state as:

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

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

$$\eta_1(\alpha_2/\alpha_1) = \bar{v}_1^{-1} - E[v_s(x)]^{-1} + \eta_3 \bar{h} \quad (77)$$

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

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

$$\eta_1 g_1(x) = \bar{v}_1^{-1} - v_1(x)^{-1} + \eta_3 [\bar{h} - h(x)] - \eta_4(\alpha_1/\alpha_2)g_2(x)h(x)$$

Dividing both sides by  $\eta_1$  we thus establish  $g_1(x) = g_1(x, \gamma^*)$ .

■

**Proof of Theorem 3.2.** The proof follows the same steps as the proof to Theorem 2.2. First we define some candidate values for the Kuhn Tucker multipliers, as a function of  $\gamma$  and establish they are positive. Then we show that if  $\gamma \in \hat{\Gamma}$  the first order conditions for the optimization problem in (29) are satisfied in both states. Finally we demonstrate the complementary slackness conditions are also satisfied. Since the objective function for the underlying maximization problem is strictly concave, and the constraints are linear, the first order and complementary slackness conditions in the Kuhn Tucker formulation uniquely determine the solution to the optimal contracting problem, thus proving the theorem.

1. Let  $\eta_0(\gamma) \equiv E[v_s(x, \gamma)]^{-1}$  and:

$$\eta_2(\gamma) \equiv \bar{v}_2(\gamma)^{-1} - E_2[v_2(x, \gamma)^{-1}]$$

Along with (37), (36) and (35), these equations defines candidate values for the five Kuhn Tucker multipliers for the hybrid model for the  $\gamma$  parameterization. By inspection both  $\eta_0(\gamma)$  and  $\eta_2(\gamma)$  are strictly positive. Also  $\eta_j(\gamma) \geq 0$  for  $j \in \{1, 3, 4\}$  from the construction of  $\widehat{\Gamma}$ .

2. From the definitions of  $\widehat{\alpha}_1(\gamma)$ ,  $\widehat{\alpha}_2(\gamma)$ ,  $\widehat{g}_2(x, \gamma)$  and  $\eta_2(\gamma)$  it follows that:

$$\eta_2(\gamma) \left[ \widehat{g}_2(x, \gamma) - \frac{\widehat{\alpha}_2(\gamma)}{\widehat{\alpha}_1(\gamma)} \right] = E_2[v_2(x, \gamma)]^{-1} - v_s(x, \gamma)^{-1}$$

From the definition of  $\eta_3(\gamma)$  we have:

$$E[v_s(x, \gamma)]^{-1} + \eta_3(\gamma) + \eta_4(\gamma) = E_2[v_2(x, \gamma)]^{-1}$$

Subtracting the first equation from the second and substituting  $\eta_0(\gamma)$  for  $E[v_s(x, \gamma)]^{-1}$  we obtain the first order condition for the second state in the hybrid model given by the second line of (30). Turning to the first state, the definition of  $\widehat{g}_1(x, \gamma)$  implies:

$$\begin{aligned} \eta_1(\gamma) \widehat{g}_1(x, \gamma) &= \bar{v}_1(\gamma)^{-1} - v_1(x, \gamma)^{-1} + \eta_3(\gamma) [\bar{h} - h(x)] \\ &\quad - \eta_4(\gamma) \widehat{g}_2(x, \gamma) h(x) \frac{\widehat{\alpha}_2(\gamma)}{\widehat{\alpha}_1(\gamma)} \end{aligned}$$

From the definition of  $\eta_1(\gamma)$ :

$$\eta_3(\gamma) \bar{h} = \eta_1(\gamma) \frac{\widehat{\alpha}_2(\gamma)}{\widehat{\alpha}_1(\gamma)} - E[v_s(x, \gamma)]^{-1} - \bar{v}_1(\gamma)^{-1}$$

Substituting out  $\eta_3(\gamma) \bar{h}$  in the expression above for  $\eta_1(\gamma) \widehat{g}_1(x, \gamma)$ , and using the fact that  $\eta_0(\gamma) \equiv E[v_s(x, \gamma)]^{-1}$  now yields the first line of (30) upon rearrangement, which is the first order condition for the first state.

3. The definition of  $\widehat{\alpha}_2(\gamma) \equiv \{E[v_s(x, \gamma)]\}^{-1}$  directly implies the participation constraint is met with equality, and hence the complementary slackness condition for participation is satisfied. The complementary slackness conditions for truth telling and sincerity constraints are directly imposed by virtue of  $\gamma \in \widehat{\Gamma}$ . We now show the remaining two complementary slackness conditions are satisfied. In the second state, we again appeal to the fact that the definitions of  $\widehat{\alpha}_1(\gamma)$ ,  $\widehat{\alpha}_2(\gamma)$ ,  $\widehat{g}_2(x, \gamma)$  and  $\eta_2(\gamma)$  are identical to their counterparts in the pure moral hazard model, which implies from Item 2 in the moral hazard case that:

$$\eta_2(\gamma) \left[ \widehat{g}_2(x, \gamma) - \frac{\widehat{\alpha}_2(\gamma)}{\widehat{\alpha}_1(\gamma)} \right] = E_2[v_2(x, \gamma)]^{-1} - v_2(x, \gamma)^{-1}$$

Multiplying this equation by  $v_2(x, \gamma)$  and taking expectations conditional on the second

state yields:

$$\begin{aligned} E_2 \left\{ \eta_2(\gamma)_2 v_2(x, \gamma) \left[ \hat{g}_2(x, \gamma) - \frac{\hat{\alpha}_2(\gamma)}{\hat{\alpha}_1(\gamma)} \right] \right\} &= E_2 \left\{ \eta_2(\gamma)_2 v_2(x, \gamma) [E_2[v_2(x, \gamma)]^{-1} - v_2(x, \gamma)^{-1}] \right\} \\ &= 0 \end{aligned}$$

proving from (24) that the complementary slackness condition for incentive compatibility in the second state holds.

Multiplying the first line of (30), the first order condition for the first state, by  $v_1(x, \gamma)$ , using the identity  $\eta_0(\gamma) \equiv E[v_s(x, \gamma)]^{-1}$ , and taking the expectation conditional on the first state yields implies:

$$\begin{aligned} &\eta_1(\gamma) E_1 \left\{ v_1(x, \gamma) \left[ \hat{g}_1(x, \gamma) - \frac{\hat{\alpha}_2(\gamma)}{\hat{\alpha}_1(\gamma)} \right] \right\} \\ &= E_1[v_1(x, \gamma)] E[v_s(x, \gamma)]^{-1} - \eta_3(\gamma) E_1[v_1(x, \gamma)h(x)] \\ &\quad - \eta_4(\gamma) \frac{\hat{\alpha}_2(\gamma)}{\hat{\alpha}_1(\gamma)} E_1[v_1(x, \gamma)\hat{g}_2(x, \gamma)h(x)] - 1 \end{aligned}$$

Successively substituting the definitions of  $\eta_3(\gamma)$  and  $\eta_4(\gamma)$  into the right side of this equation proves that both sides of the equation are zero. Comparing the left side of the equation with (24), it now follows that the complementary slackness condition for incentive compatibility in the first state also holds.

■

**Proof of Lemma 4.1.** For notational convenience, and without loss of generality, we suppress the dependence of compensation  $w_{nt}$  on  $(s_{nt}, b_t, b_{t+1})$ . Let  $\tilde{x}$  denote the net excess returns,  $x$  gross excess returns,  $w(x)$  the compensation schedule as a mapping from gross excess returns, and let  $V$  denote the value of the firm at the beginning of the period. By our definition of net and gross excess returns:

$$\tilde{x} = x - w(x)/V \tag{78}$$

Suppose there exists for some  $(\tilde{x}_0, V_0)$  two distinct values of net excess returns, denoted  $x_1 \in R$  and  $x_2 \in R$ , satisfying Equation (78). Then:

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

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

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

But this possibility is ruled out as a possibility in the premise of the Lemma. Therefore a unique solution to the relation defined by Equation (78) exists for each pair  $(\tilde{x}, V)$ , and we can denote the solution mapping by  $x \equiv X(\tilde{x}, V)$ . Substituting  $X(\tilde{x}, V)$  for  $x$  in  $w(x)$  we define  $\Lambda(\tilde{x}, V) \equiv w[X(\tilde{x}, V)]$ . The lemma now follows because the measurement error on compensation is assumed to independent of  $(\tilde{x}, V)$ , so  $E[\tilde{w}|\tilde{x}, V] = \Lambda(\tilde{x}, V)$ . ■

## B A Dynamic Hybrid Model

This appendix develops the notation for a dynamic version of the hybrid model of moral hazard, writes down the feasibility constraints for the optimization problem, and then shows that the optimal contract mimics the optimal contract for a static model under the parameter transformation given in the text.

### B.1 Assumptions and notation

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

If the manager accepts employment with the firm, so  $l_{t0} = 0$ , the prospects of the firm are now fully revealed to the manager but partially hidden to the shareholders. There are two states, and the probability the first state occurs is identically and independently distributed with probability  $\varphi_1 \in (0, 1)$ . For convenience we denote the probability of the second state occurring by  $\varphi_2 \equiv 1 - \varphi_1$ . We assume that managers privately observe the true state  $s_t \in \{1, 2\}$  in period  $t$ , information that affects the distribution of the firm's abnormal returns next period, and reports the state  $r_t \in \{1, 2\}$  to the board. If the manager discloses the second state, meaning  $r_t = 2$ , then the board can independently confirm or refute it; thus if  $s_t = 1$  he reports  $r_t = 1$ . If  $s_t = 2$  the manager then truthfully declares or lies about the firm's prospects by announcing  $r_t \in \{1, 2\}$ , effectively selecting one of two schedules,  $w_{1t}(x)$  or  $w_{2t}(x)$  in that case.

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

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

for both states  $s_t \in \{1, 2\}$ :

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

the inequality reflecting the preference of shareholders for diligent work over shirking. Since  $f_s(x) g_s(x)$  is a density,  $g_s(x)$  is positive and integrating  $f_s(x) g_s(x)$  with respect to  $x$  with respect to  $f_s(x)$  demonstrates  $E_s [g_s(x)] = 1$ . As in the text we assume:

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

for each  $s \in \{1, 2\}$ . We make similar assumptions about the weighted likelihood ratio of the second state occurring relative to the first given any observed value of excess returns  $x \in R$ , by assuming:

$$\lim_{x \rightarrow \infty} [\varphi_2 f_2(x) / \varphi_1 f_1(x)] \equiv \lim_{x \rightarrow \infty} [h(x)] = \sup_{x \in R} [h(x)] \equiv \bar{h} < \infty \quad (79)$$

The manager's wealth is endogenously determined by his consumption and compensation. We assume there are a complete set of markets for all publicly disclosed events, effectively attributes all deviations from the law of one price to the particular market imperfections under consideration. Let  $b_t$  denote the price of a bond that pays of a unit of consumption each period from period  $t$  onwards, relative to the price of a unit of consumption in period  $t$ ; to simplify the exposition we assume  $b_{t+1}$  is known at period  $t$ . Preferences over consumption and work are parameterized by a utility function exhibiting absolute risk aversion that is additively separable over periods and multiplicatively separable with respect to consumption and work activity within periods. In the model we estimate, lifetime utility can be expressed as:

$$- \sum_{t=0}^{\infty} \sum_{j=0}^J \beta^t \tilde{\alpha}_j l_{tj} \exp(-\tilde{\gamma} c_t) \quad (80)$$

where  $\beta$  is the constant subjective discount factor,  $\tilde{\gamma}$  is the constant absolute level of risk aversion, and  $\tilde{\alpha}_j$  is a utility parameter that measures the distaste from working at level  $j \in \{0, 1, 2\}$ . As in the text we assume  $\tilde{\alpha}_2 > \tilde{\alpha}_1$  and normalize  $\tilde{\alpha}_0 = 1$ .

## B.2 Feasibility constraints

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

**Lemma B.1** *If the manager, offered a contract of  $w_{rt}(x)$  for announcing  $r$ , retires in period  $t$  or period  $t+1$  by setting  $(1 - l_{t0})(1 - l_{t+1,0}) = 0$ , upon observing the state  $s$  and reporting*

$r_t(s)$  he optimally chooses  $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$  to minimize:

$$\sum_{s=1}^2 \varphi_s \left\{ \left( \frac{l_{st1}}{\tilde{\alpha}_1} + \frac{l_{st2}}{\tilde{\alpha}_2} \right)^{1/(b_t-1)} + E_s \left[ \exp \left( -\frac{\tilde{\gamma} w_{r_t(s),t}(x)}{b_{t+1}} \right) [g_s(x) l_{t1} + l_{t2}] \right] \right\} \quad (81)$$

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

$$v_{st}(x) \equiv \exp \left( -\frac{\tilde{\gamma} w_{st}(x)}{b_{t+1}} \right) \quad (82)$$

To induce an honest, diligent manager to participate, his expected utility from employment must exceed the utility he would obtain from retirement. Setting  $(l_{t2}, r_t) = (1, s_t)$  in (81) and substituting in  $v_{st}(x)$ , the participation constraint is thus:

$$\left[ \sum_{s=1}^2 \int_x^\infty \varphi_s v_{st}(x) f_s(x) dx \right] \equiv E[v_{st}(x)] \leq \tilde{\alpha}_2^{-1/(b_t-1)} \quad (83)$$

Given his decision to stay with the firm one more period, and truthfully reveal the state, the incentive compatibility constraint induces the manager to prefer working diligently to shirking. Substituting the definition of  $v_{st}(x)$  into (81) 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:

$$0 \leq \int_x^\infty \left( g_s(x) - (\tilde{\alpha}_2/\tilde{\alpha}_1)^{1/(b_t-1)} \right) v_{st}(x) f_s(x) dx \equiv E_s \left[ \left( g_s(x) - (\tilde{\alpha}_2/\tilde{\alpha}_1)^{1/(b_t-1)} \right) v_{st}(x) \right] \quad (84)$$

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

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

$$0 \leq \int [v_{1t}(x) - v_{2t}(x)] f_2(x) dx \equiv E_2 [v_{1t}(x) - v_{2t}(x)] \quad (85)$$

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

$$0 \leq \int \left[ (\tilde{\alpha}_1/\tilde{\alpha}_2)^{1/(b_t-1)} v_{1t}(x) g_2(x) - v_{2t}(x) \right] f_2(x) dx \equiv E_2 \left[ (\tilde{\alpha}_1/\tilde{\alpha}_2)^{1/(b_t-1)} v_{1t} g_2(x) - v_{2t}(x) \right] \quad (86)$$

where  $(\tilde{\alpha}_1/\tilde{\alpha}_2)^{1/(b_t-1)} v_{1t}(x)$  is proportional to the utility obtained from shirking and announcing the first state, and  $f_2(x) g_2(x)$  is the probability density function associated with shirking when the second state occurs.

### B.3 Optimal contracting

We first prove the short term optimal contract for the dynamic model has a static analogue of the form we describe in the text, and then show that the long term contract decomposes to a sequence of short term contracts. As in the static model deriving  $w_{st}(x)$  to minimize expected compensation of inducing diligent work in both states subject to the five constraints is equivalent to choosing  $v_{st}(x)$  to maximize:

$$\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s \log [v_{st}(x)] f_s(x) dx \equiv E [\log v_{st}(x)] \quad (87)$$

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

$$\begin{aligned} & \sum_{s=1}^2 \varphi_\rho \int_{\underline{x}}^{\infty} \left\{ \log [v_{st}(x)] + \eta_{0t} \left[ 1 - \tilde{\alpha}_2^{1/(b_t-1)} v_{st}(x) \right] \right\} f_s(x) dx \\ & + \sum_{s=1}^2 \varphi_s \eta_{st} \int_{\underline{x}}^{\infty} v_{st}(x) \left[ \tilde{\alpha}_1^{1/(b_t-1)} g_s(x) - \tilde{\alpha}_2^{1/(b_t-1)} \right] f_s(x) dx \\ & + \varphi_2 \eta_{3t} \int_{\underline{x}}^{\infty} [v_{1t}(x) - v_{2t}(x)] f_2(x) dx \\ & + \varphi_2 \eta_{4t} \int_{\underline{x}}^{\infty} \left[ \tilde{\alpha}_1^{1/(b_t-1)} v_{1t}(x) g_2(x) - \tilde{\alpha}_2^{1/(b_t-1)} v_{2t}(x) \right] f_2(x) dx \end{aligned} \quad (88)$$

with respect to  $v_{st}(x)$ , where  $\eta_{0t}$  through  $\eta_{4t}$  are the shadow values assigned to the linear constraints. Setting:

$$\alpha_1 = \tilde{\alpha}_1^{1/(b_t-1)} \quad \alpha_2 = \tilde{\alpha}_2^{1/(b_t-1)} \quad \gamma = \tilde{\gamma}/b_{t+1} \quad (89)$$

establishes by inspection that the solution to the static model solves the transformed problem as claimed in the text.

In this framework there are no gains from a long term arrangement between shareholders and the manager. Lemma B.2 verifies the assumptions of Fudenberg, Holmstrom and Milgrom (1990) are met, thus establishing that the long term optimal contract decentralizes to a sequence of short term contracts solved by the problem above.<sup>26</sup>

**Lemma B.2** *Denote by  $\bar{\tau}$  the manager's date of retirement. The optimal long term contract can be implemented by a  $\bar{\tau}$  period replication of the optimal short term contract.*

<sup>26</sup>Malcomson and Spinnewyn (1988), Fudenberg, Homstrom and Milgom (1990) and Rey and Salanie (1990) have independently established conditions under which long term optimal contracts can be implemented via a sequence of one period contracts in dynamic models of generalized moral hazard, and the proof of Lemma B.2 in the Appendix draws extensively upon their results.

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

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

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

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

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

$$-b_t \tilde{\alpha}_2^{1/b_t} \left\{ E \left[ \exp \left( -\frac{\tilde{\gamma} w_{r_t(s)t}(x)}{b_{t+1}} \right) \right] \right\}^{1-1/b_t} \exp \left( -\frac{q_t + \tilde{\gamma} e_{nt}}{b_{t+1}} \right)$$

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

$$-b_t \tilde{\alpha}_1^{1/b_t} \left\{ E_t \left[ \exp \left( -\frac{\tilde{\gamma} w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x)] \right] \right\}^{1-1/b_t} \exp \left( -\frac{q_t + \tilde{\gamma} e_{nt}}{b_{t+1}} \right)$$

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

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

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

$$\begin{aligned} & l_{t0} + (\tilde{\alpha}_1 l_{st1} + \tilde{\alpha}_2 l_{st2})^{1/b_t} \left\{ E \left[ \exp \left( -\frac{\tilde{\gamma} w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x) l_{st1} + l_{st2}] \right] \right\}^{1-1/b_t} \\ = & l_{t0} + \left\{ (\tilde{\alpha}_1 l_{st1} + \tilde{\alpha}_2 l_{st2})^{1/(b_t-1)} E_t \left[ \exp \left( -\frac{\tilde{\gamma} w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x) l_{st1} + l_{st2}] \right] \right\}^{(b_t-1)/b_t} \end{aligned}$$

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

$$l_{t0} + (\tilde{\alpha}_1 l_{st1} + \tilde{\alpha}_2 l_{st2})^{1/(b_t-1)} E_t \left[ \exp \left( -\frac{\tilde{\gamma} w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_s(x) l_{st1} + l_{st2}] \right]$$

Multiplying through by the factor  $(\tilde{\alpha}_1 l_{st1} + \tilde{\alpha}_2 l_{st2})^{1/(b_t-1)}$  and summing over the two states  $s \in \{1, 2\}$  yields the minimand in Lemma B.1 ■

**Proof of Lemma B.2.** In our model the proof of Proposition 5 in Margiotta and Miller (2000) can be simply adapted to show that Theorem 3 of Fudenberg, Holmstrom and Milgrom (1990) applies, thus demonstrating that the long term optimal contract can be sequentially

implemented. An induction completes the proof, by establishing that the sequential contract implementing the optimal long term contract for a manager who will retire in  $\bar{\tau}$  periods replicates the one period optimal contract. In the optimal short term contract, the participation constraint is satisfied with strict equality, which implies that at the beginning of period  $\bar{\tau} - 1$  the expected lifetime utility of the manager is determined by setting  $t = \bar{\tau} - 1$  in the equation:

$$-b_t \exp\left(-\frac{a_t + \tilde{\gamma}e_t}{b_t}\right) \quad (90)$$

Suppose that at the beginning of all periods  $t \in \{\tau + 1, \tau + 2, \dots, \bar{\tau} - 1\}$ , the expected lifetime utility of the manager is given by Equation (90). We first show the expected lifetime utility of the manager at  $\tau$  is also given by Equation (90). From Lemma 2.1 the problem shareholders solve at  $\tau$  is identical to the short term optimization problem solved in the text. In the solution to each cost minimization subproblem for the four  $(L_{1t}, L_{2t})$  choices, the manager's participation constraint is met with equality. Consequently the manager achieves an expected lifetime utility of the manager at  $\tau$  is given by Equation (90) as claimed. Therefore the problem of participating at time  $\tau$  and possibly continuing with the firm for more than one period reduces to the problem of participating at time  $\tau$  one period at most, solved in Lemma B.1. The induction step now follows. ■

## C Implementing the Restrictions of the Pure Moral Hazard and Hybrid Models

This appendix extends the discussion of Section 5.1 on estimating and testing the unrestricted pure moral hazard model to other models we analyzed in our empirical investigations. First we show the set of admissible  $\tilde{\gamma}$  shrinks when we impose the restrictions that tastes for working or shirking do not change with the state, and only vary with the bond price following Equations. Then we characterize the set of restrictions on  $\tilde{\gamma}$  implied by the fully restricted hybrid moral hazard model as defined in the text.

### C.1 Restrictions on the pure moral hazard model

In the most unrestricted pure moral hazard model it follows from Section 5.1 that:

$$\tilde{\Gamma} = \{\tilde{\gamma} > 0 : Q(\tilde{\gamma}) = 0\}$$

where

$$Q(\tilde{\gamma}) \equiv \sum_{t=1}^T \sum_{s=1}^2 \min\{0, Q_{st}(\tilde{\gamma})\}^2 \quad (91)$$

With reference to and in the text, we now define the taste parameters for the dynamic version of the pure moral hazard model as:

$$\begin{aligned} \alpha_{1st}(\tilde{\gamma}) &\equiv \alpha_1 (\tilde{\gamma}/b_{t+1})^{b_t-1} \\ \alpha_{2st}(\tilde{\gamma}) &\equiv \alpha_2 (\tilde{\gamma}/b_{t+1})^{b_t-1} \end{aligned}$$

We investigated how the confidence region for  $\tilde{\gamma}$  shrinks when we impose the restrictions that  $\alpha_{1st}(\tilde{\gamma})$  and  $\alpha_{2st}(\tilde{\gamma})$  do not change with the state  $s \in \{1, 2\}$  or with time  $t \in \{1, \dots, T\}$ .

To impose the restriction that  $\alpha_{1st}(\tilde{\gamma})$  does not vary by state, we define the real valued functions  $\Upsilon_{1t}(\tilde{\gamma})$  as:

$$\Upsilon_{1t}(\tilde{\gamma}) \equiv \left\{ \frac{1 - E_{1t} \left[ \exp \left( \frac{\tilde{\gamma}(w_{1t} - \bar{w})}{b_{t+1}} \right) \right]}{E_{1t} \left[ \exp \left( \frac{-\tilde{\gamma}w_{1t}}{b_{t+1}} \right) \right] - \exp \left( \frac{-\tilde{\gamma}\bar{w}}{b_{t+1}} \right)} \right\}^{b_t - 1} - \left\{ \frac{1 - E_{2t} \left[ \exp \left( \frac{\tilde{\gamma}(w_{2t} - \bar{w})}{b_t} \right) \right]}{E_{2t} \left[ \exp \left( \frac{-\tilde{\gamma}w_{2t}}{b_t} \right) \right] - \exp \left( \frac{-\tilde{\gamma}\bar{w}}{b_t} \right)} \right\}^{b_1 - 1}$$

and note that  $\Upsilon_{1t}(\tilde{\gamma}) = 0$  if and only if  $\alpha_{11t}(\tilde{\gamma}) = \alpha_{12t}(\tilde{\gamma})$ . Similarly  $\alpha_{21t}(\tilde{\gamma}) = \alpha_{22t}(\tilde{\gamma})$  if and only if  $\Upsilon_{2t}(\tilde{\gamma}) = 0$  where:

$$\Upsilon_{2t}(\tilde{\gamma}) \equiv \left\{ E_2 \left[ \exp \left( \frac{-\tilde{\gamma}w_{2t}}{b_{t+1}} \right) \right] \right\}^{1 - b_t} - \left\{ E_1 \left[ \exp \left( \frac{-\tilde{\gamma}w_{1t}}{b_t} \right) \right] \right\}^{1 - b_t}$$

Thus to find a confidence region for the risk parameter under the null hypothesis that tastes for shirking or working,  $\alpha_{jst}(\tilde{\gamma})$  for  $j \in \{1, 2\}$ , do not vary by state, we augment (91) and find those values of  $\tilde{\gamma}$  that achieve close to the lower bound of zero for a sample analogue of:

$$\sum_{t=1}^T \sum_{s=1}^2 [\min \{0, Q_{st}(\tilde{\gamma})\}^2 + \Upsilon_{jt}(\tilde{\gamma})^2] \quad (92)$$

The results from separately imposing these two sets of restrictions for are reported in Table 6 of the text.

Essentially the same procedure can be used to constrain  $\tilde{\alpha}_{1st}$  or  $\tilde{\alpha}_{2st}$  to remain constant over time. Defining:

$$\Upsilon_{1st}(\tilde{\gamma}) \equiv \left\{ \frac{1 - E_{st} \left[ \exp \left( \frac{\tilde{\gamma}(w_{st} - \bar{w})}{b_{t+1}} \right) \right]}{E_{st} \left[ \exp \left( \frac{-\tilde{\gamma}w_{st}}{b_{t+1}} \right) \right] - \exp \left( \frac{-\tilde{\gamma}\bar{w}}{b_{t+1}} \right)} \right\}^{b_t - 1} - \left\{ \frac{1 - E_{s1} \left[ \exp \left( \frac{\tilde{\gamma}(w_{s1} - \bar{w})}{b_2} \right) \right]}{E_{s1} \left[ \exp \left( \frac{-\tilde{\gamma}w_{s1}}{b_2} \right) \right] - \exp \left( \frac{-\tilde{\gamma}\bar{w}}{b_2} \right)} \right\}^{b_1 - 1}$$

it immediately follows from that  $\tilde{\alpha}_{1s1} = \tilde{\alpha}_{1st}$  when  $\Upsilon_{1st}(\tilde{\gamma}) = 0$ . Similarly  $\tilde{\alpha}_{2s1} = \tilde{\alpha}_{2st}$  when  $\Upsilon_{2st}(\tilde{\gamma}) = 0$ , where  $\Upsilon_{2st}(\tilde{\gamma})$  is defined as:

$$\Upsilon_{2st}(\tilde{\gamma}) \equiv \left\{ E_s \left[ \exp \left( \frac{-\tilde{\gamma}w_{st}}{b_{t+1}} \right) \right] \right\}^{1 - b_t} - \left\{ E_s \left[ \exp \left( \frac{-\tilde{\gamma}w_{s1}}{b_2} \right) \right] \right\}^{1 - b_1}$$

This restriction implies that  $\Upsilon_{2st}(\tilde{\gamma}) = 0$  for all  $t \in \{1, 2, \dots, T\}$  and  $s \in \{1, 2\}$ . Thus the confidence region for the risk parameter under the null hypothesis that  $\alpha_{jst}(\tilde{\gamma})$  does not vary over time, could be found by constructing a sample analogue of:

$$\sum_{t=1}^T \sum_{s=1}^2 [\min \{0, Q_{st}(\tilde{\gamma})\}^2 + \Upsilon_{jst}(\tilde{\gamma})^2] \quad (93)$$

and, using the methods we describe below, selecting those  $\tilde{\gamma}$  that bring the criterion function

close to zero.

Sample analogues are formed the same way as in the unrestricted model and the testing procedures followed the same steps. As reported in the text, the intersection of the regions across the firm types is empty, implying that under pure moral hazard tastes vary either over time or across state. For the sake of completeness Table A1 displays the regions that

## C.2 Restrictions on the hybrid moral hazard model

The restrictions in the hybrid model are imposed in a similar way. Here we maintain throughout our analysis of the hybrid model the null hypothesis that the that the taste parameters for working and shirking, both mappings of  $\tilde{\gamma}$ , do not vary by state or time. These restrictions are maintained because the intersection of the estimated confidence intervals for  $\tilde{\gamma}$  for the 24 sectors under the null hypothesis is not empty.

To develop the notation for the econometric framework that accommodates a panel where bonds prices over time, as opposed to a cross section or a steady state economy with constant interest rates, we extend our notation as follows. Appealing to Equations (47) and (48), define taste parameters that are independent of the states:

$$\begin{aligned}\hat{\alpha}_{1t}(\tilde{\gamma}) &\equiv \hat{\alpha}_1(\tilde{\gamma}/b_{t+1})^{b_t-1} \\ \hat{\alpha}_{2t}(\tilde{\gamma}) &\equiv \hat{\alpha}_2(\tilde{\gamma}/b_{t+1})^{b_t-1}\end{aligned}$$

Similarly the likelihood ratio for the second state is defined as

$$\hat{g}_{2t}(x, \tilde{\gamma}) \equiv \hat{g}_2(x, \tilde{\gamma}/b_{t+1})$$

We then define the Lagrange multipliers  $\eta_{1t}(\tilde{\gamma})$  through  $\eta_{4t}(\tilde{\gamma})$  by substituting  $\tilde{\gamma}/b_{t+1}$  for  $\gamma$ ,  $\hat{\alpha}_{1t}(\tilde{\gamma})$  for  $\hat{\alpha}_1(\gamma)$ ,  $\hat{\alpha}_{2t}(\tilde{\gamma})$  for  $\hat{\alpha}_2(\gamma)$  and  $\hat{g}_{2t}(x, \tilde{\gamma})$  for  $\hat{g}_2(x, \gamma)$  into Equations, and hence define:

$$\hat{g}_{1t}(x, \gamma) \equiv \hat{g}_{2t}(x, \gamma/b_{t+1})$$

We are now in a position to define  $\hat{\Gamma}_t$  by substituting  $\eta_{1t}(\tilde{\gamma})$  through  $\eta_{4t}(\tilde{\gamma})$  for  $\eta_1(\tilde{\gamma})$  through  $\eta_4(\tilde{\gamma})$ ,  $\Lambda_{it}(\tilde{\gamma})$  for  $\Lambda_i(\tilde{\gamma})$ , and  $\Psi_{kt}(\tilde{\gamma})$  for  $\Psi_k(\tilde{\gamma})$  in the definition of  $\hat{\Gamma}$  and replacing  $\gamma$  with  $\tilde{\gamma}$ . To impose the restriction that none of the parameters vary over time we take the intersection:

$$\begin{aligned}\hat{\Gamma}(T) &\equiv \bigcap_{t=1}^T \hat{\Gamma}_t \\ &= \left\{ \tilde{\gamma} > 0 : \hat{Q}(\tilde{\gamma}) = 0 \right\}\end{aligned}$$

where:

$$\begin{aligned}\hat{Q}(\tilde{\gamma}) &\equiv \sum_{t=1}^T \sum_{j=5}^9 \min[0, \Psi_{jt}(\tilde{\gamma})]^2 + \sum_{t=1}^T \sum_{j=6}^7 [\Psi_{5t}(\tilde{\gamma}) \Psi_{jt}(\tilde{\gamma})]^2 \\ &\quad + \sum_{t=1}^T \Psi_{4t}(\tilde{\gamma})^2 + \sum_{t=1}^T [\Psi_{6t}(\tilde{\gamma}) \Psi_{8t}(\tilde{\gamma})]^2 + \sum_{t=1}^T \sum_{k=3}^5 \min[0, \Lambda_{kt}(\tilde{\gamma})]^2\end{aligned}\quad (94)$$

In the hybrid model, the components of  $\widehat{Q}(\tilde{\gamma})$  are formed from the probability density functions characterizing abnormal returns, conditional on the firm's characteristics and the manager's report, that is  $f_s(x)$ , and the nonlinear regression function of compensation on abnormal returns the same set of variables, denoted by  $w_{st}(x)$ . In the previous sections we described our estimates of the compensation scheme,  $w_{st}^{(N)}(x)$ , the probability densities,  $f_s^{(N)}(x)$ , and the probabilities,  $\varphi_s^{(N)}$ . From these estimated functions, we directly form the estimated weighted ratio  $h^{(N)}(x)$ . Our structural analysis inputs vectors of the form  $(w_{nt}^{(N)}, x_{nt}^{(N)}, r_{nt})$  and the subsampling methods we use to obtain test statistics compute the vectors in each subsample. Denote by  $f_s(x)$  the conditional density of abnormal returns  $x$  given the true state  $s$  and the firm's characteristics  $z$ . Under the null hypothesis that  $s_t = r_t(s_t)$  consistent estimates of  $f_s(x)$  can be obtained from nonparametric density estimator of  $x_{nt}^{(N)}$  on  $r_{nt}$ . Similarly estimates of the compensation schedule for period  $t$  are given by the nonparametric regression estimator of  $w_{nt}^{(N)}$  on  $r_{nt}$  and  $x_{nt}^{(N)}$ .

To impose the restrictions embodied in the dynamic version of the hybrid model we form nonparametric estimators  $\Psi_{jt}^{(N)}(\tilde{\gamma})$  and  $\Lambda_{kt}^{(N)}(\tilde{\gamma})$  for  $\Psi_{jt}(\tilde{\gamma})$  and  $\Lambda_{kt}(\tilde{\gamma})$  from estimates of their components. In the previous section we described our estimates of the compensation scheme,  $w_{st}^{(N)}(x)$ , the probability densities,  $f_s^{(N)}(x)$ , and the probabilities,  $\varphi_s^{(N)}$ . From these estimated functions, we directly form the estimated weighted ratio  $h^{(N)}(x)$ ,  $Q_{st}^{(N)}(\tilde{\gamma})$  as well as  $\Psi_{jt}^{(N)}(\tilde{\gamma})$  for  $j \in \{1, 2, 5\}$  using the definitions of  $\Psi_{jt}(\tilde{\gamma})$  given in the previous section. In addition to  $\bar{w}_{st}^{(N)}$  we also require an estimate of  $\bar{h}(z)$  and  $\bar{h}^{(N)}$  which we obtained using the estimators described in the text. Defining the sample analogue to  $\widehat{Q}(\tilde{\gamma})$  as:

$$\begin{aligned} \widehat{Q}^{(N)}(\tilde{\gamma}) \equiv & \sum_{t=1}^T \sum_{j=5}^9 \min \left[ 0, \Psi_{jt}^{(N)}(\tilde{\gamma}) \right]^2 + \sum_{t=1}^T \sum_{j=6}^7 [\Psi_{5t}^{(N)}(\tilde{\gamma}) \Psi_{jt}^{(N)}(\tilde{\gamma})]^2 \\ & + \sum_{t=1}^T \Psi_{4t}^{(N)}(\tilde{\gamma})^2 + \sum_{t=1}^T [\Psi_{6t}^{(N)}(\tilde{\gamma}) \Psi_{8t}^{(N)}(\tilde{\gamma})]^2 + \sum_{t=1}^T \sum_{k=3}^5 \min \left[ 0, \Lambda_{kt}^{(N)}(\tilde{\gamma}) \right]^2 \quad (95) \end{aligned}$$

we constructed a confidence region for the hybrid model in a similar way to the pure moral hazard model. That is substituting  $\widehat{Q}^{(N)}(\tilde{\gamma})$  into Equation ( ) in the text, we derived the rate of convergence and numerically computed the critical value using the same procedures laid out in the text.

Computing  $\Lambda_{2t}(\tilde{\gamma})$  and  $\Lambda_{3t}(\tilde{\gamma})$  requires us to solve  $w_s^{(1,0)}(x, \tilde{\gamma}/b_{t+1})$  and  $w_s^{(0,1)}(x, \tilde{\gamma}/b_{t+1})$  for each candidate value of  $\tilde{\gamma}$ , a nonlinear problem that includes two Lagrange multipliers. If the states  $s \in \{1, 2\}$  and the effort level  $(l_1, l_2)$  were observed by shareholders, then they would optimally offer  $b_{t+1} \log[\hat{\alpha}_{1t}(\tilde{\gamma})]/(b_t - 1)\tilde{\gamma}$  for shirking and  $b_{t+1} \log[\hat{\alpha}_{2t}(\tilde{\gamma})]/(b_t - 1)\tilde{\gamma}$  for diligence. The profits from this hypothetical arrangement are therefore:

$$\begin{aligned} \Lambda'_{2t}(\tilde{\gamma}) = & \varphi_1 \{ E[x] - b_{t+1} \log[\hat{\alpha}_{2t}(\tilde{\gamma})]/(b_t - 1)\tilde{\gamma} \} \\ & + \varphi_2 E_2 \{ x \hat{g}_{2t}(x, \tilde{\gamma}) - b_{t+1} \log[\hat{\alpha}_{1t}(\tilde{\gamma})]/(b_t - 1)\tilde{\gamma} \} \end{aligned}$$

from shirking in the second state and working diligently in the first, and:

$$\begin{aligned}\Lambda'_{3t}(\tilde{\gamma}) &= \varphi_2 \{E[x] - b_{t+1} \log[\hat{\alpha}_{2t}(\tilde{\gamma})] / (b_t - 1) \tilde{\gamma}\} \\ &\quad + \varphi_1 E_1[x \hat{g}_{1t}(x, \tilde{\gamma}) - b_{t+1} \log[\hat{\alpha}_{1t}(\tilde{\gamma})] / (b_t - 1) \tilde{\gamma}]\end{aligned}$$

from shirking in the first state and working diligently in the second. Since neither cost minimization problem imposes the truth telling, sincerity and one incentive compatibility constraint, but has the same objective function, it now follows that  $\Lambda'_{2t}(\tilde{\gamma}) \geq \Lambda_{2t}(\tilde{\gamma})$  and  $\Lambda'_{3t}(\tilde{\gamma}) \geq \Lambda_{3t}(\tilde{\gamma})$ . Let  $\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t})$  denote the set of  $\tilde{\gamma}$  formed from excluding  $\Lambda_{2t}(\tilde{\gamma})$  and  $\Lambda_{3t}(\tilde{\gamma})$  for all  $t$ . By construction:

$$\hat{\Gamma}(\tilde{\gamma}, T) \subseteq \hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t})$$

Now let  $\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t}, \Lambda'_{2t}, \Lambda'_{3t})$  denote the set of  $\tilde{\gamma}$  formed from intersecting  $\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t})$  with  $\Lambda'_{2t}(\tilde{\gamma})$  and  $\Lambda'_{3t}(\tilde{\gamma})$  for all  $t \in \{1, 2, \dots, T\}$ . Since  $\Lambda'_{2t}(\tilde{\gamma}) \geq \Lambda_{2t}(\tilde{\gamma})$  and  $\Lambda'_{3t}(\tilde{\gamma}) \geq \Lambda_{3t}(\tilde{\gamma})$  it immediately follows that:

$$\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t}, \Lambda'_{2t}, \Lambda'_{3t}) \subseteq \hat{\Gamma}(\tilde{\gamma}, T)$$

Thus:

$$\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t}, \Lambda'_{2t}, \Lambda'_{3t}) \subseteq \hat{\Gamma}(\tilde{\gamma}, T) \subseteq \hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t}) \quad (96)$$

In our empirical application we found the confidence region for  $\tilde{\gamma}$  obtained from imposing  $\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t})$  coincided with the confidence region obtained from imposing  $\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t}, \Lambda'_{2t}, \Lambda'_{3t})$ . In other words imposing the restrictions  $\Lambda'_{2t}(\tilde{\gamma})$  and  $\Lambda'_{3t}(\tilde{\gamma})$  for all  $t \in \{1, 2, \dots, T\}$  did not shrink  $\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t})$ , implying from (96) that:

$$\hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t}, \Lambda'_{2t}, \Lambda'_{3t}) = \hat{\Gamma}(\tilde{\gamma}, T) = \hat{\Gamma}(\tilde{\gamma}, T / \Lambda_{2t}, \Lambda_{3t})$$

In this way we computed the confidence region for  $\hat{\Gamma}(\tilde{\gamma}, T)$  without solving for  $w_s^{(1,0)}(x, \tilde{\gamma} / b_{t+1})$  and  $w_s^{(0,1)}(x, \tilde{\gamma} / b_{t+1})$ .

## References

- [1] Andersen, S., G. W. Harrison, M.I. Lau, and E.E. Rutstrom (2008): "Eliciting Risk and Time Preferences", *Econometrica*, **76(3)**, 583-618.
- [2] Andrews, D., S. Berry and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location", Department of Economics, Yale University.
- [3] Antle, R. and Smith, A. (1985): "Measuring Executive Compensation: Methods and an Application", *Journal of Accounting Research*, **23**, 296-325.
- [4] Antle, R. and Smith, A. (1986): "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives", *Journal of Accounting Research*, **24**, 1-39.

- [5] Bajari, P., L. Bankard and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition", *Econometrica*, **75**, 1331-1371.
- [6] Baker, G., M. Gibbs and B. Holmstrom (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data", *Quarterly Journal of Economics*, **109**, 881-919.
- [7] Baker, G., M. Gibbs and B. Holmstrom (1994b): "The Wage Policy of a Firm", *Quarterly Journal of Economics*, **109**, 921-955.
- [8] Berle, A. A., and G. C. Means (1932): *The Modern Corporation and Private Property*. Harcourt, Brace, & World Inc., New York.
- [9] Bebchuk, L. and J. Fried (2003): "Executive Compensation as an Agency Problem", *Journal of Economic Perspectives*, **17**, 71-92.
- [10] Bertrand, M. and S. Mullainathan (2001): "Are CEOs Rewarded for Luck? The Ones without Principals Are", *Quarterly Journal of Economics*, **116**, 901-932.
- [11] Bolton, P. and M. Dewatripont (2005): *Contract Theory*, MIT Press, Cambridge.
- [12] Brunk, H. D. (1958): "On the Estimation of Parameters Restricted by Inequalities", *Annals of Mathematical Statistics*, **29**, 437-454.
- [13] Chernozhukov, V., H. Hong and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243-1284.
- [14] Chetty, R. (2006): "A New Method of Estimating Risk Aversion", *American Economic Review*, **96**(5), 1821-1834.
- [15] Chiappori, P.A. and B. Salanie (2000): "Testing Contract Theory: A Survey of Some Recent Work", in *Advances in Economics and Econometrics*, **1**, M. Dewatripont, L.Hansen and S. Turnovsky eds. , Cambridge University Press.
- [16] Cohen, A. and L. Einav (2007): "Estimating Risk Preferences from Deductible Choice", *American Economic Review*, **97**(3) , 745-788
- [17] Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets", *Econometrica*, **77**, 1791–828.
- [18] Dohmen, T., A. Falk, D. Huffman, U. Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?", *American Economic Review*, **100** (3), 1238-1260.
- [19] Dubois P. and T. Vukina (2009): "Optimal Incentives under Moral Hazard and Heterogeneous Agents: Evidence from Production Contracts Data", *International Journal of Industrial Organization*, **4**, 489-500.
- [20] Dufflo E., R. Hanna and S. Ryan (2011): "Incentives Work: Getting Teachers to Go to School", Forthcoming *American Economic Review*..

- [21] Dye, R. A. and S. S. Sridhar (2005): "Moral Hazard Severity and Contract Design", *RAND Journal of Economics*, **36**, 78-92.
- [22] Ferrall C. and B. Shearer (1999): "Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records", *Review of Economic Studies*, **66**, 309-338.
- [23] Finnerty, J.E. (1976): "Insiders and Market Efficiency", *Journal of Business*, **31**, 1141-1148.
- [24] Fudenberg, D., B. Holmstrom and P. Milgrom (1990): "Short-Term Contracts and Long-Term Agency Relationships", *Journal of Economic Theory*, **50**, 1-31.
- [25] Gayle, G.-L. and R. A. Miller (2009a): "Insider Information and Performance Pay", *CEsifo Economic Studies*, **55(3-4)**, 515-541.
- [26] Gayle, G.-L. and R. A. Miller (2009b): "Has Moral Hazard become a More important Factor in Managerial Compensation?", *American Economic Review*, **99(5)**: 1740–1769.
- [27] Gayle, G.-L., L. Golan and R. A. Miller (2011): "Promotion, Turnover and Compensation in the Executive Labor Market", Tepper School of Business, Carnegie Mellon University.
- [28] Gillan, S. L., Hartzell, J. C. and R. Parrino (2009): "Explicit versus Implicit Contracts: Evidence from CEO Employment Agreements", *The Journal of Finance*, **64**, 1629–1655.
- [29] Green, J. R. and J.-J. Laffont (1986): "Partially Verifiable Information and Mechanism Design", *Review of Economic Studies*, **53**, 447-456.
- [30] Grossman, S. and O. Hart (1983): "An Analysis of the Principal-Agent Problem", *Econometrica*, **51**, 7-46.
- [31] Hall, B. J. and J. B. Liebman, (1998): "Are CEOs Really Paid Like Bureaucrats?", *Quarterly Journal of Economics*, **113**, 653-680.
- [32] Harrison, G.W., E. Johnson, M.M. McInnes, E.E. Rutstrom (2005): "Risk Aversion and Incentive Effects: Comment", *American Economic Review*, **95(3)**, 897-901.
- [33] Harrison, G.W., J.A. List, C. Towe (2007): "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion", *Econometrica*, **75(2)**, 433-458.
- [34] Haubrich, J. (1994): "Risk Aversion, Performance Pay, and the Principal Agent Problem," *Journal of Political Economy*, **102**, 258-276.
- [35] Hayes, R. M. and S. Schaefer (2000): "Implicit Contracts and the Explanatory Power of Top Executive Compensation for Future Performance," *RAND Journal of Economics*, **31**, 273-293.

- [36] Ho, K. (2009): "Insurer-Provider Networks in the Medical Care Market", *American Economic Review*, 99(1): 393-430.
- [37] Ho, K., J. Ho, and J. Mortimer (2011): "The Effects of Full-Line Forcing Contracts", Forthcoming *American Economic Review*.
- [38] Holt, C. and S. K. Laury (2002): "Risk Aversion and Incentive Effects", *American Economic Review*, **92(5)**, 1644-1655.
- [39] Holt, C. and S. K. Laury (2005): "Risk Aversion and Incentive Effects: New Data without Order Effects", *American Economic Review*, **95(3)**, 902-904.
- [40] Huang Y., I. Perrigne, and Q. Vuong (2007): "Nonlinear Pricing in Yellow Pages", Department of Economics, Pennsylvania State University.
- [41] Jaffe, J.F. (1974): "Special information and Insider Trading", *Journal of Business*, **47**, 410-428.
- [42] Laffont, J.-J. and D. Martimort (2002): *The Theory of Incentives*, Princeton University Press, Princeton.
- [43] Levine, A. (2008): "Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry", Department of Economics, Stanford University.
- [44] Lorie, J.H. and V. Niederhoffer (1968): "Predictive and Statistical Properties of Insider Trading", *Journal of Law and Economics*, **11**, 35-51.
- [45] Malcomson, J. M. and F. Spinnewyn (1988): "The Multiperiod Principal-Agent Problem", *Review of Economic Studies*, **55**, 391-407.
- [46] Margiotta, M. and R. A. Miller (2000): "Managerial Compensation and The Cost of Moral Hazard", *International Economic Review*, **41**, 669-719.
- [47] Mirrlees, J. A. (1975): "The Theory of Moral Hazard and Unobservable Behavior, Part 1", mimeo, Oxford.
- [48] Murphy, K. (1999): "Executive Compensation", in *Handbook of Labor Economics*, 3, O. Ashenfelter and D. Card, eds., Amsterdam: North-Holland, 2485-2563.
- [49] Myerson, R. (1982): "Optimal Coordination Mechanisms in Generalized Principal Agent Problem", *Journal of Mathematical Economics*, **10**, 67-81.
- [50] Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in *Handbook of Econometrics*, eds. R. F. Engle and D. L. McFadden. Amsterdam: North-Holland. 2111-2245.
- [51] Pakes, A., J. Porter, K. Ho and J. Ishii (2006): "Moment Inequalities and Their Application", CEMMAP Working Paper.

- [52] Parsons, V. L. (1978): "The Limiting Distribution of the Isotonic Estimator at a Point", Unpublished Manuscript.
- [53] Perrigne, I. and Q. Vuong (2011): "Nonparametric Identification of A Contract Model with Adverse Selection and Moral Hazard", Forthcoming, *Econometrica*.
- [54] Prendergast, C. (1999): "The Provision of Incentives in Firms", *Journal of Economic Literature*, **37**, 7-63.
- [55] Rey, P. and B. Salanie (1990): "Long-term, Short-term and Renegotiation: On the Value of Commitment in Contracting", *Econometrica*, **58**, 597-619.
- [56] Romano, J.P. and A.M. Shaikh (2006): "Inference for Identifiable Parameters in Partially Identified Econometric Models", working paper.
- [57] Seyhun, H. N. (1986): "Insiders' Profits, Cost of Trading, and Market Efficiency", *Journal of Financial Economics*, **16**, 189-212.
- [58] Seyhun, H. N (1992a): "The Effectiveness of Insider Trading Sanctions", *Journal of Law and Economics*, **35**, 149-182.
- [59] Seyhun, H. N (1992b): "Why Does Aggregate Insider Trading Predict Future Stock Returns?", *Quarterly Journal of Economics*, **107**, 1303-1331.
- [60] Shearer, B. (2004): "Piece Rates, Fixed Wages and Incentives: Evidence From a Field Experiment", *Review of Economic Studies*, **71**, 513-534.
- [61] Wright, F.T. (1981): "The Asymptotic Behavior of Monotone Regression Estimates", *Annals of Statistics*, **9**, 443-448.

TABLE 1: CROSS SECTION SUMMARY OF FIRM CHARACTERISTICS BY SECTOR

Variable	Primary	Consumer	Services
Observations	8,980	6,762	11,144
Assets (millions of 2000 \$US)	6,322 (27,773)	5,277 (22,124)	17,776 (67,133)
Market Value (millions of 2000 \$US)	6,480 (25,160)	7,811 (21,975)	11,664 (35,002)
Employees (thousands)	15.8 (40.8)	32.23 (78.75)	11.9 (26.59)
Debt/Equity Ratio	2.07 (40.9)	1.94 (26.21)	4.56 (50.63)
Accounting Return	1.15 (4.54)	1.13 (1.68)	1.28 (7.26)

TABLE 2: TIME SERIES SUMMARY OF FIRM CHARACTERISTICS AND COMPENSATION

Year	Bond	Assets	Employees	$\frac{\text{Debt}}{\text{Equity}}$	$x_{nt}$	$r_{nt}$	Compensation	Observations
1993	15.90	8,896 (26,269)	18.02 (46.15)	2.83 (7.24)	1.19 (0.45)	1.18 (0.51)	1,854 (12,412)	1574
1994	13.72	7,770 (25,284)	16.18 (43.41)	2.87 (5.04)	0.97 (0.29)	1.07 (2.52)	2,714 (10,909)	1876
1995	14.00	8,187 (28,650)	16.43 (44.41)	3.45 (33.40)	1.26 (0.47)	1.18 (0.64)	1,781 (13,252)	1867
1996	13.79	8,357 (29,029)	17.31 (45.92)	2.41 (17.20)	1.16 (0.38)	1.17 (0.87)	3,257 (14,824)	1926
1997	13.67	8,770 (31,797)	17.94 (47.96)	2.76 (41.40)	1.30 (0.48)	1.22 (3.06)	4,691 (17,791)	1997
1998	15.00	9,486 (40,145)	17.67 (45.91)	3.91 (71.30)	1.05 (0.53)	1.20 (1.11)	2,726 (18,530)	2012
1999	13.97	10,303 (43,087)	18.34 (45.75)	2.84 (11.57)	1.14 (0.76)	1.31 (8.27)	1,652 (21,631)	1970
2000	13.18	10,484 (45,936)	19.59 (54.08)	2.64 (8.31)	1.14 (0.68)	1.18 (1.50)	4,624 (21,641)	1865
2001	14.16	12,015 (52,064)	20.10 (56.50)	2.69 (14.90)	1.08 (0.54)	1.17 (1.86)	3,314 (18,842)	1851
2002	14.32	12,115 (57,166)	19.47 (54.51)	4.69 (105.00)	0.86 (0.42)	1.00 (2.43)	3,165 (16,077)	1877
2003	14.87	13,869 (66,331)	19.15 (52.85)	2.51 (35.20)	1.45 (0.64)	1.53 (16.10)	3,151 (18,830)	1814
2004	14.17	14,429 (70,812)	21.05 (64.83)	2.77 (9.39)	1.16 (0.37)	1.11 (1.38)	4,069 (17,195)	1687
2005	13.89	20,925 (89,832)	22.19 (52.34)	2.63 (12.27)	1.07 (0.36)	1.16 (1.63)	4,397 (19,992)	751

Assets in millions of 2000 \$US, Employees in thousands, Compensation in thousands of \$US.  
 $x_{nt}$  is excess financial return and  $r_{nt}$  is accounting return.

TABLE 3: ESTIMATE OF  $\varphi(j)$  PROBABILITIES

(A, W, D) <sup>†</sup>	Primary			Consumer			Services			Total	
	<i>N</i>	Good	Bad	<i>N</i>	Good	Bad	<i>N</i>	Good	Bad		
(S, S, S)	1	2,598	0.0917	0.1975	2,023	0.1227	0.1764	3,483	0.1249	0.1877	8,103
(S, L, S)	2	319	0.0141	0.0214	268	0.0121	0.0275	210	0.0040	0.0149	797
(S, L, L)	3	469	0.0257	0.0266	418	0.0229	0.0389	1,210	0.0337	0.0749	2,097
(S, S, L)	4	1,326	0.0763	0.0713	961	0.0725	0.0696	952	0.0434	0.0421	3,239
(L, S, S)	5	541	0.0272	0.0331	498	0.0308	0.0427	760	0.0248	0.0434	1,799
(L, L, S)	6	1,105	0.0635	0.0595	734	0.0593	0.0493	927	0.0164	0.0668	2,766
(L, L, L)	7	2,398	0.1118	0.1552	1,686	0.0879	0.1614	3,056	0.0865	0.1878	7,140
(L, S, L)	8	224	0.0127	0.0123	175	0.0145	0.0114	546	0.0262	0.0227	945
Total		8,980	0.423	0.577	6,762	0.423	0.577	11,144	0.360	0.640	26,886

<sup>†</sup>Firm Type is measured by the triplicate (A, W, D) where each element corresponds to (i.e., A is assets, W is the number of workers, and D is the debt to equity ratio) whether that element is above (i.e., L) or below (i.e., S) its industry average.

TABLE 4: CROSS SECTION SUMMARY OF RETURNS AND COMPENSATION BY STATE AND SECTOR

(A, W, D) <sup>†</sup>	Variable	Primary		Consumer		Services	
		Good	Bad	Good	Bad	Good	Bad
Return							
(S,S,S)	1	0.11 (0.56)	-0.10 (0.43)	0.10 (0.57)	-0.15 (0.47)	0.28 (0.94)	-0.05 (0.70)
(S,L,S)	2	-0.03 (0.40)	-0.14 (0.34)	-0.01 (0.31)	-0.07 (0.35)	0.02 (0.43)	-0.10 (0.46)
(S,L,L)	3	-0.03 (0.36)	-0.11 (0.37)	0.07 (0.38)	-0.12 (0.36)	0.10 (0.41)	-0.05 (0.34)
(S,S,L)	4	0.07 (0.52)	-0.11 (0.44)	0.05 (0.62)	-0.11 (0.55)	0.20 (0.75)	-0.11 (0.82)
(L,S,S)	5	-0.01 (0.34)	-0.10 (0.39)	0.01 (0.41)	-0.12 (0.40)	0.08 (0.71)	-0.09 (0.52)
(L,L,S)	6	-0.07 (0.29)	-0.11 (0.33)	-0.06 (0.32)	-0.12 (0.34)	0.15 (0.61)	-0.05 (0.47)
(L,L,L)	7	-0.03 (0.27)	-0.13 (0.30)	-0.01 (0.41)	-0.16 (0.38)	0.02 (0.32)	-0.06 (0.37)
(L,S,L)	8	0.02 (0.30)	-0.13 (0.40)	0.07 (0.47)	-0.24 (0.49)	0.13 (0.85)	-0.12 (0.59)
Compensation							
(S,S,S)	1	3,889 (14,651)	670 (10,779)	3,397 (19,178)	-1,501 (15,235)	6,063 (20,034)	1,701 (17,316)
(S,L,S)	2	4,384 (9,381)	2,339 (14,243)	4,922 (30,677)	-486 (23,882)	8,015 (24,615)	-1,183 (25,740)
(S,L,L)	3	3,742 (11,903)	521 (15,710)	9,194 (19,898)	821 (11,820)	7,096 (14,740)	2,274 (14,363)
(S,S,L)	4	2,522 (9,855)	721 (8,851)	3,977 (14,844)	908 (11,504)	4,154 (16,068)	-150 (14,255)
(L,S,S)	5	3,079 (20,381)	-850 (15,773)	4,235 (20,107)	-510 (16,940)	3,386 (18,844)	1,629 (19,287)
(L,L,S)	6	4,154 (13,375)	2,422 (16,220)	4,727 (20,989)	-429 (21,784)	8,035 (24,244)	5,496 (26,472)
(L,L,L)	7	5,781 (12,807)	2,200 (12,208)	6,897 (19,288)	2,775 (19,118)	9,846 (24,075)	5,595 (19,936)
(L,S,L)	8	4,396 (14,831)	-3,729 (18,890)	4,742 (19,288)	-2,442 (14,448)	5,647 (20,347)	1,718 (17,612)

Compensation in thousands of 2000 \$US. <sup>†</sup>Firm Type is measured by the triplicate (A, W, D) where each element corresponds to (i.e., A is assets, W is the number of workers, and D is the debt to equity ratio) whether that element is above (i.e., L) or below (i.e., S) its industry average.

TABLE 5: SUBSAMPLING<sup>‡</sup> ESTIMATES OF THE RISK AVERSION PARAMETER'S 95% CONFIDENCE REGION FOR THE PURE MORAL HAZARD MODEL (Based on Profit Maximization)

Firm Type (A, W, D) <sup>†</sup>	Sector		
	Primary	Consumer Goods	Service
(S,S,S)	(0.01, 13.4)	(0.01, 0.43)	(0.01, 1.61)
(S,L,S)	(0.01, 1.61)	(0.01, 6.61)	(0.01, 1.78)
(S,L,L)	(0.01, 2.66)	(0.01, 3.61)	(0.01, 0.24)
(S,S,L)	(0.01, 4.88)	(0.01, 16.4)	(0.01, 3.26)
(L,S,S)	(0.01, 9.9)	(0.01, 0.29)	(0.01, 0.21)
(L,L,S)	(0.02, 4.0)	(0.02, 20.1)	(0.01, 0.35)
(L,L,L)	(0.02, 2.66)	(0.01, 4.88)	(0.01, 0.43)
(L,S,L)	(0.01, 4.88)	(0.01, 0.39)	(0.01, 18.2)
Observations	7,796	5,600	8,536

<sup>‡</sup>The subsampling procedure was done using 100 replications of subsamples with 3,000 observations each. <sup>†</sup>Firm Type is measured by the triplicate (A, W, D) where each element corresponds to (i.e., A is assets, W is the number of workers, and D is the debt to equity ratio) whether that element is above (i.e., L) or below (i.e., S) its industry average. All specification automatically imposed cost minimization.

TABLE 6: SUBSAMPLING<sup>†</sup> ESTIMATES OF THE RISK AVERSION PARAMETER'S 95% CONFIDENCE REGION FOR THE PURE MORAL HAZARD MODEL (Exclusion Restriction of Independence of the Cost of Effort and Accounting Return)

Firm Type (A, W, D) <sup>†</sup>	Cost of Shirking			Cost of Working		
	$\alpha_{11t} = \alpha_{12t}$			$\alpha_{21t} = \alpha_{22t}$		
	Primary	Consumer Goods	Service	Primary	Consumer Goods	Service
(S,S,S)	(0.4, 0.5)∪(2.0, 13.4)	(0.01, 0.43)	(0.29, 1.61)	(2.18, 8.09)	(0.26, 0.43)	(0.48, 1.61)
(S,L,S)	(0.16, 0.39)	(0.01, 0.48)	(0.06, 0.21)	(0.01, 0.02)	(1.31, 1.97)	(0.01, 0.03)
(S,L,L)	(0.01, 0.43)	(0.07, 0.16)	(0.01, 0.24)	(0.03, 2.18)	(2.41, 3.99)	(0.06, 0.24)
(S,S,L)	(0.17, 0.72)	(0.17, 0.59)	(2.18, 3.26)	(0.01, 4.88)	(1.45, 16.4)	(2.41, 3.26)
(L,S,S)	(0.19, 9.9)	(0.09, 0.29)	(0.13, 0.21)	(0.01, 0.03)∪(1.31, 2.95)	(0.17, 0.29)	(0.05, 0.21)
(L,L,S)	(0.02, 0.43)	(0.02, 0.59)	(0.06, 0.26)	(0.02, 0.07)	(0.02, 0.03)	(0.03, 0.35)
(L,L,L)	(0.02, 0.29)	(0.01, 0.43)	(0.01, 0.21)	(0.02, 0.05)	(0.02, 0.04)	(0.17, 0.43)
(L,S,L)	(0.08, 0.53)	(0.16, 0.24)	(3.26, 16.4)	(0.01, 0.02)	(0.01, 0.02)	(5.40, 18.2)
Observations	7,796	5,600	8,536	7,796	5,600	8,536

<sup>†</sup>The subsampling procedure was done using 100 replications of subsamples with 3,000 observations each. <sup>†</sup>Firm Type is measured by the triplicate (A, W, D) where each element corresponds to (i.e., A is assets, W is the number of workers, and D is the debt to equity ratio) whether that element is above (i.e., L) or below (i.e., S) its industry average. All specifications imposed cost minimization and profit maximization. The risk aversion parameter from the interaction across sector, size and leverage is  $\gamma \in [0.02, 0.21]$  for the pure model hazard based only cost minimization and profit maximization.

TABLE 7: SUBSAMPLING ESTIMATES OF THE 95% CONFIDENCE REGION OF THE RISK AVERSION PARAMETER FOR THE MOST<sup>‡</sup> RESTRICTED HYBRID MORAL HAZARD MODEL

Sector	Observations	Confidence Region
Primary	7,796	(0.002 , 0.26)∪(0.37 , 0.42)
Consumer Goods	5,600	(0.002 , 0.13)∪(0.19 , 0.57)
Service	8,536	(0.27 , 0.53)

<sup>‡</sup>The restrictions imposed are profit maximization and equalization of preference parameters across size, leverage, and time. The subsampling procedure was done using 100 replications of subsamples with 3000 observations each.

TABLE 8: CONFIDENCE REGIONS OF CERTAINTY EQUIVALENT WAGE OF SHIRKING AND WORKING (Measured in millions of 2000 \$US)

Firm Type (A, W, D) <sup>†</sup>	Primary			Consumer			Services		
	Pure		Hybrid	Pure		Hybrid	Pure		Hybrid
	Bad	Good	All	Bad	Good	All	Bad	Good	All
	Wage for shirking								
(S,S,S)	(-0.78, -0.78)	(0.67, 0.67)	(-1.20, -1.18)	(-3.67, -3.66)	(-1.99, -1.98)	(-4.62, -4.59)	(-1.13, -1.10)	(0.87, 0.88)	(-1.37, -1.36)
(S,L,S)	(1.21, 1.24)	(2.96, 2.96)	(0.49, 0.50)	(-6.05, -6.04)	(0.70, 0.81)	(-4.27, -4.18)	(-5.53, -5.43)	(3.82, 3.87)	(0.09, 0.12)
(S,L,L)	(-0.24, -0.22)	(2.17, 2.18)	(0.70, 0.72)	(-0.72, -0.71)	(6.88, 6.91)	(1.84, 1.84)	(0.68, 0.70)	(3.55, 3.56)	(0.40, 0.41)
(S,S,L)	(-0.42, -0.42)	(1.34, 1.34)	(0.54, 0.54)	(-1.40, -1.40)	(0.69, 0.70)	(-0.27, -0.26)	(-2.21, -2.21)	(0.76, 0.78)	(-1.34, -1.34)
(L,S,S)	(-3.84, -3.84)	(-2.68, -2.65)	(-4.46, -4.43)	(-3.00, -2.98)	(1.65, 1.71)	(-1.21, -1.20)	(-1.19, -1.16)	(0.66, 0.70)	(-0.61, -0.57)
(L,L,S)	(0.72, 0.74)	(2.96, 2.98)	(1.81, 1.83)	(-4.66, -4.65)	(1.38, 1.41)	(-0.78, -0.74)	(1.09, 1.20)	(3.90, 3.94)	(1.89, 2.05)
(L,L,L)	(1.24, 1.26)	(4.63, 4.65)	(2.60, 2.62)	(0.51, 0.54)	(3.21, 3.25)	(0.78, 0.79)	(1.45, 1.48)	(6.34, 6.41)	(3.42, 3.44)
(L,S,L)	(-5.60, -5.57)	(1.07, 1.08)	(-3.07, -3.01)	(-4.28, -4.28)	(3.50, 3.51)	(0.23, 0.26)	(-1.59, -1.57)	(2.44, 2.46)	(0.31, 0.32)
	Wage for working under perfect monitoring								
(S,S,S)	(0.36, 0.48)	(3.16, 3.43)	(1.05, 1.09)	(-2.19, -1.94)	(2.40, 2.99)	(-0.77, 0.67)	(0.42, 0.81)	(4.07, 4.49)	(1.41, 1.51)
(S,L,S)	(1.78, 2.17)	(4.32, 4.39)	(2.52, 2.58)	(-1.46, -0.83)	(5.37, 6.66)	(-0.36, 0.15)	(-1.47, -0.24)	(4.98, 5.44)	(-1.38, 1.09)
(S,L,L)	(0.16, 0.37)	(2.94, 3.14)	(1.30, 1.35)	(0.68, 0.85)	(8.52, 8.95)	(3.10, 3.19)	(1.72, 1.96)	(6.14, 6.36)	(2.77, 2.84)
(S,S,L)	(0.30, 0.35)	(1.96, 2.04)	(1.08, 1.10)	(0.60, 0.77)	(2.58, 2.85)	(1.36, 1.42)	(-1.02, -0.88)	(3.20, 3.46)	(8.37, 8.95)
(L,S,S)	(-1.22, -0.97)	(2.19, 2.81)	(-0.17, 0.06)	(-1.09, -0.75)	(3.76, 4.35)	(0.39, 0.51)	(0.91, 1.46)	(2.65, 3.10)	(0.98, 1.11)
(L,L,S)	(1.67, 1.93)	(3.88, 4.12)	(2.52, 2.59)	(-0.87, -0.30)	(3.54, 4.10)	(0.89, 1.03)	(4.25, 5.41)	(5.92, 6.44)	(3.47, 3.73)
(L,L,L)	(1.84, 2.06)	(5.15, 5.32)	(2.97, 3.03)	(2.44, 2.90)	(6.11, 6.71)	(3.15, 3.28)	(4.69, 5.27)	(8.85, 9.76)	(5.22, 5.40)
(L,S,L)	(-4.22, -3.75)	(3.46, 3.67)	(-0.89, 0.78)	(-2.55, -2.34)	(4.41, 4.62)	(0.95, 1.02)	(0.52, 0.83)	(4.92, 5.27)	(2.44, 2.54)

<sup>†</sup>Firm Type is measured by the triplicate (A, W, D) where each element corresponds to (i.e., A is assets, W is the number of workers, and D is the debt to equity ratio) whether that element is above (i.e., L) or below (i.e., S) its industry average. The number observations are: 7,796, 5,600, and 8,536 in the Primary, Consumer, and Service sectors, respectively. The Shirking Wage is  $w_t^{(1)} = [\gamma(b_t - 1)]^{-1} b_{t+1} \log(\alpha_1)$  and the Working Wage under perfect monitoring is  $w_t^{(2)} = [\gamma(b_t - 1)]^{-1} b_{t+1} \log(\alpha_2)$ . All calculations in this table are done using the median bond price in the data. The risk aversion parameters used are the interaction across sector, size and leverage of the most parasitonomous non-rejected specification—i.e.,  $\gamma \in (0.02, 0.21)$  for the pure model hazard and  $\gamma \in (0.37, 0.42)$  for the hybrid moral hazard model.

TABLE 9: CONFIDENCE REGIONS OF THE AGENCY COSTS (Gross Losses are measured in percentage and Risk Premium is in millions of 2000 US\$)

(A, W, D)	Primary		Consumer		Services	
	Pure	Hybrid	Pure	Hybrid	Pure	Hybrid
Gross Losses to Firms from Shirking						
(S,S,S)	(12.53, 13.36)	(12.22, 12.44)	(15.15, 16.35)	(14.21, 14.53)	(15.32, 17.05)	(13.80, 14.19)
(S,L,S)	(7.27, 8.09)	(6.57, 6.72)	(14.11, 15.22)	(14.74, 14.99)	(11.72, 14.11)	(1.46, 1.53)
(S,L,L)	(3.45, 4.17)	(5.97, 6.42)	(8.71, 9.86)	(19.41, 20.87)	(8.19, 9.14)	(6.97, 7.15)
(S,S,L)	(8.00, 8.58)	(14.44, 14.76)	(16.55, 17.46)	(17.45, 17.60)	(16.07, 17.32)	(23.22, 23.35)
(L,S,S)	(17.73, 18.88)	(17.35, 17.65)	(10.14, 11.72)	(21.90, 22.45)	(11.29, 13.30)	(9.60, 10.03)
(L,L,S)	(4.64, 5.52)	(4.60, 4.63)	(10.45, 11.52)	[0.00, 3.83]	(11.23, 13.79)	(9.01, 9.51)
(L,L,L)	(3.17, 3.99)	[0.00, 6.74]	(9.09, 10.47)	(7.97, 8.29)	(10.54, 12.00)	(10.77, 11.01)
(L,S,L)	(10.61, 11.66)	(10.58, 10.72)	(9.89, 10.80)	(7.04, 7.21)	(16.13, 17.65)	(22.62, 26.11)
Risk Premium from Agency						
(S,S,S)	(0.020, 0.201)	(0.369, 0.416)	(0.042, 0.435)	(0.807, 0.909)	(0.044, 0.451)	(0.813, 0.916)
(S,L,S)	(0.033, 0.308)	(0.526, 0.586)	(0.092, 0.939)	(1.716, 1.930)	(0.113, 1.172)	(2.140, 2.425)
(S,L,L)	(0.025, 0.240)	(0.425, 0.476)	(0.029, 0.297)	(0.692, 0.781)	(0.026, 0.274)	(0.533, 0.604)
(S,S,L)	(0.007, 0.076)	(0.141, 0.159)	(0.024, 0.247)	(0.438, 0.493)	(0.021, 0.222)	(0.446, 0.505)
(L,S,S)	(0.046, 0.477)	(0.868, 0.981)	(0.048, 0.503)	(0.947, 1.070)	(0.056, 0.581)	(1.017, 1.149)
(L,L,S)	(0.028, 0.288)	(0.511, 0.577)	(0.062, 0.629)	(1.144, 1.288)	(0.113, 1.169)	(2.036, 2.300)
(L,L,L)	(0.023, 0.234)	(0.443, 0.500)	(0.056, 0.580)	(1.046, 1.182)	(0.075, 0.767)	(1.371, 1.548)
(L,S,L)	(0.037, 0.376)	(0.848, 0.960)	(0.023, 0.233)	(0.552, 0.622)	(0.035, 0.367)	(0.703, 0.795)

Firm size and leverage is measured by the triplicate (A, W, D) where each element corresponds to (i.e., A is assets, W is the number of workers, and D is the debt to equity ratio) whether that element is above (i.e., L) or below (i.e., S) its industry average. The number observations are: 7,796, 5,600, and 8,536 in the Primary, Consumer, and Service sectors respectively. All calculations in this table are done using the average over bond price in the data. The risk aversion parameter used are the interaction across sector, size and leverage of the most parsimonious non-rejected specification—i.e.,  $\gamma \in [0.02, 0.21]$  for the pure model hazard and  $\gamma \in [0.37, 0.42]$  for the hybrid moral hazard model.

Figure 1: Abnormal Return Densities and Compensation Schedule by Accounting Return State

