

Identifying and Testing Models of Managerial Compensation

GEORGE-LEVI GAYLE

Department of Economics, Washington University in St Louis

and

ROBERT A. MILLER

Tepper School of Business, Carnegie Mellon University

First version received June 2009; final version accepted November 2014 (Eds.)

This article analyses the identification and empirical content of the pure moral hazard (PMH) and the hybrid moral hazard (HMH) principal–agent models. The PMH model has hidden actions, while the HMH model has hidden information in addition to hidden actions. In both models, agents are risk averse and principals are risk neutral. The article derives the equilibrium restrictions from the optimal contract and uses the restrictions to show that the models have empirical content. For any given risk-aversion parameter, the models' other parameters are non-parametrically point identified. The risk-aversion parameter—and hence the model—are, however, only partially identified. Management's ability to manipulate accounting reports arises endogenously within HMH models, but not in all versions of PMH models. We use our framework to investigate whether shareholders contract with management recognizing that accounting reports are susceptible to manipulation and, therefore, endogenous to the incentives offered to management. The data reject all models in which accounting reports are verifiable. Furthermore, the version of the PMH in which accounting reports can be manipulated is rejected if expected compensation is restricted to be positive.

Key words: Managerial compensation, Moral hazard, Hidden information, Empirical content, Partial identification, Semiparametric

JEL Codes: C10, C12, C13, C14, J30, J33, M50, M52, M55

1. INTRODUCTION

The principal-agent model is the main theoretical underpinning for why managers are compensated with stocks, options, and bonuses instead of a flat salary: asymmetric information considerations are prevalent in the market for top managers of firms with dispersed ownership. The activities of these managers, paid to create value by organizing the firm's resources, are rarely observed or are hard for shareholders to monitor. Through their everyday work, managers come to know more about the state of the firm and its future profits than the shareholders do. These issues give rise to an asymmetric-information problem. Empirical research on managerial compensation seeks to identify and quantify the effects of asymmetric information and assess its impact on welfare, competition and policy.

Many useful tests of the theory of agency can be derived and performed without estimating a structural model (*e.g.* Jensen and Murphy, 1990; Hall and Liebman, 1998; Aggarwal and Samwick, 1999). However, a more structured approach is needed to quantify

asymmetric information's welfare loss, the efficiency of current pay practices, and the potential impact of policy reforms:¹ the return to managerial effort, the cost of managerial effort, and other information relevant for evaluating public policy reforms is unobserved by researchers. Recently, a handful of papers have specified, estimated, and conducted welfare analysis of executive-compensation contracting models (*e.g.* Margiotta and Miller, 2000; Gayle and Miller, 2009a, 2009b; Li, 2013; Gayle *et al.*, 2014), but none analyses the identification and empirical content of the standard paradigm for this literature, where agents are risk averse and principals are risk neutral. Identification presumes the probability distribution defining the population for the data comes directly from an unknown model belonging to a known class of models, and determines how many models within that class generate the same probability distribution. Empirical content determines whether the class of models imposes any restrictions that can be falsified by probability distributions generating the data. Therefore, identification and empirical content are fundamental to empirical research on the principal-agent model, even if a structural approach is not pursued in estimation.

This article analyses the identification and empirical content of the pure moral hazard (PMH) and the hybrid moral hazard (HMH) principal-agent models. The PMH model has hidden actions, while the HMH model has hidden information as well. In both models, the agent is risk averse and the principal is risk neutral. We derive the equilibrium restrictions from optimal contracting to predict the shape of the compensation schedule and fully characterize the empirical content of these models. We show that all the other parameters can be expressed as mappings of the risk-aversion parameter and the probability distribution of the data-generating process. This proves the model is non-parametrically point identified up to a known risk-aversion parameter, for which we establish sharp and tight bounds. We fully characterize the empirical content of this class of models by proving that some probability distributions cannot be rationalized by any risk-aversion parameter. We propose estimation and testing procedures, by first inferring bounds for the risk-aversion parameter, and then non-parametrically estimating the model's remaining parameters for all risk-aversion parameters satisfying the bounds. Our analysis can accommodate observed heterogeneity, dynamics, and - as shown in the [Supplementary Appendix](#) - some forms of unobserved heterogeneity.

The second part of the article illustrates the application of these methods to executive compensation. We investigate whether there is evidence that shareholders contract with management recognizing that financial reporting is susceptible to manipulation by managers and therefore endogenous to the incentives offered to management. Within the basic PMH model, which is the dominant paradigm of the empirical managerial compensation literature, managers lack the discretion to manipulate their reports on accounting earnings (and other financial events); all financial reports are interpreted as truthful disclosures enforced by auditors regardless of the incentives provided by shareholders to elicit information. In contrast, the HMH model can rationalize why managerial compensation is affected by accounting earnings statements when auditing conventions permit managerial discretion in deciding how to measure earnings. In the HMH model, accounting reports are truthful because the optimal contract satisfies the truth-telling and sincerity constraints.² There are other versions of the PMH model in which accounting reports can be manipulated; they however have different implications for the relationship between accounting reports and the primitives of the model than both the basic PMH and HMH models

1. Current proposals for reforming public policy on managerial compensation include capping their total compensation, restricting performance-based compensation, and differentially taxing the various components of total compensation.

2. The truth-telling constraint guarantees that the manager finds it in his best interest to report truthfully and the sincerity constraint ensures that he finds it optimal to report the truth and work in the shareholders best interest.

do. Reduced-form econometric investigations show that accounting returns affect managerial compensation, but cannot resolve whether managers exercise discretion in their accounting reports: both models can rationalize why accounting earnings affects compensation. This is why structural econometric analysis is required to determine whether managers need an incentive in the compensation package to truthfully reveal the firm's financial position. Our data set, covering 2610 firms and their chief executive officers and spanning the years 1992 to 2005, was constructed from Standard & Poor's ExecuComp, S&P COMPUSTAT, and the Center for Research in Security Prices (CRSP) databases. We test various specifications of the PMH and HMH models and use the unrejected models to estimate the cost of asymmetric information. The data reject all versions of the PMH model in which accounting reports are verifiable, but not the HMH model, and a version of the PMH model in which accounting reports can be manipulated. In this version of the PMH model, we find that managers are willing to pay shareholders for the privilege of holding the job of CEO. This version of the PMH model can be interpreted as a specification of the HMH in which the truth-telling and sincerity constraints are either redundant or ignored. They might be redundant because a monitoring technology imposing disutility on the principal renders these constraints redundant, or ignored because shareholders do not fully optimize over the contract space. The HMH model rationalizes the premium paid to managers for truthfully reporting good news in return for being held to a higher standard, without resorting to such unbelievable work ethics; consequently we reject this specialization.

The rest of the article is organized as follows; Section 2 presents the analysis of the static PMH model while Section 3 presents the analysis of the static HMH model. Section 4 extends the PMH and HMH models to a dynamic environment. Section 5 outlines the estimation and testing procedure while illustrating the methods in an application. Section 6 concludes and discusses possible extensions. The proofs to the results in the main text are collected in an Appendix while a [Supplementary Appendix](#) presents the empirical implementation of the application and the analysis of possible extensions in more detail. Below we discuss the related literature.

Related literature: This article is related to a very small literature on the identification of principal-agent models and a slightly larger empirical literature on the estimation of contract-theory models of managerial compensation. The closest paper to ours is [Perrigne and Vuong \(2011\)](#). They showed that static contract models with adverse selection and moral hazard are non-parametrically point identified, yet our results show that the PMH and HMH models are only set identified. There are several important differences between these non-nested specifications that explain why. The PMH model in this article is based on [Grossman and Hart \(1983\)](#) and the HMH model is based on [Myerson \(1982\)](#), whereas [Perrigne and Vuong \(2011\)](#) used another model based on [Laffont and Tirole \(1986\)](#). In [Laffont and Tirole \(1986\)](#) a first-best allocation is achieved in the absence of hidden information; for this reason [Perrigne and Vuong \(2011\)](#) called it a model of false moral hazard (FMH). In contrast, the HMH model specializes to the PMH model absent hidden information, so neither achieves a first-best allocation. In the FMH model hidden actions and hidden information enter additively into the production function, so in the principal's optimization problem all the private information can be reconciled through one incentive-compatibility constraint. However, in the HMH model, hidden actions and hidden information do not enter additively into the output production function, so several constraints are required in the optimal programming problem to induce the agent to produce a given level of effort and also truthfully report on private information. Furthermore, [Perrigne and Vuong \(2011\)](#) analysed identification when contracts are linear, one of many optimal contracts in the FMH model. In our framework, the optimal allocation is unique and non-linear. Finally, both papers make parametric assumptions. [Perrigne and Vuong \(2011\)](#) assumed hidden action and hidden information enter the production of output in an additively separable manner and that agents

are risk neutral. We allow hidden action and hidden information to enter the output production function in a general way but limit action and information sets to discrete support. We also allow agents to be risk averse, but restrict utility to the constant absolute risk-aversion (CARA) class.³

Despite the differences between the two sets of models, some features of the results are comparable. We show that the model's other parameters are non-parametrically point identified if the risk preference is known, and in [Perrigne and Vuong \(2011\)](#) this is indeed the case since they assumed agents are risk neutral. We further show that if the econometrician does not know risk preferences, the model is only semiparametrically set identified, an identification result that has analogues in the literature on first-price auctions. [Guerre et al. \(2009\)](#) showed that the general first-price auction model with risk-averse bidders is not identified, while [Campo et al. \(2011\)](#) showed that first-price auction models are semiparametrically point identified if bidders' utility functions are specified parametrically. [Campo et al. \(2011\)](#) achieved point identification by imposing exclusion restrictions on the distribution of valuations. In contrast, we achieve set identification using only the contract-selection criteria endogenous to the model. As our application demonstrates, this weaker but more robust result suffices to generate empirical content with welfare implications pertaining to the importance of asymmetric information.

Our article also contributes to previous work that quantifies the economic significance of incentives in the executive labour market, adding to work by [Haubrich \(1994\)](#), [Margiotta and Miller \(2000\)](#), and [Gayle and Miller \(2009b\)](#). [Haubrich \(1994\)](#) calibrated a PMH model to demonstrate that the observed pay for performance in the executive market is consistent with that generated by the optimal contract. [Margiotta and Miller \(2000\)](#) developed and estimated welfare measures to evaluate the importance of hidden information. The main finding of [Gayle and Miller \(2009b\)](#) is that the greater volatility in managerial compensation over the last 50 years can be directly attributed to increased firm size. All three studies used fully parametric models and deployed nested-fixed-point full-solution techniques to structurally estimate a PMH model using different estimation methods and data from different industrial sectors, executive ranks, and time periods. The set identification results in this article imply that the PMH models in [Margiotta and Miller \(2000\)](#) and [Gayle and Miller \(2009b\)](#) achieve point identification from the functional form assumptions: that the distribution of output and the likelihood ratio are derived from a truncated normal parent distribution and that risk preferences are constant over time.

We use managerial compensation as the motivation for and the empirical example of the moral hazard models studied in this article. However, the results obtained and the techniques developed in this article apply to other principal-agent settings where the compensation (or payment) and some ex post performance indicator are observed. Empirical models that fit these conditions have been used in many areas of economics including agricultural economics, health economics, financial economics, industrial organization, and labour economics. These applications include the livestock production contracts ([Dubois and Vukina, 2004](#)), sharecropping contracts ([Laffont and Matoussi, 1995](#); [Akerberg and Botticini, 2002](#)), moral hazard and financial structure ([Biais et al., 2000](#)), medical insurance ([Vera-Hernandez, 2003](#), [Cardon and Hendel, 2001](#)), and bonus and piece rates for non-executive worker (a non-exhaustive list includes [Ferrall and Shearer, 1999](#); [Shearer, 1996, 2004](#); [Paarsch and Shearer, 2000](#); [Copeland and Monnet, 2009](#); [Lazear, 2000](#)).

Starting with [Myerson \(1982\)](#), the HMM model has been extensively studied in the theoretical literature, but we believe ours is the first article to empirically test and estimate it.⁴ Likewise, this is the first article to incorporate accounting information into an empirical structural model of

3. Risk aversion is a standard assumption in the principal-agent setting when studying executive compensation. If agents are risk neutral, then a first-best outcome can be achieved by the principal selling the firm to the agent.

4. See [Mas-Colell, Whinston and Green \(1995\)](#), Chapter 14 for a theoretical treatment of the hybrid model.

contracting, and we find data on accounting returns plays a subtle yet critical role. If accounting data are not used—or, alternatively, are treated as a report on hidden information in the HMH model—estimates obtained from structural models of executive compensation are robust to a variety of welfare measures, econometric techniques, and data sources. For example, our estimates in the HMH model for the risk premium, and for losses firms would incur if they ignored moral hazard, are quite close to those found in the PMH models by Margiotta and Miller (2000) and Gayle and Miller (2009b) using different estimation methods and different data sets.⁵ If, however, accounting data are used, stable risk preferences over time are compatible with the HMH model but not the PMH model, and the estimates of the non-pecuniary benefits to the manager from shirking are very sensitive to assumptions about hidden information.

2. PURE MORAL HAZARD

To explain our approach to identification, estimation and testing, we first analyse a simple principal-agent model. We set up a static model of pure moral hazard and derive the principal's cost-minimizing optimal contract for two effort levels by the agent, working or shirking. Comparing the expected revenue of both contracts yields the profit-maximizing contract. Then we analyse the identification and empirical context of this model. For notational simplicity, we assume that the principal is profit maximizing and that the optimal contract is based on revenue; there are no costs aside from the agent's compensation.⁶

2.1. A benchmark model

At the beginning of the period, a risk-neutral principal proposes to a risk-averse agent a compensation plan that depends on the future realization of 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 maximizing the principal's expected revenue, called working, and accepting employment from the principal but following the objectives he would pursue if he were paid a fixed wage, called shirking. The decision to accept or reject the offer is observed by the principal, but the 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 remaining revenue is profit to the principal. We introduce the notation and the model and then solve for the cost-minimizing contracts that elicit diligence and shirking.

Notation: We denote the agent's workplace employment decision 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, $l_0 + l_1 + l_2 = 1$. Revenue to the principal is denoted by X , a random variable drawn from a probability distribution determined by the agent's work routine. After x , a realization of X , is revealed to 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

5. Margiotta and Miller (2000) used data for from a subset of the primary sector for the period 1944–1979. Gayle and Miller (2009b) compared results from this period of time with later data for the period 1993–2004.

6. More generally, the principal observes a signal correlated with the agent's action. Using revenue and ignoring other costs simplifies the presentation by eliminating the need to deal with additional conditional expectation functions without affecting the basic arguments.

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:

$$E[xg(x)] \equiv \int xf(x)g(x)dx < \int xf(x)dx \equiv E[x]. \quad (1)$$

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

$$E[g(x)] \equiv \int g(x)f(x)dx = 1. \quad (2)$$

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

$$\lim_{x \rightarrow \infty} [g(x)] = 0. \quad (3)$$

Intuitively, this assumption states that a truly extraordinary performance can only be attained if the agent works. We also 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.⁷ Given regularity condition (3), we assume for notational ease that the support of X is the real line and that $f(x)$ and $g(x)f(x)$ are defined on the full support of X . Without this common support assumption the first-best resource allocation is obtained by the principal severely punishing the agent when x in the support of $g(x)f(x)$ but not in the support of $f(x)$ occurs.

We assume the agent maximizes utility, which is exponential in compensation:

$$-l_0 - l_1\alpha_1 E \left[e^{-\gamma w(x)} g(x) \right] - l_2\alpha_2 E \left[e^{-\gamma w(x)} \right], \quad (4)$$

where, without further loss of generality, we normalize the utility of the outside option to negative one. Thus, γ is the coefficient of absolute risk aversion, and α_j is a utility parameter with consumption equivalent $-\gamma^{-1} \log(\alpha_j)$ measuring 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 working, giving rise to a conflict of interest between the principal and the agent: The latter prefers shirking ($\alpha_1 < \alpha_2$), yet the principal prefers the agent to work ($E[xg(x)] < E[x]$).

The assumption of CARA utility pervades the empirical literature on executive compensation and contract theory.⁸ The basic setup of the PMH model assumes that revenue, x , is continuous, but this assumption is not critical to our derivations and results. We could instead assume that x has discrete support and interpret $f(x)$ and $f(x)g(x)$ as probability mass functions instead of densities. To rule out the possibility of a trivial first-best allocation, we would maintain the assumption that each point in the support of x has positive mass under both $f(x)$ and $f(x)g(x)$ and replace regularity condition (3) with $g(\max\{x\}) = 0$. The HMH model we develop in the next section

7. The assumption that $g(x)$ is monotone decreasing in x , frequently used in the theoretical literature, is stronger than condition (3) imposed in this article.

8. Data on wealth are rarely available, yet except in the case of CARA utility, the optimal contract in principal-agent models typically depends on wealth.

could also be modified to handle the discrete support case without materially affecting the results we derive.

Optimal contracting: To induce the agent to accept the principal's offer and shirk, it suffices to propose a contract that gives the agent an expected utility of at least minus one.⁹ In this case, we require $w(x)$ to satisfy the inequality

$$\alpha_1 E \left[e^{-\gamma w(x)} g(x) \right] \leq 1. \tag{5}$$

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

$$\alpha_2 E \left[e^{-\gamma w(x)} \right] \leq 1 \tag{6}$$

and

$$\alpha_2 E \left[e^{-\gamma w(x)} \right] \leq \alpha_1 E \left[e^{-\gamma w(x)} g(x) \right]. \tag{7}$$

To attain expected revenue of $E[x]$ at minimal expected cost, the principal chooses a schedule $w(x)$ to minimize expected compensation, denoted by $E[w(x)]$, subject to inequalities (6) and (7). Alternatively, to attain expected revenue of $E[xg(x)]$ at minimal expected cost, the principal chooses a schedule $w(x)$ to minimize $E[w(x)]$ subject to inequality (5). 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 an agent to shirk is $\gamma^{-1} \ln(\alpha_1)$. To minimize the cost of inducing the agent to accept employment and work, the principal offers the contract*

$$w^o(x) \equiv \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[1 + \theta \left(\frac{\alpha_2}{\alpha_1} \right) - \theta g(x) \right], \tag{8}$$

where θ is the unique positive solution to

$$E \left[\frac{g(x)}{\alpha_2 + \theta[(\alpha_2/\alpha_1) - g(x)]} \right] = E \left[\frac{(\alpha_2/\alpha_1)}{\alpha_2 + \theta[(\alpha_2/\alpha_1) - g(x)]} \right]. \tag{9}$$

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 agent to uncertainty in a shirking contract by tying compensation to revenue. Hence, an agent paid to shirk is offered a fixed wage that just offsets his non-pecuniary benefits, $\gamma^{-1} \ln \alpha_1$. The certainty equivalent of the cost-minimizing contract that induces 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 $E[w^o(x)] - \gamma^{-1} \ln \alpha_2$.¹⁰ These two factors, that working is less enjoyable than shirking and less uncertainty in compensation is preferable, explain why compensating an

9. Following the literature, we assume that all contracts between the principal and agent are honoured.

10. 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, (6), is met with equality. This implies $\alpha_2 E \left[e^{-\gamma w(x)} \right] = 1$ or $\gamma^{-1} \ln \alpha_2 = -\gamma^{-1} \ln E \left[e^{-\gamma w(x)} \right]$. Thus, $E[w^o(x)]$ exceeds $\gamma^{-1} \ln \alpha_2$ if and only if $\exp E[\gamma w^o(x)]$ exceeds $E \left[e^{-\gamma w(x)} \right]$, which is true by Jensen's inequality if and only if $\gamma > 0$.

agent to align his interests with those of the principal is more expensive than merely paying him enough to accept employment.

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

$$\max\{0, \gamma E[xg(x)] - \ln(\alpha_1)\} \leq \gamma E[x - w^o(x)], \quad (10)$$

while a shirking contract is offered if and only if

$$\max\{0, \gamma E[x - w^o(x)]\} \leq \gamma E[xg(x)] - \ln(\alpha_1). \quad (11)$$

Otherwise, the principal does not employ the agent. Throughout the article, we analyse several versions of the PMH model; we denote the static homogenous model as PMH1 and reserve PMH for the generic class of pure moral hazard models.

2.2. Identification and empirical content in the PMH1 model

In this section, we specify the econometric model for the observed variables, taking account of possible measurement errors. The observables are ex post measures of compensation and revenue, (\tilde{w}, x) . We assume observed compensation is an error ridden measure of true compensation, w , defined either by the optimal contract from working in equation (8), $w = w^o(x)$, or the optimal shirking contract, $w = \gamma^{-1} \ln \alpha_1$. The error $\varepsilon \equiv \tilde{w} - w$ is assumed to be orthogonal to all variables of interest, specifically x . We do not observe the effort level demanded by the principal; our only premise is that the equilibrium distribution of (W, X) is identified from the observables.¹¹ The parameters of the model are characterized by $f(x) \in F$ and $g(x) \in G$, which together define the probability density functions of revenue, $(\alpha_1, \alpha_2) \in A$, the preference parameters for shirking and working (relative to the normalized utility from taking the outside option), and the risk-aversion parameter, $\gamma \in \Gamma_1$. Thus the identification problem reduces to whether a uniquely defined PMH1 model can be recovered from the structure $[F, G, A, \Gamma_1]$ with knowledge of (W, X) ; the empirical content problem is whether or not the structure $[F, G, A, \Gamma_1]$ can rationalize any distribution of (W, X) .

Appealing to the compensation equation (8), the regularity condition on $g(x)$ given by (3) and the fact that $g(x)$ is nonnegative, the agent's maximum compensation is

$$\lim_{x \rightarrow \infty} w^o(x) = \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[1 + \theta \left(\frac{\alpha_2}{\alpha_1} \right) \right] \equiv \bar{w}. \quad (12)$$

Thus, \bar{w} is identified by the maximum of the support of W .

There are three cases of possible effort choice by the principal to investigate; when it is optimal for managers to work, when it is optimal for managers to shirk, and when it is optimal for one type of principal to induce working and another type to induce shirking. Whether the agent works or shirks is identified from the distribution of (W, X) . When the agent works, equation (10) holds and w depends nontrivially on revenue. Because of the relevance of managerial compensation where managers are compensated with stocks, options and bonuses instead of a flat salary, we focus on the first case in the main text, in which principal seeks to overcome a moral hazard problem through the provision of appropriate incentives.¹² In this case, $f(x)$ is identified by the

11. In this section, we assume that w instead of \tilde{w} is observed because by assumption $w = E[\tilde{w}|X=x]$. In the [supplementary appendix](#), we consider a case normally encountered in practice where both ex post measures of compensation and revenue are measured with error.

12. In the [supplementary appendix](#), we also analyse identification in the two other cases.

marginal distribution of X and $w^o(x)$ are identified by the conditional expectation of W given X , *i.e.* $w^o(x) = E[W|X=x]$. In equilibrium, the principal selects the distribution from which x is drawn depending on the profit maximization conditions (10) and (11) and this decision is not observed by the econometrician. Therefore, identification of $f(x)$ is subject to a selection problem because depending on which contract is demanded by the principal x is drawn from different distributions, $f(x)$ or $f(x)g(x)$. However, under the assumption that the principal finds it optimal to induce working then the marginal distribution of X identifies $f(x)$.

Identifying $f(x)$ essentially reduces the structure to $[G, A, \Gamma_1]$. First, we show that if γ is known, then α_1 , α_2 , and $g(x)$ are point identified from the cost-minimization problem. This means that the set of observationally equivalent parameters can be indexed by the positive real number γ , the risk-aversion parameter. Secondly, we show that the principal's preference for working over shirking provides an additional inequality that helps delineate the values of observationally equivalent γ . Thirdly, we prove that the set of restrictions we have derived in the first two steps fully characterize the identified set. Finally, we show that there exist distributions of (W, X) that cannot be rationalized by the structure $[F, G, A, \Gamma_1]$.

Restrictions from cost minimization: Suppose γ is known, and define the mappings $g(x, \gamma)$, $\alpha_1(\gamma)$, and $\alpha_2(\gamma)$ as

$$g(x, \gamma) \equiv \frac{e^{\gamma \bar{w}} - e^{\gamma w^o(x)}}{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]} \tag{13}$$

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

$$\alpha_2(\gamma) \equiv \left\{ E[e^{-\gamma w^o(x)}] \right\}^{-1}. \tag{15}$$

These mappings are derived from the compensation equation (8), the participation constraint (6), and the incentive-compatibility constraint (7); their derivations are formally proved in Theorem 2.1 below. All three mappings inherit the basic structure of the model for any positive value of γ . That is, $g(x, \gamma)$ is a likelihood ratio, $\alpha_1(\gamma)$ and $\alpha_2(\gamma)$ are positive, and $\alpha_1(\gamma) < \alpha_2(\gamma)$. Integrating (13) over x , we get 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 (12), we see that $w(x) \rightarrow \bar{w}$, and hence $g(x, \gamma) \rightarrow 0$, as stipulated by the regularity condition in equation (3). This proves $g(x, \gamma)$ can be interpreted as a likelihood ratio satisfying (3) 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 γ , so $\alpha_1(\gamma)$ is also positive. Rearranging the expression for the ratio of $\alpha_1(\gamma)$ and $\alpha_2(\gamma)$, we obtain

$$\frac{\alpha_1(\gamma)}{\alpha_2(\gamma)} = \frac{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]}{e^{\gamma \bar{w}} - \left\{ E[e^{-\gamma w^o(x)}] \right\}^{-1}}. \tag{16}$$

Since the inverse function is convex, Jensen's inequality gives $E[e^{-\gamma w^o(x)}] > \left\{ E[e^{\gamma w^o(x)}] \right\}^{-1}$ or $\left\{ E[e^{-\gamma w^o(x)}] \right\}^{-1} < E[e^{\gamma w^o(x)}]$, and consequently $\alpha_1(\gamma) < \alpha_2(\gamma)$ for all positive γ as stipulated by the theoretical model.

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 from the distribution of (W, X) , we can construct, for any positive γ , 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 from data on compensation and revenue using equations (13), (14), and (15).

Theorem 2.1. *Suppose the distribution of (W, X) is generated by a parameterization of a pure moral hazard model with risk aversion γ^* . Then*

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

When x is continuous the basic ideas for the proof of this theorem are straightforward. Making $g(x)$ the subject of the compensation equation (8) and differentiating with respect to x yields

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

From this equation, it is evident that the slope is defined up to an unknown constant, $(\theta\alpha_2)^{-1}$; by the fundamental theorem of calculus a second unknown constant determines the level of $g(x)$. In our setup, the regularity condition (3) identifies one of the unknown constants; the fact that $E[g(x)] = 1$ identifies the other.

To obtain the level normalization, we exponentiate the compensation equation (8) and rearrange it to obtain

$$e^{\gamma w^o(x)} = \alpha_2 [1 + \theta(\alpha_2/\alpha_1) - \theta g(x)] = \alpha_2 [1 + \theta(\alpha_2/\alpha_1)] - \alpha_2 \theta g(x). \quad (17)$$

Note from equations (12) and (3) that $w^o(x) \rightarrow \bar{w}$ and $g(x) \rightarrow 0$ as $x \rightarrow \infty$, so equation (17) implies

$$e^{\gamma^* \bar{w}} = \alpha_2 [1 + \theta(\alpha_2/\alpha_1)]. \quad (18)$$

Subtracting equation (17) from (18) we obtain

$$\alpha_2 \theta g(x) = e^{\gamma^* \bar{w}} - e^{\gamma^* w^o(x)}. \quad (19)$$

equation (19) shows that the regularity condition (3) determines the level of $g(x)$ up to the slope normalization, $(\theta\alpha_2)^{-1}$.

Now taking expectations over x in equation (17) and appealing to the likelihood ratio property that $E[g(x)] = 1$ yields

$$E[e^{\gamma^* w^o(x)}] = \alpha_2 [1 + \theta(\alpha_2/\alpha_1) - \theta]. \quad (20)$$

Subtracting equation (20) from (18), we obtain the slope normalization:

$$\alpha_2 \theta = e^{\gamma^* \bar{w}} - E[e^{\gamma^* w^o(x)}]. \quad (21)$$

The proof of Lemma 2.1 shows that the participation constraint (6) is met with equality:

$$\alpha_2 E \left[e^{-\gamma w^o(x)} \right] = 1. \tag{22}$$

Rearranging (22) gives the formula for $\alpha_2(\gamma^*)$. Finally, the incentive-compatibility constraint is also met with equality, so

$$\alpha_1 E \left[e^{-\gamma w^o(x)} g(x) \right] = \alpha_2 E \left[e^{-\gamma w^o(x)} \right] = 1. \tag{23}$$

Substituting for $g(x)$ from equation (13) on the left side and rearranging to make α_1 the subject of the equation yields equation (14) evaluated at γ^* .

Restrictions from the principal’s choice of contract: The restrictions from cost minimization tie down all the parameters up to γ , but place no restrictions at all on γ . Imposing profit maximization, as opposed to cost minimization only, does limit the set of admissible γ . Since profit maximization implies that the expected profits from paying the agent $w^o(x)$ are higher than paying him $\gamma^{-1} \ln(\alpha_1)$, it follows from (10) that

$$E[x] - E[w^o(x)] - E[xg(x)] + \gamma^{-1} \ln(\alpha_1) \geq 0. \tag{24}$$

Substituting for $g(x)$ and α_1 from (13) and (14) into the LHS of (24) defines

$$Q_0(\gamma) \equiv \frac{\text{cov}(x, e^{\gamma w^o(x)})}{e^{\gamma \bar{w}} - E[e^{\gamma w^o(x)}]} - E[w^o(x)] + \gamma^{-1} \ln \left(\frac{1 - E[e^{\gamma w^o(x) - \gamma \bar{w}}]}{E[e^{-\gamma w^o(x)}] - e^{-\gamma \bar{w}}} \right) \tag{25}$$

and implies that $Q_0(\gamma^*) \geq 0$. This inequality restricts the set of γ that are admissible for the data-generating process.¹³

Tight and sharp bounds: Theorem 2.1 exploits the first-order conditions, the participation and incentive-compatibility conditions, and an inequality derived from 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 γ satisfying (25) is admissible. Thus Γ_1 , a Borel set of risk-aversion parameters defined as

$$\Gamma_1 \equiv \{\gamma > 0 : Q_0(\gamma) \geq 0\}, \tag{26}$$

indexes all the parameterizations that are observationally equivalent to the true model.

Tight means that Γ_1 covers the identified set. By construction Γ_1 is tight. A set is sharp if any value in the set, including end points, cannot be rejected as the value of the true model that generates the (W, X) distribution. Succinctly, every element in a sharp set is, by definition, observationally equivalent.¹⁴ Therefore, when a subset of the restrictions from the model define a sharp set, the set does not shrink from imposing extra restrictions derived from the same model to the set of constraints that the parameters must satisfy.

13. 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 demonstrates, this inequality does not impose any further restrictions.

14. For example see Frisch (1934, p. 86).

Theorem 2.2 establishes that Γ_1 is also sharp. That is, for any $\gamma > 0$ satisfying (26), equations (13), (14), and (15) define the primitives for the principal-agent models considered here that generate a compensation schedule from the optimal contract, given in Lemma 2.1, that matches the data-generating process. The constructed α_1 and α_2 satisfy the inequalities $0 < \alpha_1 < \alpha_2$; the constructed $g(x)$ is positive with $E[g(x)] = 1$ and $\lim_{x \rightarrow \infty} [g(x)] = 0$; every draw from the data set (x, w) satisfies $w = w(x)$, where $w(x)$ is the optimal contract for the constructed model. With reference to the previous paragraph, the restrictions embodied in inequality (26) suffice to obtain an observationally equivalent set of risk-aversion parameters, each indexing a different parameterization: additional restrictions derived from the structure cannot shrink (26).

Theorem 2.2. Γ_1 is sharp.

Four steps summarize how to identify the PMH1 model. The first three apply to any given γ . Suppose $f(x)$ and $w(x)$ are both known. First, α_2 is recovered, up to a normalization reflecting the outside option, from the participation constraint in equation (6) by computing $\int e^{-\gamma w(x)} f(x) dx$. Secondly, $g(x)$ comes from the mapping from x in to w as defined in the first-order condition for cost minimization, equation (8), plus the regularity condition in equation (3). Multiplying $f(x)$ with $g(x)$ yields the revenue density conditional on shirking. Thirdly, the incentive-compatibility constraint, equation (7), is met with equality in the cost-minimizing contract, so α_1 is recovered in the same way as α_2 , except $\int e^{-\gamma w(x)} f(x) g(x) dx$ instead of $\int e^{-\gamma w(x)} f(x) dx$ is computed. The fourth step exploits the profit-maximization condition in equation (10) to partially identify γ , and imparts empirical content to the PMH1 model: only a subset of positive real numbers, Γ_1 , satisfies Inequality (24) reflecting the principal's greater profits from $w(x)$ over a fixed-wage shirking contract, $\gamma^{-1} \ln \alpha_1$.

Empirical content of PMH1: Identification is concerned with recovering parameters of interest from data generated by a model: empirical content determines whether the model can be rejected by data generated by a different model. We now suppose the data are not necessarily generated by the PMH1 model. Under the null hypothesis, that the data could have been generated by a PMH1 model, Γ_1 is not empty; under the alternative hypothesis the data could not have been generated by a PMH1 model and Γ_1 is empty. This section concludes with a corollary that establishes PMH1 models have empirical content: they can be rejected.

The PMH1 model is flexible enough to entertain a non-monotone mapping from revenue to compensation in the equilibrium optimal contract. Nevertheless, an agency problem only exists because the principal expects higher revenue if the agent works, but the agent prefers to shirk. Frequently observing high levels of compensation paired with low revenue outcomes and vice-versa seem counterintuitive to the predictions of a PMH1 model. Such empirical regularities provide the basis for rejecting it. Using only conditions derived from the cost minimization problem, Theorem 2.1 demonstrates the data-generating process identifies a set of parameters that fully characterizes a well specified PMH1 model up to any positive γ . Only the profit condition (24) could be violated. If so, maintaining PMH1 requires the principal to motivate the agent to work, even though inducing work is more expensive to the principal and their goals are perfectly aligned when the agent shirks! The potential for observing this contradiction underlies its empirical content. Formally, the proof to Corollary 2.1 below shows that a profit condition is violated if compensation is monotone decreasing in revenue.

Corollary 2.1. *There exist joint distributions of (W, X) such that Γ_1 is empty.*

3. HYBRID MORAL HAZARD

The hidden action framework of the PMH model is the primary paradigm for rationalizing why executives are compensated in firm denominated securities instead of by a fixed wage. Many real-world situations with moral hazard also have hidden information. For example, it is widely believed that accounting reports can be manipulated by managers. A hybrid model incorporating both hidden actions and hidden information offers an attractive way of rationalizing the correlation between managerial compensation and accounting reports, because it concentrates attention on what compensation committees can accomplish with executives of their firms through optimal contracting.

It is straightforward to generalize the previous analysis to multiple states by indexing the probability distribution from which revenue is drawn. An HMH model differs from a PMH model with multiple states because the agent is not only subject to moral hazard, but also has private information about the state. This section presents an HMH model in order to explain how our analysis of the previous section must be modified to account for the information asymmetry; how the information asymmetry is captured by inequality constraints that restrict the range of feasible contracts; and, finally, how the principal's optimization problem is affected. We also present the differences that emerge in identification and empirical content.

3.1. *Theoretical framework*

The assumptions and preferences of the agent are unchanged from the previous sections: $l_0 = 1$ means the agent rejects the principal's offer; $l_1 = 1$ means the agent accepts the principal's offer and shirks; $l_2 = 1$ means the agent accepts the principal's offer and works; if $l_i = 1$ then $l_k = 0$ for $i \neq k$. The agent is an expected utility maximizer with utility exponential in compensation, and α_j is a utility parameter that measures the distaste from working at level $j \in \{1, 2\}$, where $\alpha_2 > \alpha_1$.

Output and states: We now assume there are two states $s \in \{1, 2\}$, and the probability state s occurs is identically and independently distributed with probability $\varphi_s \in (0, 1)$. If state s occurs, revenue is drawn from the probability density function $f_s(x)$ if the agent works and from $g_s(x)f_s(x)$ if the agent shirks. As in the previous section, we assume

$$\int xf_s(x)g_s(x)dx < \int xf_s(x)dx, \int g_s(x)f_s(x)dx = 1 \quad \text{and} \quad \lim_{x \rightarrow \infty} [g_s(x)] = 0 \quad (27)$$

for $s \in \{1, 2\}$. The distribution functions, $f_s(x)$ and $f_s(x)g_s(x)$, are common knowledge.¹⁵

Hidden information: As in the previous section, the principal observes l_0 but not $l_j \in \{0, 1\}$ for $j \in \{1, 2\}$. In the HMH model, the agent privately observes $s \in \{1, 2\}$, after the employment decision, l_0 , but before the effort choice.¹⁶ We assume the agent reports the state, $r \in \{1, 2\}$, to the principal before making his effort choice. If the agent reports the second state, $r = 2$, 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 constraint prevents the agent from lying when the first state occurs and

15. We assume throughout our analysis of the HMH model that α_2 and α_1 do not depend on s , but extending the analysis to deal with state-dependent preferences is straightforward.

16. Our model should be distinguished from the mixed models reviewed in Chapter 7 of Laffont and Martimort (2002), and Chapter 6 of Bolton and Dewatripont (2005). Their work addresses adverse selection, where the agent learns the state before making his participation decision.

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.¹⁷ If $s=2$ the agent then truthfully declares or lies about the firm's prospects by announcing $r \in \{1, 2\}$, but if $s=1$ he reports $r=1$. Define $h(x) \equiv \varphi_2 f_2(x) / \varphi_1 f_1(x)$ as the weighted likelihood ratio of the second state occurring relative to the first given any observed value of excess returns $x \in R$. We assume that

$$\lim_{x \rightarrow \infty} [h(x)] = \sup_{x \in R} [h(x)] \equiv \bar{h} < \infty. \quad (28)$$

If $h(x)$ was unbounded for some value of x' (because $f_2(x') > 0$ and $f_1(x') = 0$), then truth telling about $s=2$ could be enforced without cost because the principal would promise to severely punish the agent if the $r=1$ is reported but x' is subsequently drawn as the revenue outcome. Thus, (28) rules out this possibility by bounding $h(x)$. The equality in (28) captures the idea that if the agent works, then the likelihood of the second state is highest relative to the first-state likelihood when the revenue attains its highest values.

The agent's compensation is determined by what he discloses about the probability distribution of 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 $w_r(x)$.

Truth telling and sincerity constraints: Contracts between the principal and the agent that induce honest reporting in state 2 and working in both states must satisfy a participation constraint plus two incentive-compatibility constraints (one for each state) and two additional conditions 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

$$\int [1 - (\alpha_1/\alpha_2) g_s(x)] v_s(x) f_s(x) dx \equiv E_s \{ [1 - (\alpha_1/\alpha_2) g_s(x)] v_s(x) \} \leq 0, \quad s \in \{1, 2\} \quad (29)$$

and the participation constraint for working as

$$\sum_{s=1}^2 \varphi_s \int [v_s(x)] f_s(x) dx \equiv E [v_s(x)] \leq \alpha_2^{-1}. \quad (30)$$

There are two incentive-compatibility constraints, not one as in the previous section, because now there are two states, not one.

In the HMH model, we append (29) and (30) with two further constraints. Comparing the expected value to the agent from lying about the second state and working with the expected utility from reporting honestly in the second state and working, the principal can prevent the former by requiring contracts to satisfy

$$\int [v_2(x) - v_1(x)] f_2(x) dx \equiv E_2 [v_2(x) - v_1(x)] \leq 0. \quad (31)$$

An optimal contract also induces the agent 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

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

utility from reporting honestly and working, the sincerity condition reduces to

$$\int [v_2(x) - (\alpha_1/\alpha_2)v_1(x)g_2(x)]f_2(x)dx \equiv E_2[v_2(x) - (\alpha_1/\alpha_2)v_1(x)g_2(x)] \leq 0, \quad (32)$$

where $-\alpha_1v_1(x)$ 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.

Optimal contracting in the HMH model: Since $v_s(x)$ is monotone decreasing in $w_s(x)$, deriving $w_s(x)$ to minimize expected compensation for inducing work in both states subject to the five constraints is tantamount to choosing $v_s(x)$ for each (s, x) to maximize

$$\sum_{s=1}^2 \int \varphi_s \log[v_s(x)]f_s(x)dx \equiv E[\log v_s(x)] \quad (33)$$

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

$$\begin{aligned} & \sum_{s=1}^2 \varphi_s \int \left\{ \log[v_s(x)] + \eta_0 [\alpha_2^{-1} - v_s(x)] + \eta_s v_s(x) [(\alpha_1/\alpha_2)g_s(x) - 1] \right\} f_s(x) dx \\ & + \varphi_2 \int \left\{ \eta_3 [v_1(x) - v_2(x)] + \eta_4 [(\alpha_1/\alpha_2)v_1(x)g_2(x) - v_2(x)] \right\} f_2(x) dx \quad (34) \end{aligned}$$

with respect to $v_s(x)$, where η_0 through η_4 are the shadow values assigned to the linear constraints. Since each constraint is a convex set, their intersection is too. Also, $\log v$ is concave increasing in v , the expectation 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 PMH and HMH models are evident from (34). In the PMH 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

$$\begin{aligned} v_1(x)^{-1} &= \eta_0 + \eta_1 [(\alpha_2/\alpha_1) - g_1(x)] - \eta_3 h(x) - \eta_4 (\alpha_1/\alpha_2) g_2(x) h(x) \\ v_2(x)^{-1} &= \eta_0 + \eta_2 [(\alpha_2/\alpha_1) - g_2(x)] + \eta_3 + \eta_4. \end{aligned} \quad (35)$$

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 PMH 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 PMH model applies, and expected utility is equalized across

states. Otherwise $(\eta_3 + \eta_4)$ is strictly positive, implying that expected utility from the PMH model straddles the expected utility attained in the HMH 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 $s=2$ and penalized for $s=1$; in other words, the optimal contract pays him for luck.

There are three other contracts the principal might design. All three involve the agent shirking in at least one state. The cost-minimizing contract for shirking in both states is found by setting $\eta_1 = \eta_2 = 0$, and in both states the agent is paid $\gamma^{-1} \ln(\alpha_1)$. To make the agent work in the first state and shirk in the second, the principal sets $\eta_2 = 0$ in the cost-minimization problem. From the second part of Lemma 3.1, the agent receives a certain utility in the second state that exceeds his expected utility in the first state because to install incentives in the first state the agent must be rewarded to reveal when it does not occur. Finally, when the agent chooses work in the second state and shirking in the first, at least one of the multipliers, η_3 or η_4 , is strictly positive: from its first-order condition, $v_1(x)$ also depends on revenue, through $h(x)$ and possibly $g_2(x)$. Rather than load all the risk premiums into the second state, compensation in the first state optimally depends on revenue, not to induce work, but to induce truth telling and sincerity. The principal completes the optimization by comparing profits from each of the four contract types using the solutions to the respective cost-minimization problems.

3.2. Identification and empirical content in the HMH model

The observables in the econometric model of HMH are ex post measures of compensation, revenue and accounting reports, (\tilde{w}, x, r) . As in the PMH1 model, we assume that the ex post measure of compensation is measured with error but the measurement is independent of the all variables of interest, in this case x and r . The true compensation, w_r , is defined either by the optimal contract under work in equation (35), or the optimal shirking contract, $w_r = \gamma^{-1} \ln \alpha_1$. As is the case in the PMH1 model the effort level is not observed. There is another unobservable in the HMH model: the realization of the true state $s \in \{1, 2\}$. Our starting point is the premise that the equilibrium distribution of (W, X, R) is identified.¹⁸ The parameters of the HMH model are characterized by $f_s(x) \in F_s$, $g_s(x) \in G_s$, and $\varphi_s \in \Phi_s$ for $s \in \{1, 2\}$, which together defined the probability density functions for revenue in each state and the probability of each state occurring, $(\alpha_1, \alpha_2) \in \hat{A}$, the preference parameters for shirking and working (relative to the normalized utility from taking the outside option), and the risk-aversion parameter, $\gamma \in \Gamma_H$. Thus, the identification problem reduces to whether the structure $[F_s, G_s, \Phi_s, \hat{A}, \Gamma_H : s \in \{1, 2\}]$ can be recovered from knowledge of distribution of (W, X, R) . While the empirical content problem is whether any distribution of (W, X, R) can be rationalized by the structure $[F_s, G_s, \Phi_s, \hat{A}, \Gamma_H : s \in \{1, 2\}]$.

In equilibrium, agents truthfully reveal the state, implying $s=r(s)$, so $r=s$ in data generated by the HMH model. Consequently, φ_s is identified from the marginal distribution of R . As in the PMH1 model we focus on when work is demanded for each type of report. Therefore $f_s(x)$ is identified by the conditional distribution of X given $R=r(s)$, and similarly $w_s(x) = w_r(x)$ is identified from conditional expectation of W given X and R , *i.e.* $w_r(x) = E[W|X=x, R=r]$. The structure $[G_s, \hat{A}, \Gamma_H : s \in \{1, 2\}]$ is all that remains is to be identified from the distribution of (W, X, R) given that we have already identified $f_s(x)$, φ_s , and $w_s(x)$. We follow the same procedure as in the previous section.

18. Although (\tilde{w}, x, r) rather than (w, x, r) is observed, there is no loss in generality from assuming (w, x, r) is observed because $w_r(x) = E[\tilde{w}|X=x, R=r]$.

The most important differences between the analysis of the previous section and this one arise from the inequalities and equations that define equilibrium. These differences complicate the analysis of identification in the HMH model. Nevertheless, the main thrust of the results derived for the easier PMH1 model also holds for the HMH model. First, if γ , the risk parameter, is known, then the remaining parameters are non-parametrically point identified. Secondly, if γ is unknown then all the parameters are only set identified. Thirdly, we obtain sharp and tight bounds.

To set the stage for the theorem on tightness, we highlight the role of γ by defining $v_s(x, \gamma) \equiv e^{-\gamma w_s(x)}$ and $\bar{v}_s(\gamma) \equiv \sup_x [e^{-\gamma w_s(x)}]$. We also define the real valued mappings:

$$\hat{\alpha}_2(\gamma) \equiv \left[\int \sum_{s=1}^2 \varphi_s v_s(x, \gamma) f_s(x) dx \right]^{-1} \equiv E[v_s(x, \gamma)]^{-1} \tag{36}$$

$$\hat{\alpha}_1(\gamma) \equiv \hat{\alpha}_2(\gamma) \left\{ \frac{[\bar{v}_2(\gamma)]^{-1} - E_2[v_2(x, \gamma)^{-1}]}{[\bar{v}_2(\gamma)]^{-1} - E_2[v_2(x, \gamma)]^{-1}} \right\}. \tag{37}$$

For any given γ , we interpret $\hat{\alpha}_1(\gamma)$ and $\hat{\alpha}_2(\gamma)$ as taste parameters because one can show, by following the same arguments used to characterize $\alpha_1(\gamma)$ and $\alpha_2(\gamma)$ in the PMH1 case, that $0 < \hat{\alpha}_1(\gamma) < \hat{\alpha}_2(\gamma)$ for all $\gamma > 0$. Also let

$$g_2(x, \gamma) \equiv \frac{\bar{v}_2(\gamma)^{-1} - v_2(x, \gamma)^{-1}}{\bar{v}_2(\gamma)^{-1} - E_2[v_2(x, \gamma)^{-1}]} \tag{38}$$

As in the previous section, $g_2(x, \gamma)$ is positive with $E_2[g_2(x, \gamma)] = 1$ and so can be interpreted as a likelihood ratio function of x for all $\gamma > 0$. Finally, we sequentially define $g_1(x, \gamma)$ by first defining $\eta_4(\gamma)$, then $\eta_3(\gamma)$ and $\eta_1(\gamma)$ as

$$g_1(x, \gamma) \equiv \frac{\bar{v}_1(\gamma)^{-1} - v_1(x, \gamma)^{-1} + \eta_3(\gamma)[\bar{h} - h(x)] - \eta_4(\gamma)g_2(x, \gamma)h(x)\frac{\hat{\alpha}_1(\gamma)}{\hat{\alpha}_2(\gamma)}}{\eta_1(\gamma)}, \tag{39}$$

where:

$$\eta_4(\gamma) \equiv \frac{\frac{E_1[v_1(x, \gamma)]}{E[v_s(x, \gamma)]} - 1 - E_1[v_1(x, \gamma)h(x)] \left\{ E_2[v_2(x, \gamma)]^{-1} - E[v_s(x, \gamma)]^{-1} \right\}}{\frac{\hat{\alpha}_1(\gamma)}{\hat{\alpha}_2(\gamma)} E_1[v_1(x, \gamma)g_2(x, \gamma)h(x)] - E_1[v_1(x, \gamma)h(x)]} \tag{40}$$

$$\eta_3(\gamma) \equiv E_2[v_2(x, \gamma)]^{-1} - \eta_4(\gamma) - E[v_s(x, \gamma)]^{-1} \tag{41}$$

$$\eta_1(\gamma) \equiv \frac{\hat{\alpha}_1(\gamma)}{\hat{\alpha}_2(\gamma)} \left\{ \bar{v}_1(\gamma)^{-1} - E[v_s(x, \gamma)]^{-1} + \eta_3(\gamma)\bar{h} \right\}. \tag{42}$$

By inspection, all the mappings above can be computed as population moments given a value for γ .¹⁹ We do not claim that $g_1(x, \gamma)$ is a likelihood ratio for all $\gamma > 0$, nor that $\eta_i(\gamma) \geq 0$ for each $i \in \{1, 3, 4\}$ so we cannot necessarily interpret them as Kuhn Tucker multipliers for all $\gamma > 0$.

19. Note that $E[v_s(x, \gamma)]$, $E_s[v_s(x, \gamma)^{-1}]$, $E_s[v_s(x, \gamma)]$, $\bar{v}_s(\gamma)$, \bar{h} , $h(x)$, φ_s and $f_s(x)$ can be expressed as population moments of the data-generating process given γ .

Nevertheless, $g_1(x, \gamma^*)$ is a likelihood ratio, by Theorem 3.2 below. This theorem is the analogue to Theorem 2.1. It shows that if an HMH model with parameter γ^* generates the data, then the remaining parameters are point identified by $\widehat{\alpha}_1(\gamma^*)$, $\widehat{\alpha}_2(\gamma^*)$ and $g_s(x, \gamma^*)$. In other words if the risk parameter is known, the HMH model is also point identified without making any further parametric assumptions.

Theorem 3.2. *Suppose the distribution of (W, X, R) is generated by a parameterization of the HMH model with positive risk-aversion parameter γ^* . Then,*

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

Additional restrictions: The HMH model imposes truth-telling and sincerity constraints. Since these constraints help shape the optimal contract as a function of the parameters, they provide several restrictions on the population moments that do not hold in the PMH1 model. Define $\Psi_2(\gamma)$ through $\Psi_4(\gamma)$ as

$$\begin{aligned}\Psi_2(\gamma) &\equiv E_1[1\{g_1(x, \gamma)\} - 1] \\ \Psi_3(\gamma) &\equiv E_2[v_1(x, \gamma) - v_2(x, \gamma)] \\ \Psi_4(\gamma) &\equiv E_2[\widehat{\alpha}_1(\gamma)v_1(x, \gamma)g_2(x, \gamma) - \widehat{\alpha}_2(\gamma)v_2(x, \gamma)].\end{aligned}\quad (43)$$

The truth-telling constraint (31) implies $\Psi_3(\gamma^*) \geq 0$, while the sincerity constraint (32) implies $\Psi_4(\gamma^*) \geq 0$. The equality $\Psi_3(\gamma^*)\Psi_4(\gamma^*) = 0$ guarantees at least one of the constraints holds strictly. Since $g_1(x)$ is a likelihood ratio in the HMH model, we ensure $\widehat{g}_1(x, \gamma^*) \geq 0$ with unit mass by imposing the restriction that $\Psi_2(\gamma^*) \geq 0$. Three more inequalities ensure $\eta_1(\gamma^*)$, $\eta_3(\gamma^*)$ and $\eta_4(\gamma^*)$ are positive, a necessary condition for being Kuhn Tucker multipliers. Similarly, complementary-slackness conditions for truth telling and sincerity must be satisfied, meaning $\Psi_3(\gamma^*)\eta_3(\gamma^*) = 0$ and $\Psi_4(\gamma^*)\eta_4(\gamma^*) = 0$.

Another exclusion restriction imposed throughout is that α_1 does not depend on the state.²⁰ The HMH model yields this restriction from the same value of α_1 appearing in the incentive compatibility conditions for both states, define

$$\Psi_1(\gamma) \equiv \frac{1 - \frac{E_1[v_1(x, \gamma)]}{E[v_s(x, \gamma)]} + \eta_3(\gamma)E_1[h(x)v_1(x, \gamma)] + \eta_4(\gamma)\frac{\widehat{\alpha}_1(\gamma)}{\widehat{\alpha}_2(\gamma)}E_1[g_2(x, \gamma)h(x)v_1(x, \gamma)]}{E_1[v_1(x, \gamma)]}.\quad (44)$$

Theorem 3.2 shows that if an HMH model with parameter γ^* generates the data, then $\Psi_1(\gamma^*) = 0$.

Turning now to the effort level induced by the principal in the HMH model, we first remark that if shirking is demanded in both states, then compensation is determined in Lemma 2.1 for the PMH1 model. Since this is suboptimal,

$$\Lambda_1(\gamma) \equiv E[x - w_s(x)] - E[xg_s(x, \gamma)] + \gamma^{-1} \ln[\widehat{\alpha}_1(\gamma)]\quad (45)$$

is positive at γ^* . The principal induces work in both states when the expected profits from doing so are higher than when the agent shirks in one state; this remark yields two extra restrictions on

20. This exclusion restriction is a natural one to impose in our application, but is easy to relax. Noting (44) defines $\Psi_1(\gamma^*)$, and (48) defines the constraint set ΓH , we redefine ΓH by omitting the equation $\Psi_1(\gamma^*) = 0$ from ΓH .

γ^* to be utilized in identification. For any $\gamma \in R^+$, we denote by $w_s^{(s')}(x, \gamma)$ the cost-minimizing compensation in state s when the agent works in state $s' \in \{1, 2\}$ and shirks in the other state. Given the parameterization indexed by γ , the difference in value to the principal from demanding work in both states versus working in the first state only is

$$\Lambda_2(\gamma) = \varphi_1 E_1 \left[w_1^{(1)}(x, \gamma) - w_1(x) \right] + \varphi_2 E_2 \left\{ x - w_2(x) - g_2(x, \gamma) \left[x - w_2^{(1)}(x, \gamma) \right] \right\}. \quad (46)$$

At $\gamma = \gamma^*$, this expression is positive when it is optimal to require work in both states. In similar fashion, we define

$$\Lambda_3(\gamma) = \varphi_1 E_1 \left\{ x - w_1(x) - g_1(x, \gamma) \left[x - w_1^{(2)}(x, \gamma) \right] \right\} + \varphi_2 E_2 \left[w_2^{(2)}(x, \gamma) - w_2(x) \right] \quad (47)$$

and note that $\Lambda_3(\gamma^*) \geq 0$ for the same reason.

Sharp and tight bounds: Consolidating the restrictions directly applied to the HMH model, we define Γ_H , a Borel set of risk-aversion parameters, as

$$\Gamma_H \equiv \left\{ \gamma > 0 : \begin{array}{l} \Lambda_i(\gamma) \geq 0 \text{ for } i \in \{1, 2, 3\} \\ \eta_j(\gamma) \geq 0 \text{ for } j \in \{1, 3, 4\} \\ \Psi_1(\gamma) = 0 \text{ and } \Psi_k(\gamma) \geq 0 \text{ for } k \in \{3, 4\} \\ \Psi_3(\gamma)\Psi_4(\gamma) = \Psi_3(\gamma)\eta_3(\gamma) = \Psi_4(\gamma)\eta_4(\gamma) = 0 \end{array} \right\}. \quad (48)$$

By construction, Γ_H is tight because every observationally equivalent parameterization must satisfy its restrictions. Its restrictions consist of both inequalities and equalities; it is possible that Γ_H is a singleton but this cannot be guaranteed. Our last theorem establishes a result analogous to Theorem 2.2. We demonstrate that the tight bounds constructed for the HMH model also exclude every parameterization that cannot be rationalized by the data.

Theorem 3.2. Γ_H is sharp.

Nesting PMH within HMH: As the proof to the next corollary shows, when compensation in the HMH model depends on revenue x but not on the state s , it specializes to the PMH model. This nesting result immediately implies that the conditions for rejecting PMH1 model apply to this specialization of the HMH model as well. More generally, one can show by a continuity argument that the HMH model has empirical content.

Corollary 3.1. If $w_1(x) = w_2(x)$ then $\Gamma_H = \Gamma_1$. Hence there exist distributions of (W, X, R) such that Γ_H is empty.

A more interesting comparison is between the HMH model and a two-state analogue of the corresponding PMH model, which we abbreviate by PMH2; it is identical to HMH aside from its information structure, and compensation typically differs by state in both models. The PMH2 model assumes that both agent and principal observe $s \in \{1, 2\}$. We establish in the proof of Corollary 3.2 below that the observationally equivalent set of γ for the PMH2 model, denoted by Γ_2 , is essentially found by duplicating the elements used in deriving Γ_1 , and adding a condition to ensure expected utility is equated across states.²¹

21. It is possible for Γ_2 to be a singleton but this cannot be guaranteed.

There are three key differences in the optimal contracts of the PMH2 and HMH models, each of which is grounded in the information structure. While the outside option value equals the expected utility from working when averaged over both states in both models, under PMH2 the expected utility of the agent from working in each state is equated with the outside option value, whereas under HMH his expected utility from working is higher in the second state than the first. Since contracts are contingent on what the principal observes, differences in risk attitude lead the principal to insure the agent against the exogenous stochastic process determining the state. However, if the information is privy to the agent and is payoff relevant in guarding against moral hazard, the principal provides incentives for the agent to reveal it.

The HMH and PMH2 models are non-nested, because they impose different restrictions on the parameter space that arise from the contractual differences. By definition, the expected utility of an agent with risk parameter γ from working in state s is $-\alpha_2 E_2[e^{-\gamma w_s}]$, so the restriction that expected utilities are equated across states is simply

$$E_1[e^{-\gamma w_1}] = E_2[e^{-\gamma w_2}] = E[e^{-\gamma w_s}]. \quad (49)$$

This equality helps determine Γ_2 but not Γ_H . The other two differences can be traced back to the truth telling and sincerity constraints that apply in HMH but not PMH2. We have already showed that in HMH the truth telling constraint implies $E_2[e^{-\gamma w_1}] \geq E_2[e^{-\gamma w_2}]$ and the sincerity constraint implies $E_2[\hat{\alpha}_1(\gamma)g_2(x, \gamma)e^{-\gamma w_1}] \geq E_2[\hat{\alpha}_2(\gamma)e^{-\gamma w_2}]$ with at least one inequality holding with equality. These restrictions on the parameter space help determine Γ_H but not Γ_2 .

Corollary 3.2. $\Gamma_H \not\subseteq \Gamma_2 \not\subseteq \Gamma_H$.

To summarize our analysis of empirical content and nesting, Corollaries 2.1 and 3.1 justify testing each model individually, while Corollary 3.2 demonstrates that the nature and extent of agency cannot be refuted without taking the data to both models.

4. EXTENSIONS

For expositional ease, we have maintained the assumption up until now that the model is static: agents consume their compensation immediately, and the principal maximizes expected revenue net of compensation. Yet applied to the context of executive compensation, shareholders and managers solve dynamic problems: in reality, managers typically accumulate wealth during their highest earning years, and following the financial economics literature, it is more reasonable to assume that shareholders are forward looking and are diversified against idiosyncratic shocks that affect firm value, maximizing expected returns (net of managerial compensation). Another dimension along which our canonical model is clearly counterfactual is that it does not permit heterogeneity across principal–agent pairs. Yet firms differ in size and scope, and the data exhibit considerable heterogeneity across industrial sector and firm size within sector. For these two reasons, we extend our framework in order to account for heterogeneity and dynamic contracting. The extension builds on the dynamic PMH models analyzed in [Malcomson and Spinnewyn \(1988\)](#); [Fudenberg et al. \(1990\)](#); and [Rey and Salanie \(1990\)](#). This body of theoretical work has been applied in the empirical studies of executive compensation of [Margiotta and Miller \(2000\)](#); [Gayle and Miller \(2009a, 2009b\)](#); [Edmans et al. \(2012\)](#); and [Gayle et al. \(2014\)](#).

The identification and empirical content results in Sections 3 and 4 generalize to the dynamic analogue of the PMH and HMH models. The notable exception—apart from minor modifications to the formulae—is that the consumption smoothing problem gives rise to additional restrictions

that can help identify γ . These additional restrictions do not guarantee point identification. Below we outline how the PMH and HMH models are modified to account for dynamics and provide a brief overview of the identification and empirical content results. The details of these extensions and results are collected in the [Supplementary Appendix](#).

The extension of the framework to accommodate observed heterogeneity are straight forward and hence are only in the [Supplementary Appendix](#). The identification and empirical content results extend to environments where unobserved (to the researcher) heterogeneity helps determined the manager’s preferences. We consider the situation where the researcher has panel data that tracks agents for T periods, and the risk-aversion parameter, γ_n , is manager specific and time invariant. If T is large relative to N (the number of agents) then our identification and empirical-content results apply. See the [Supplementary Appendix](#) for more details.

4.1. A dynamic PMH model

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 $(\alpha_1, \alpha_2, \gamma)$ in the PMH model to depend on bond prices by setting

$$\alpha_1 \equiv \tilde{\alpha}_1^{1/(b_t-1)}, \alpha_2 \equiv \tilde{\alpha}_2^{1/(b_t-1)}, \gamma \equiv \tilde{\gamma}/b_{t+1}, \tag{50}$$

where $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\gamma})$ become the primitive preference parameters. The agent’s preferences become:

$$-l_{0t} - l_{1t} \tilde{\alpha}_1^{1/(b_t-1)} E \left[g(x) \exp \left(\frac{-\tilde{\gamma} w_{t+1}(x)}{b_{t+1}} \right) \right] - l_{2t} \tilde{\alpha}_2^{1/(b_t-1)} E \left[\exp \left(\frac{-\tilde{\gamma} w_{t+1}(x)}{b_{t+1}} \right) \right]. \tag{51}$$

Comparing equation (4) with (51), instead of agents with γ receiving $w(x)$, they have $\tilde{\gamma}$ and receive only the interest on the bonds purchased with the compensation, namely $w_{t+1}(x)/b_{t+1}$. Similarly instead of receiving the cash certainty equivalent of α_j , which is $\gamma^{-1} \ln \alpha_j$, the agent receives the one-period-deferred cash certainty equivalent of $\tilde{\alpha}_j$, which is $(b_{t+1}/b_t - 1) \ln \tilde{\alpha}_j$.²²

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

$$-E \left[\sum_{t=0}^{\infty} \beta^t (l_{0t} + l_{1t} \tilde{\alpha}_1 + l_{2t} \tilde{\alpha}_2) e^{-\tilde{\gamma} c_t} \right], \tag{52}$$

where $(l_{0t}, l_{1t}, l_{2t}, c_t)$ are the choice variables for each period t , and $\beta \in (0, 1)$ is the agent’s subjective discount factor. Mirroring the static model, $l_{0t} \in \{0, 1\}$ is an indicator variable for accepting an employment contract or taking the outside option in the t^{th} period, $l_{1t} \in \{0, 1\}$ indicates whether the agent shirks or not in that period, $l_{2t} \in \{0, 1\}$ indicates whether the agent works or shirks, 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 revenue the principal receives at the end of period t , $f(x_t)$ denote the density of revenue in period t under diligence, and $f(x_t)g(x_t)$ the revenue under shirking. Reinterpreted within this dynamic setting, [Margiotta and Miller \(2000\)](#) proved that the long-term contract for the PMH model in this dynamic framework decentralizes to a sequence of short-term contracts that mimic the contract described in Section 2.

22. 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.

4.2. A dynamic HMH model

This section develops the notation for a dynamic version of the HMH model, lays out the feasibility constraints for the optimization problem, and then shows that the optimal contract mimics the optimal contract for a static model under a simple parameter transformation.

Assumptions and notation: At the beginning of period t , the agent 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, denoted by c_t , and the principal proposes a new contract. The principal announces how the agent's 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 revenue 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 agent chooses whether to be engaged by the firm or not (either with another firm or in retirement). Denote this decision by the indicator $l_{t0} \in \{0, 1\}$, where $l_{t0} = 1$ if the agent chooses to be engaged outside the firm and $l_{t0} = 0$ if he chooses to be engaged inside the firm.

The agent then makes his unobserved labour 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. At the beginning of period $t+1$, revenue for the firm, x_{t+1} , is drawn from $f_s(x)$ if $l_{st2} = 1$ and $f_s(x)g_s(x)$ if $l_{st1} = 1$. We maintain the same assumptions on output and states, equation (27), and private information, equation (28), as in the static version of the HMH model.

The agent's wealth is endogenously determined by his consumption and compensation. We assume that 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 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. The lifetime utility is expressed as

$$-\sum_{s=1}^2 \sum_{t=0}^{\infty} \sum_{j=0}^J \beta^t \tilde{\alpha}_j l_{stj} \exp(-\tilde{\gamma} c_t), \quad (53)$$

where β 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 static models, we assume $\tilde{\alpha}_2 > \tilde{\alpha}_1$ and normalize $\tilde{\alpha}_0 = 1$.

The cornerstone for formulating the constraint that circumscribes the minimization problem the principal solves is the indirect utility function for a agent choosing between immediate retirement and retirement one period hence. Lemma 4.1 states this indirect utility function in terms of the utility received from retiring immediately. To state the lemma, let $r_t(s)$ denote the agent's disclosure rule about the state when the true state is $s \in \{1, 2\}$.

Lemma 4.1. *If the agent, offered a contract of $w_{rt}(x)$ for announcing r , retires in period t or $t+1$ 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

$$\sum_{s=1}^2 \varphi_s \left\{ \left(\frac{l_{st1}}{\tilde{\alpha}_1} + \frac{l_{st2}}{\tilde{\alpha}_2} \right)^{1/(b_t-1)} + E_s \left[\exp \left(-\frac{\tilde{\gamma} w_{r_t(s),t}(x)}{b_{t+1}} \right) [g_s(x) l_{t1} + l_{t2}] \right] \right\}. \quad (54)$$

Had he truthfully disclosed the true state s_t in period t , the agent would actually receive $w_{st}(x)$ as compensation if revenue x is realized at the end of the period. 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 $w_{st}(x)$:

$$v_{st}(x) \equiv \exp \left(-\frac{\tilde{\gamma} w_{st}(x)}{b_{t+1}} \right). \quad (55)$$

Substituting from (50) for α_1, α_2 , and γ into equations (29), (30), (31), and (32) gives the dynamic versions of the incentive-compatibility, participation, truth-telling, and sincerity constraints.

We first prove that the short-term optimal contract for the dynamic model has a static analogue. We 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 the expected compensation for inducing work in both states subject to the five constraints is equivalent to choosing $v_{st}(x)$ to maximize

$$\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s \ln [v_{st}(x)] f_s(x) dx \equiv E [\ln v_{st}(x)] \quad (56)$$

subject to the same four constraints. In this framework, there are no gains from a long-term arrangement between the principal and the agent. This claim is established by verifying that Fudenberg *et al.*'s (1990) assumptions are met, thus establishing that the long-term optimal contract decentralizes to a sequence of short-term contracts solved by the problem above. See the [Supplementary Appendix](#) for the proof of all these results pertaining to the dynamic HMH model.

4.3. Identification of the dynamic PMH and HMH models

Regarding identification PMH model, equation (50) treats bond prices as observed variables entering preferences restrictively, thus providing a further source of identification.²³ In the PMH model, we can substitute from (50) for α_1, α_2 , and γ into equations (13) (14), and (15), to prove that Theorem 2.1 implies

$$g(x, \tilde{\gamma}) \equiv \frac{e^{\tilde{\gamma}/b_{t+1} \bar{w}_{t+1}} - e^{\tilde{\gamma}/b_{t+1} w_{t+1}^o(x)}}{e^{\tilde{\gamma}/b_{t+1} \bar{w}_{t+1}} - E \left[e^{\tilde{\gamma}/b_{t+1} w_{t+1}^o(x)} \right]} \quad (57)$$

$$\tilde{\alpha}_1^{1/(b_t-1)} \equiv \frac{1 - E \left[e^{\tilde{\gamma}/b_{t+1} (w_{t+1}^o(x) - \bar{w}_{t+1})} \right]}{E \left[e^{-\tilde{\gamma}/b_{t+1} w_{t+1}^o(x)} \right] - e^{-\tilde{\gamma}/b_{t+1} \bar{w}_{t+1}}} \quad (58)$$

23. While structures in PMH and HMH remain $[F, G, A, \Gamma]$ and $[F_s, G_s, \Phi_s, \hat{A}, \Gamma H; s \in \{1, 2\}]$ the observable distributions now relate to the random variables (W, X, B) and (W, X, R, B) in the PMH and HMH models respectively.

$$\tilde{\alpha}_2^{1/(b_t-1)} \equiv \left\{ E \left[e^{-\tilde{\gamma}/b_{t+1}w_{t+1}^o(x)} \right] \right\}^{-1} \quad (59)$$

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 (58) and (59), and then first differencing over t yields $T - 1$ further restrictions that aid identification of $\tilde{\gamma}$, and hence the other structural parameters. These additional restrictions are equalities, however, because they are non-linear in $\tilde{\gamma}$, they do not imply point identification. Similar procedure is followed in the dynamic HMH model by substituting from (50) for α_1 , α_2 , and γ into the equations in Theorem 3.1 implies that dynamic HMH models is identified if $\tilde{\gamma}$ is known. While by substituting from (50) for α_1 , α_2 , and γ into restrictions defined in equation (48) implies $\tilde{\gamma}$ is set identified. These restrictions are all presented in more details in the [supplementary appendix](#).

5. AN EMPIRICAL APPLICATION

This section outlines a general approach to estimation, testing and inference based on our identification analysis, and illustrates the approach within the context of executive compensation. The application tests PMH and HMH models of executive compensation and estimates the importance of asymmetric information in the variations not rejected by the data.

5.1. Motivation and data

Top executives in publicly traded companies are paid mostly with firm-denominated securities, but their total compensation also varies positively with accounting benchmarks. The PMH model is the standard paradigm for explaining why executives are paid in firm-denominated securities instead of a fixed wage. Yet in PMH1 α_1 , α_2 , $f(x)$, and $g(x)$ do not depend on unobserved states, and compensation does not depend on accounting reports.

The PMH2 model discussed in Section 3.2 can simultaneously justify why compensation consists mostly of firm denominated securities and depends on accounting benchmark if accounting reports are verifiable. Compensation is a function of both financial returns and accounting information: $f_s(x)$, and $g_s(x)$ are indexed by the states, but α_1 and α_2 are not. However in PMH2 accounting information is not subject to manipulation, in seeming contradiction to the widespread belief that managers have some discretion in setting accounting benchmarks and manipulating them. To address this limitation, several theoretical papers use the HMH model to study executives compensation when accounting reports can be manipulated.²⁴ In these HMH models, the principal designs an optimal contract, not only to provide the agent with an incentive to take actions that enhance profitability, but also to minimize the agent's incentive to manipulate accounting benchmarks. The structure of the primitives in the HMH model is the same as PMH2; only the contract space is smaller because of the additional constraints imposed on the principal.

A parallel body of theory treats managerial manipulation of accounting information as another hidden action within a PMH model.²⁵ The information structure is identical to the HMH model, but the agent has an unobserved manipulation effort choice, as well as an unobserved value-enhancing effort choice. Since some states are easier to hide than others the preference parameters of this model also depend on the state. We denote a PMH model with state-dependent preferences by PMH3: letting α_{js} denote the taste parameter associated with choosing action j in state s ,

24. See for example [Lacker and Weinberg \(1989\)](#), [Crocker and Morgan \(1998\)](#), [Crocker and Slemrod \(2007\)](#), [Maggi and Rodriguez-Clare \(1995\)](#) and [Kartik \(2009\)](#).

25. See for example [Peng and Roell \(2008\)](#) and [Benmelech et al. \(2010\)](#).

the parameters α_{1s} and α_{2s} are functions of accounting reports where presumably $\alpha_{j2} > \alpha_{j1}$ for $j \in \{1, 2\}$ because it is more costly to manipulate accounting reports when $s=2$. In the PMH3 model $f_s(x)$, and $g_s(x)$ are also functions of accounting reports; accounting reports are only partially reflected in stock returns because investors are unable to back out the true state due to uncertainty caused by executives manipulating the accounting reports. This reduces the precision of stock returns as a signal of managerial effort.

Using reduced form regression techniques we can reject the PHM1 model, since compensation should be independent of accounting reports, but the PHM2, PMH3, and HMM models cannot be rejected that easily, because accounting information is a valid regressor in the compensation equation. To test, and distinguish between, the PMH2, PMH3, and HMM models, structural econometrics is required; a byproduct of this approach is that we can conduct a welfare analysis with the unrejected models.

There are three main ways to measure the welfare cost of asymmetric information in managerial compensation settings: the expected gross output loss to the firm for switching from the distribution of abnormal returns for working to the distribution for shirking, denoted Δ_1 ; the non-pecuniary benefits to the manager from shirking, denoted Δ_2 ; and the firm's willingness to pay to eliminate the agency problem, denoted Δ_3 .²⁶ Let V denote the value of the firm at the beginning of the period, and let x denote the firm's gross excess return realized at the end of the period. Δ_1 is defined as

$$\Delta_1 = E[x|\text{manager works}]V - E[x|\text{manager shirks}]V. \quad (60)$$

Let $w^{(2)}$ denote the manager's reservation compensation to work under perfect monitoring or if there were no moral hazard problem, and let $w^{(1)}$ denote the manager's reservation compensation to shirk. Then Δ_2 , the compensating differential for these two activities, can be expressed as the difference:

$$\Delta_2 = w^{(2)} - w^{(1)}. \quad (61)$$

If managerial effort were observed by the shareholders, then the firm would pay the manager a fixed salary of $w^{(2)}$. However, if managerial effort is not observed by the shareholders and shareholders want the manager to work, then the manager is paid according to the optimal compensation schedule $w(x)$. Thus, Δ_3 is defined as

$$\Delta_3 = E[w(x)] - w^{(2)}. \quad (62)$$

Both of our models share the prediction with the literature on this topic that a manager would be paid a fixed salary if he shirks, that x is drawn from the probability distribution conditional on him working, and that asymmetric information explains why managerial compensation varies with abnormal firm returns, an almost universal finding of a large empirical literature subject only to the caveat that all the components comprising CEO compensation be included in the definition.²⁷ In equilibrium $E[x|\text{manager works}]$ can be estimated, but not $E[x|\text{manager shirks}]$, so with reference to (60), the data does not yield a direct estimate of Δ_1 . Similarly, while average compensation gives an unbiased and consistent estimator of $E[w(x)]$, its certainty equivalent, $w^{(2)}$, cannot be estimated without a measure of risk aversion, one of the model's primitives.

26. These three measures directly address the six questions posed in Abowd and Kaplan's (1999) survey, and are estimated in parametric PMH models by Margiotta and Miller (2000) and Gayle and Miller (2009b).

27. See for example Hall and Liebman (1998), Margiotta and Miller (2000), and Gayle and Miller (2009b) who followed Antle and Smith's (1985) outline of the key components of managerial compensation.

TABLE 1
Cross-sectional summary of firm characteristics by sector

Variable	Primary	Consumer	Service
Observations	8980	6762	11,144
Assets (millions of 2000 US\$)	6322 (27,773)	5277 (22,124)	17,776 (67,133)
Market Value (millions of 2000 US\$)	6480 (25,160)	7811 (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)

Note: Standard deviations are in parentheses.

Likewise $w^{(1)}$ depends on the non-pecuniary benefits from shirking, another model primitive. To summarize: $E[x|\text{manager shirks}]$, $w^{(2)}$ and $w^{(1)}$ are counterfactual objects used as inputs to define the welfare measures Δ_1 through Δ_3 , which cannot be freely estimated from the data on compensation, abnormal return, and firms' balance sheet items without imposing structure. Details for computing Δ_1 , Δ_2 , and Δ_3 as mappings of γ for the different models are in the [supplementary appendix](#).

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 CRSP database. The sample was partitioned into three industrial sectors by GICS code. Sector 1, called primary, includes firms in energy, materials, industrials, and utilities. Sector 2, consumer, comprises firms from consumer discretionary and consumer staples. Firms in health care, financial services, information technology and telecommunication services comprise Sector 3, which we call service.

Table 1 summarizes the cross-sectional features of our data. Average firm size by total assets is highest in the service 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. This ranking is reversed when employment is used to measure firm size instead. For this reason, we used total assets and employment as our 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.

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. In the sample period, financial returns from the stock market to diversified shareholders ranged from a yield of 45% in one year to a loss of 14% 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 their actions, financial returns to the firm net of the share market index return. Average accounting returns is highly correlated with financial returns, almost without exception rising and falling together. The term structure of interest rates underlying the bond price series was constructed from data on Treasury bills of varying maturities, and the prices were derived using methods described in [Gayle and Miller \(2009b\)](#).

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	8896 (26,269)	18.02 (46.15)	2.83 (7.24)	1.19 (0.45)	1.18 (0.51)	1854 (12,412)	1574
1994	13.72	7770 (25,284)	16.18 (43.41)	2.87 (5.04)	0.97 (0.29)	1.07 (2.52)	2714 (10,909)	1876
1995	14.00	8187 (28,650)	16.43 (44.41)	3.45 (33.40)	1.26 (0.47)	1.18 (0.64)	1781 (13,252)	1867
1996	13.79	8357 (29,029)	17.31 (45.92)	2.41 (17.20)	1.16 (0.38)	1.17 (0.87)	3257 (14,824)	1926
1997	13.67	8770 (31,797)	17.94 (47.96)	2.76 (41.40)	1.30 (0.48)	1.22 (3.06)	4691 (17,791)	1997
1998	15.00	9486 (40,145)	17.67 (45.91)	3.91 (71.30)	1.05 (0.53)	1.20 (1.11)	2726 (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)	4624 (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)	3314 (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)	3165 (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)	3151 (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)	4069 (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)	4397 (19,992)	751

Note: Standard deviations are in parentheses. 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.

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, total employment, and debt–equity ratio. 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 was less than or greater than median employment for firms in the sector, and whether its debt–equity ratio was less than or greater than the median debt–equity ratio for firms in the sector. Therefore firm type is measured by the triplicate (A,W,D), where A is assets, W is the number of workers, and D is the debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its median of the industry. For example (S,S,L) denote lower total assets and employment than the median firm in the sector, but a higher debt–equity ratio than the median debt–equity ratio for firms in the sector. Similarly (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 reported. They have many ways of directly affecting the firm’s balance sheets, choosing, for example, among different valuation methods for credits and liabilities and using discretionary timing when writing off non-performing 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

TABLE 3
Estimates of the probability distribution of accounting reports

Firm type (A,W,D)	Primary			Consumer			Service			Total
	N	Good	Bad	N	Good	Bad	N	Good	Bad	
(S,S,S)	2598	0.0917	0.1975	2023	0.1227	0.1764	3483	0.1249	0.1877	8103
(S,L,S)	319	0.0141	0.0214	268	0.0121	0.0275	210	0.0040	0.0149	797
(S,L,L)	469	0.0257	0.0266	418	0.0229	0.0389	1210	0.0337	0.0749	2097
(S,S,L)	1326	0.0763	0.0713	961	0.0725	0.0696	952	0.0434	0.0421	3239
(L,S,S)	541	0.0272	0.0331	498	0.0308	0.0427	760	0.0248	0.0434	1799
(L,L,S)	1105	0.0635	0.0595	734	0.0593	0.0493	927	0.0164	0.0668	2766
(L,L,L)	2398	0.1118	0.1552	1686	0.0879	0.1614	3056	0.0865	0.1878	7140
(L,S,L)	224	0.0127	0.0123	175	0.0145	0.0114	546	0.0262	0.0227	945
Total	8980	0.423	0.577	6762	0.423	0.577	11,144	0.360	0.640	26,886

Note: A report is classified as ‘Good’ if the firm’s accounting is higher the expected value of accounting return—the yearly sample average for a firm type and sector—and ‘Bad’ otherwise. N is the number of observations. Firm type is measured by the triplicate (A,W,D), where A is assets, W is the number of workers, and D is the debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry median.

net assets as reported in the annual report of the n^{th} firm up to the end of period t . Normalizing comprehensive income, we define the accounting return π_{nt} for the firm in period t as

$$\pi_{nt} = (A_{nt} - Debt_{nt} + Dividend_{nt}) / (A_{n,t-1} - Debt_{n,t-1}). \quad (63)$$

These variables are used to form our measure of the indicator variable $r_{nt} \in \{1, 2\}$. For a given firm type, let $E[\pi_{nt}]$ denote the expected accounting return of π_{nt} for firm n at the beginning of period t before the manager announces total assets A_{nt} . In our empirical analysis, we set $r_{nt} = 2$ if the firm’s accounting return is higher than the expected value of accounting returns in period t , and $r_{nt} = 1$ if lower. The expected value of accounting returns is calculated as the yearly sample average for firm type and sector.

Table 3 displays the number of observations 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. Table 4 provides a cross-sectional summary of the average abnormal returns and compensation 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.

5.2. Estimation, testing and inference

The static PMH3 model where both taste parameters are state dependent is a convenient prototype for explaining the empirical methods applied in this article. Appealing to the results in Section 3, we can easily prove using (25) that the identified set for this model is

$$\Gamma_3 \equiv \left\{ \gamma > 0 : \sum_{r=1}^2 \min\{0, Q_r(\gamma)\}^2 = 0 \right\}, \quad (64)$$

where

$$Q_r(\gamma) \equiv \frac{\text{cov}(x, e^{\gamma w_r(x)} | r)}{e^{\gamma \bar{w}_r} - E_r[e^{\gamma w_r(x)}]} - E_r[w_r(x)] + \gamma^{-1} \log \left(\frac{e^{\gamma \bar{w}_r} - E_r[e^{\gamma w_r(x)}]}{e^{\gamma \bar{w}_r} E_r[e^{-\gamma w_r(x)}] - 1} \right). \quad (65)$$

TABLE 4
Cross-sectional summary of returns and compensation

Firm type (A,W,D)	Primary		Consumer		Service	
	Good	Bad	Good	Bad	Good	Bad
Abnormal return						
(S,S,S)	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)	-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)	-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)	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)	-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)	-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)	-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)	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)	3889 (14,651)	670 (10,779)	3397 (19,178)	-1501 (15,235)	6063 (20,034)	1701 (17,316)
(S,L,S)	4384 (9,381)	2339 (14,243)	4922 (30,677)	-486 (23,882)	8015 (24,615)	-1183 (25,740)
(S,L,L)	3742 (11,903)	521 (15,710)	9194 (19,898)	821 (11,820)	7096 (14,740)	2274 (14,363)
(S,S,L)	2522 (9,855)	721 (8,851)	3977 (14,844)	908 (11,504)	4154 (16,068)	-150 (14,255)
(L,S,S)	3079 (20,381)	-850 (15,773)	4235 (20,107)	-510 (16,940)	3386 (18,844)	1629 (19,287)
(L,L,S)	4154 (13,375)	2422 (16,220)	4727 (20,989)	-429 (21,784)	8035 (24,244)	5496 (26,472)
(L,L,L)	5781 (12,807)	2200 (12,208)	6897 (19,288)	2775 (19,118)	9846 (24,075)	5595 (19,936)
(L,S,L)	4396 (14,831)	-3729 (18,890)	4742 (19,288)	-2442 (14,448)	5647 (20,347)	1718 (17,612)

Note: A report is classified as ‘Good’ if the firm’s accounting is higher the expected value of accounting return—the yearly sample average for a firm type and sector—and ‘Bad’ otherwise. Abnormal return is firm’s stock returns less the return on the market portfolio. Compensation in thousands of 2000 US\$. Firm type is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. Standard deviations are in parentheses.

To estimate $Q_r(\gamma)$ from a cross-section of N observations on (r_n, x_n, w_n) for otherwise identical firms and their managers, we replace \bar{w}_r with its super-consistent estimate

$$\hat{w}_r \equiv \max_{n \in \{1, \dots, N\}} \{w_n : r_n = r\} \quad (66)$$

and $w_r(x)$ and $f_r(x)$ with their non-parametric estimates and substitute sample moments for the expectation terms in (65) as a function of γ , to form a sample analogue denoted by $Q_r^{(N)}(\gamma)$. Given appropriate regularity conditions, the law of large numbers implies $Q_r^{(N)}(\gamma)$ converges to

$Q_r(\gamma)$, and we denote its rate of convergence by N^a . Our tests are based on the fact that if $\gamma \in \Gamma_3$ then sampling error is the only explanation for why $Q_r^{(N)}(\gamma)$ might be negative. Let $\Gamma_{3\delta}^{(N)}$ denote the set of risk-aversion parameters that asymptotically cover the observationally equivalent set of $\gamma > 0$ with probability $1 - \delta$. Let $c_\delta^{(N)}$ denote a consistent estimator for c_δ , the critical value associated with test size δ , and define $\Gamma_{3\delta}^{(N)}$ as

$$\Gamma_{3\delta}^{(N)} \equiv \left\{ \gamma > 0 : \sum_{r=1}^2 \min \left\{ 0, N^a Q_r^{(N)}(\gamma) \right\}^2 \leq c_\delta^{(N)} \right\}.$$

Thus $\Gamma_{3\delta}^{(N)}$ is a consistent estimator of the identified set Γ_3 . Intuitively, if $N^a Q_r^{(N)}(\gamma)$ 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. If $N^a Q_r^{(N)}(\gamma^*)$ is small in absolute value, or positive, we do not reject the null hypothesis that γ^* belongs to the identified set. In practice we reject the specification if the estimated confidence interval, $\Gamma_{3\delta}^{(N)}$, is empty. There are several methods for obtaining $c_\delta^{(N)}$; we modify a subsampling procedure of Chernozhukov *et al.* (2007).²⁸ Several of the components to the test statistic are ill defined when $v_{rr}(x, 0) = 1$ for all x , therefore our modification of the subsampling procedure bounds the set of γ considered away from zero; the subsampling procedures used in our empirical application are described in the Appendix.

The estimation and testing procedures for the models we actually take to data are more involved than those for the prototype in four respects. The prototype is a static model, but we implemented dynamic versions by including the bond price variables in the estimation equations wherever appropriate, as indicated by equation (50). Secondly, the prototype only sums over two quadratic terms, the states, whereas the inequalities and equalities defining the HMH model in the definition of Γ_H , given in (48) yield eleven such quadratic expressions to be summed. Computation of these expressions require additional estimates of $h(x)$, \bar{h} and φ_r . The next issue arises from observed heterogeneity included in the estimated models but not in the prototype. If the model was saturated, we could separately estimate it for each industry, firm size and calendar time period: however we assume preferences are stable over time, an assumption that generates additional restrictions of the form (58), (59) and (57) which must hold for each time period sampled in the panel; we also assume the risk-aversion parameter is homogeneous, so our estimator requires all the restrictions to be satisfied by the same value of γ across all industries, firm sizes and time periods. Finally, we assume w_n is measured with error, independent and identically distributed about its true value, implying \hat{w}_r is biased upwards in this setting. The treatment of these issues is extensively discussed in the [supplementary appendix](#).

5.3. Empirical results

This section presents our empirical results from our structural model. First, we explain which models were rejected by the data and which ones were not. Then we compare the estimates of γ from the nonrejected models to those found in the literature. Finally, we report the estimates of the welfare measures Δ_1 through Δ_3 obtained for the unrejected models.

28. This is called the moment selection t -test (MMM). See Andrews and Soares (2010) for a discussion of this class of critical-value functions. An alternative statistic is the quaslikelihood ratio (QLR) statistic, defined as $T_{\text{QLR}}^{(N)}(\gamma) = \inf_{t \in \mathbb{R}_+, \infty} (N^{1/2} Q^{(N)}(\gamma) - t)^2$. See Pakes *et al.* (2006); Chernozhukov *et al.* (2007); and Romano and Shaikh (2010) for studies using MMM. See Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), and Andrews and Barwick (2012) for studies using QLR.

TABLE 5
The 95% confidence region of risk-aversion for the PMH3 model

Firm type (A, W, D)	Sector		
	Primary	Consumer	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	7796	5600	8536

Note: The subsampling procedure was performed using 100 replications of subsamples with 3000 observations each using a grid of 1000 equally spaced points on the interval [9.112E-04, 50]. Firm type is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. All specifications automatically imposed cost minimization.

Specification tests: First we report the least restricted version of the PMH3 model. In Table 5, γ is allowed to vary over firm type and sector, but not over time or by accounting return; α_1 and α_2 are permitted to vary with firm type and sector, by accounting return and also across periods in an unrestricted way.²⁹ We do not reject the PMH3 model at the 5% level. The intersection of the estimated intervals is (0.02, 0.21), non-empty. Therefore, we cannot reject the hypothesis that a common γ applies to all firm types within all sectors.³⁰

Table 6 reports the results from imposing additional exclusion restrictions that eliminate the dependence of the taste parameters α_1 and α_2 on accounting returns. The left panel of Table 6 shows what happens to the estimated identified set of γ when we impose the restriction that α_1 is not affected by accounting returns. Providing γ is permitted to vary with firm and sector type, this model is not rejected. However, there is no common region of overlap for γ across all 24 firm and sector types, and the model is rejected when we impose the further restriction that γ is common across firm and sector types. The right panel presents our results from imposing the restriction that α_2 is equal across the two accounting return states. Here again, we cannot reject the model if γ is allowed to vary across firm and sector type, but it is rejected if we maintain the assumption that γ does not depend on the type of firm or sector managers select into.

Saturating the risk-aversion parameter by allowing γ to depend on each of the 24 firm and sector combinations, does not give the PMH2 model enough flexibility to accept the hypothesis that α_1 and α_2 are independent of accounting returns because the estimated identified set of γ is empty for three of the firm- and sector-specific cells.³¹ Inspecting the columns for the primary and service sectors in the left panel, we reject the hypothesis that a sector-specific γ can reconcile the data when we impose the additional restriction that the common α_1 does not depend on accounting returns. Similarly, there is no overlap in each the right three columns; we reject the model that a working parameter does not depend on accounting returns for a sector-specific γ . Performing

29. Baker and Hall (2004) discuss why the risk-aversion parameter may vary with firm size. Oyer (2004) explains why the cost of effort and the outside option may vary with risk factors beyond managers' control.

30. Moreover, we find that attributing all time variation in α_1 and α_2 to the dynamics of the expanded model, as reflected in bond price movements, does not shrink any of the 24 size- and type-specific confidence regions for γ . Thus, we find no evidence in the dynamic model that either tastes or the value of the outside option shifted within the time frame of our panel.

31. The three firm–sector combinations in which γ 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.

TABLE 6

The 95% confidence region of risk-aversion for the PMH2 model (exclusion restriction of independence of the cost of effort and accounting return)

Firm type (A,W,D)	Cost of shirking $\alpha_{11t} = \alpha_{12t}$ Sector			Cost of working $\alpha_{21t} = \alpha_{22t}$ Sector		
	Primary	Consumer	Service	Primary	Consumer	Service
(S,S,S)	(0.4, 0.5) \cup (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) \cup (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	7796	5600	8536	7796	5600	8536

Note: The subsampling procedure was performed using 100 replications of subsamples with 3000 observations each using a grid of 1000 equally spaced points on the interval [9.112E-04, 50]. Firm type is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (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 moral hazard based on only cost minimization and profit maximization.

TABLE 7

95% confidence region of risk-aversion for the HMH model

Sector	Observations	Confidence region
Primary	7796	(0.002, 0.26) \cup (0.37, 0.42)
Consumer	5600	(0.002, 0.13) \cup (0.19, 0.57)
Service	8536	(0.27, 0.53)

Note: The subsampling procedure was performed using 100 replications of subsamples with 3000 observations each using a grid of 1000 equally spaced points on the interval [9.112E-04, 50]. The restrictions imposed are profit maximization and equalization of preference parameters across size, leverage, and time.

a similar exercise on each row reveals that for a given firm type, the model rejects a common γ across sectors if we impose the additional joint restrictions that α_1 and α_2 are common within a sector. In other words, imposing restrictions on α_1 and α_2 across sectors or across firm types is inconsistent with a common γ in the selected types. Summarizing Table 6, the PMH2 model is rejected.

Table 7 presents the 95% confidence interval of the γ identified set for a restricted HMH model. Here we assume that α_1 and α_2 do not depend on accounting return or calendar time (but only firm sector and type), and we also impose a common γ across firm and sector type. We cannot reject this model at the 5% confidence level in any sector.

Risk aversion estimates: The PMH3 and HMH models are not rejected. 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 PMH3 given in Table 5, is less than the lower bound of (0.37, 0.42), the intersection of the sector types for the HMH 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 \$8849 and \$92390 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. The risk-aversion parameter set for the HMH model are quite close to those found for PMH model by Margiotta and Miller (2000), Gayle and Miller (2009) and Gayle *et al.* (2014), although these papers use different estimation methods and data from industrial sectors, executive ranks, and time periods.³² Thus if accounting reports can be manipulated as in a HMH model, or if accounting reports are 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 reports are manipulated as in the PMH3 model the estimated risk-aversion parameters are lower than those obtained in Margiotta and Miller (2000), Gayle and Miller (2009) and Gayle *et al.* (2014).

Gross loss from shirking (Δ_1): The top panel of Table 8 presents our estimated set of gross losses for both the PMH3 and the HMH models. The differences between the two model specifications are relatively small when compared to variation over firm type. For example, the median minimum distance between the confidence intervals for the PMH3 and HMH models is 0.32% in the primary sector, 1.74% in the service sector, and 1.49% in the consumer sector. By way of comparison, the variation of the confidence interval across firm type is several orders of magnitude larger; for the PMH3 it ranges between 7.84% and 14.89%, and for the HMH between 17.35% and 24.57%, depending on the sector. Since the average stock market return over this period was roughly 10% per annum, the expected gross return would have been negative for more than half the firm and sector types in both specifications if shareholders had ignored the moral-hazard problem.

To get a rough sense of the annual dollar losses implied by Δ_1 within each sector, we averaged the bounds of the confidence intervals over firm type within sectors, and then multiplied the average bounds by the average market value given in Table 2. In the PMH3 model, the estimated average annual loss to firms varies between \$545 million and \$601 million in the primary sector, \$918 million and \$1.00 billion in the consumer sector, and \$1.46 billion to \$1.66 billion in the service sector. Similarly, the loss varies from \$580 million to \$648 million in the primary sector, \$1.00 billion to \$1.07 billion in the consumer sector, and \$1.42 billion to \$1.49 billion in the service sector per year in the HMH model.

Certainty equivalent wages (Δ_2): The manager's compensating differential from shirking versus working, Δ_2 , is the difference between $w^{(2)}$ and $w^{(1)}$. Table 9 presents the estimated identified set of the manager's reservation wages, for shirking and working, for both the PMH3 model and the HMH model.³³ Because the unrestricted model has a reservation compensation for each accounting report r , but α_1 and α_2 do not vary by r in the restricted model, there are twice as many regions for the PMH3 model to report as for the HMH model.

The top panel of Table 9 presents the estimated identified set of $w^{(1)}$ for both models. In the PMH3 model, $w^{(1)}$ is always higher in the good state than in the bad. In 18 out of 24 firm and sector types the hybrid $w^{(1)}$ lies between the two estimates of $w^{(1)}$ for the PMH3; in the remaining, six the HMH $w^{(1)}$ is below the region for the $w^{(1)}$ in the bad state of the PMH3.³⁴ In the PMH3

32. Margiotta and Miller (2000) use data 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 *et al.* (2014) covers the estimated identified set for the HMH model, yet they estimate a dynamic moral hazard model with human capital accumulation and sorting over firm and ranks.

33. To estimate the confidence regions for Δ_2 and Δ_3 we set the bond price at the median for the sample period.

34. Three of the six exceptions occur in the (S,S,S) firm type; that is, for all three sectors.

TABLE 8

Confidence regions for the agency costs (gross losses are measured as percentages of assets and the risk premium is in millions of 2000 US\$)

Firm type (A,W,D)	Primary		Consumer		Service	
	PMH3	HMH	PMH3	HMH	PMH3	HMH
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)

Note: Firm size and leverage is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. The numbers of observations are 7796, 5600 and 8536 in the primary, consumer, and service sectors, respectively. All calculations in this table were performed using the median bond price in the data. The risk-aversion parameters used are the interaction across sector, size and leverage of the most parsimonious unrejected specification—i.e. $\gamma \in (0.02, 0.21)$ for the pure moral hazard model and $\gamma \in (0.37, 0.42)$ for the hybrid moral hazard model.

models, $w^{(1)}$ is negative³⁵ 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 of the estimates of $w^{(1)}$ in the HMH model are positive; because the estimated regions typically lie between the PMH3 $w^{(1)}$ for the two states, the HMH magnitudes are lower.

The bottom panel of Table 9 presents the identified set of $w^{(2)}$. In the PMH3 model, $w^{(2)}$ is negative for 9 out of 24 firm types in the bad state; in those cases, the manager would be willing to pay for employment when the firm reports it is doing poorly relative to the industry! This striking finding is not surprising. As shown in Table 4, the same nine firm and sector types have negative average 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. Turning to the HMH model, although 4 out of 24 of the firm types have a negative lower bound, the confidence region under the HMH model always has an interval containing positive numbers. In words, managers are always paid compensation with a positive certainty equivalent in the HMH model, but the parameter estimates of the PMH3 model imply that managers are paid compensation that has a negative $w^{(2)}$ in the bad state, apparently because they enjoy the challenge of working for firms that are poorly positioned within the industry. We view this finding as evidence favoring the restricted HMH model over the unrestricted PMH3 model.

35. A negative w_2 or w_1 means that in equilibrium the manager would pay shareholders for the privilege of holding the job.

TABLE 9
Confidence regions of certainty-equivalent wage for shirking and working (measured in millions of 2000 US\$)

Firm type (A,W,D)	Primary			Consumer			Service		
	PMH3		HMH	PHM3		HMH	PHM3		HMH
	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)

Note: A report is classified as 'Good' if the firm's accounting is higher the expected value of accounting return—the yearly sample average for a firm type and sector— and 'Bad' otherwise. Firm type is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt-equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. The numbers of observations are 7796, 5600 and 8536 in the primary, consumer, and service sectors, respectively. The shirking wage is $w_i^{(1)} = [\gamma(b_i - 1)]^{-1} b_{i+1} \log(\alpha_1)$ and the working wage under perfect monitoring is $w_i^{(2)} = [\gamma(b_i - 1)]^{-1} b_{i+1} \log(\alpha_2)$. All calculations in this table were performed using the median bond pure moral hazard model and $\gamma \in (0.37, 0.42)$ for the hybrid moral hazard model.

Willingness to pay to eliminate the agency cost (Δ_3): The bottom panel of Table 8 presents our estimates of the identified set of Δ_3 . We find that Δ_3 is higher in the HMH model than in the PMH3 model for every firm type, by several hundred thousand dollars. Despite the quantitative differences between the PMH and the HMH models, most of the qualitative comparisons between firm types and industry sectors match up. For both the HMH and the PMH3 specifications, after conditioning on firm type, Δ_3 is lower in the primary than the consumer sector with just one exception, (L,S,L). In both specifications, Δ_3 for the consumer sector is generally lower than for the service sectors. Controlling for assets and employment, all firm types with a higher debt–equity have a lower Δ_3 than their counterparts with a lower debt–equity ratio in both the pure and the 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 Δ_3 , given total assets and the debt–equity category. Not only does this hold for both the PMH3 and HMH specifications; 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 firm assets and Δ_3 is somewhat weaker, but generally speaking, higher firm assets are associated with a higher Δ_3 .

6. CONCLUSION

We develop semiparametric methods for identifying, estimating, and testing principal-agent models and derive the equilibrium restrictions from optimal contracting to predict the shape of the compensation schedule, when there are only hidden actions and when there is hidden information as well. These restrictions fully characterize the empirical content of our models. We show that models of moral hazard studied in this article are only partially identified because every risk-aversion parameter that satisfies the inequality derived from profit maximization generates an observationally equivalent model. We then establish sharp and tight bounds for the risk-aversion parameter, and show that all the other parameters, including the probability density function of revenue from an activity that is deterred in equilibrium and consequently never sampled from, can be expressed as mappings of the risk-aversion parameter and probability distribution of the data-generating process. The null hypothesis, that the data are generated by any one of the risk-aversion parameters in the identified set, is rejected if the estimated confidence interval for that set is empty. For the unrejected models, we infer confidence intervals for economic parameters of interest from the estimated bounds of the risk-aversion parameter. The empirical managerial compensation literature came to the conclusion that the main barrier to the resolution of whether current compensation practices are efficient boils down to a measurement problem. This is because the main elements needed to assess the efficiency of executives’ compensation are unobservable. Therefore, a critical question is whether one can always come up with models that can match any empirical regularity.³⁶ This article takes the two main theoretical models in this literature and shows that they are identified and have empirical content. Hence, they cannot match any and all empirical regularity. The basic results in the model apply to extensions in several directions. *Gayle et al.* (2014) draw upon the PMH model and extends it to include job turnover and career concerns, while *Li* (2013) investigates team production within the PMH paradigm.

The second part of the article applies the models to executive compensation in the presence of possible falsification of accounting reports to illustrate the methodology developed in the first

36. See *Abowd and Kaplan* (1999) and *Oyer and Schaefer* (2011) for summaries of this literature.

part of the article. We reject models in which accounting reports are verifiable, but we cannot reject models in which they can be manipulated. The HMH model is never rejected; it yields plausible estimates, indirectly confirming the widespread belief that managers are indeed capable of manipulating accounting numbers and that shareholders discourage this practice through the design of their compensation schemes. Finally, the PMH model in which accounting reports can be manipulated and the cost of reporting depends on the state, can be interpreted as a specialization of an HMH model in which the truth-telling and sincerity constraints are either redundant or ignored. They might be redundant because a monitoring technology imposing disutility on the principal renders the revelation principle redundant, or ignored because shareholders do not fully optimize over the contract space. This PMH model yields counterintuitive results—managers are willing to pay shareholders for the privilege of holding the job in bad accounting states—evidence leading us to conclude that principal-agent models seeking to explain accounting reporting should explicitly incorporate the constraints on the contract space arising from hidden information.

APPENDIX

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 expressed as

$$\alpha_2 E[v(x)] \leq 1. \quad (\text{A.1})$$

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

$$\alpha_2 E[v(x)] \leq \alpha_1 E[v(x)g(x)]. \quad (\text{A.2})$$

To minimize expected compensation subject to (A.1) and (A.2), we choose $v(x)$ to maximize

$$E\{\ln[v(x)]\} + \theta_0 E[1 - \alpha_2 v(x)] + \theta_1 E[\alpha_1 g(x)v(x) - \alpha_2 v(x)]. \quad (\text{A.3})$$

The first-order condition is given by

$$v(x)^{-1} = \theta_0 \alpha_2 + \theta_1 \alpha_2 - \theta_1 \alpha_1 g(x). \quad (\text{A.4})$$

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

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

The complementary-slackness condition for the participation constraint is given by

$$\theta_0 \alpha_2 E[v(x)] - \theta_0 = 0,$$

and substituting 1 for $\theta_0 \alpha_2 E[v(x)]$ into the above proves that $\theta_0 = 1$ and consequently

$$\alpha_2 E[v(x)] = 1. \quad (\text{A.5})$$

Thus, the first-order condition simplifies to

$$\begin{aligned} v(x)^{-1} &= \alpha_2 + \theta_1 \alpha_2 - \theta_1 \alpha_1 g(x) \\ &= \alpha_2 [1 + \theta (\alpha_2 / \alpha_1) - \theta g(x)], \end{aligned} \quad (\text{A.6})$$

where $\theta \equiv \theta_1 \alpha_1 / \alpha_2$. Substituting for $v(x) \equiv \exp[-\gamma w(x)]$ and taking logarithms then yields (8), the optimal work compensation equation. A contradiction argument establishes that the incentive-compatibility constraint holds with equality too. Substituting equation (A.6) into the incentive-compatibility condition and imposing equality gives the solution to θ , namely (9). Finally the optimal contract for shirking is found by setting $\theta_1 = 0$ and substituting α_1 for α_2 in (A.3) and solving for the first-order condition to obtain $\gamma^{-1} \ln(\alpha_1)$. \parallel

Proof of Theorem 2.1. Upon substituting γ^* for γ , equation (15), the expression for $\alpha_2(\gamma^*)$, follows directly from (A.5). Rearranging equation (12) yields

$$e^{\gamma^* \bar{w}} = \alpha_2 [1 + \theta (\alpha_2 / \alpha_1)]. \quad (\text{A.7})$$

Subtracting equation (A.6) from (A.7), we obtain

$$\alpha_2 \theta g(x) = e^{\gamma^* \bar{w}} - e^{\gamma^* w^\theta(x)}. \quad (\text{A.8})$$

Taking the expectation of equation (A.6) and noting that $E[g(x)] = 1$ gives

$$E[e^{\gamma^* w^\theta(x)}] = \alpha_2 [1 + \theta (\alpha_2 / \alpha_1) - \theta]. \quad (\text{A.9})$$

Subtracting equation (A.9) from (A.7), we obtain

$$\alpha_2 \theta = e^{\gamma^* \bar{w}} - E[e^{\gamma^* w^\theta(x)}]. \quad (\text{A.10})$$

Substituting for $\alpha_2 \theta$ using (A.10) in (A.8) and making $g(x)$ the subject of the equation yields the expression for $g(x, \gamma^*)$, given in (13). Substituting for $\alpha_2 \theta$ using (A.10), and also for α_2 using (A.5), in equation (A.9) yields, upon rearrangement, the expression for $\alpha_1(\gamma^*)$ given in (14). \parallel

Proof of Theorem 2.2. Suppose (X, W) is generated from revenue density $f(x)$ and compensation schedule $w(x)$, where the latter is the optimal contract for a PMH model parameterized by $(\gamma, \alpha_1, \alpha_2, g(x))$. We seek to prove Γ is sharp, meaning every $\gamma^* \in \Gamma$ is observationally equivalent, or more precisely, every element in Γ indexes by γ^* a PMH model, in which it is optimal for the principal to induce work, that is observationally equivalent to $(\gamma, \alpha_1, \alpha_2, g(x))$. It is convenient to divide the proof into four steps:

- (1). We define the model indexed by γ^* using equations (14), (15), and (13). Given any $\gamma^* > 0$, real numbers $\alpha_1^* \equiv \alpha_1(\gamma^*)$ and $\alpha_2^* \equiv \alpha_2(\gamma^*)$, and a mapping $g^*(x) \equiv g(x, \gamma^*)$ are defined with respect to the joint probability distribution for (X, W) . Following the arguments in the text, $\alpha_2^* > \alpha_1^* > 0$ and given γ^* , the mapping $g^*: \mathcal{R} \rightarrow \mathcal{R}^+$, satisfies (2) and (3). In this sense γ^* indexes a PMH model defined by $(\gamma^*, \alpha_1^*, \alpha_2^*, g^*(x))$.
- (2). In the paragraphs immediately following the description of the four steps, we show that $w(x)$ is the cost minimizing contract for inducing work when the agent's preferences are α_1^* and α_2^* , and the probability density for revenue if the agent shirks is $g^*(x)f(x)$.
- (3). Appealing to (25) $Q_0(\gamma^*) \geq 0$ if and only if

$$\begin{aligned} 0 &\leq E[x] - E[w(x)] - E \left[x \frac{e^{\gamma^* \bar{w}} - e^{\gamma^* w(x)}}{e^{\gamma^* \bar{w}} - E[e^{\gamma^* w(x)}]} \right] + \gamma^{-1} \ln \left(\frac{1 - E[e^{\gamma^* w(x) - \gamma^* \bar{w}}]}{E[e^{-\gamma^* w(x)}] - e^{-\gamma^* \bar{w}}} \right) \\ &= E[x] - E[w(x)] - E[xg^*(x)] + \gamma^{-1} \ln(\alpha_1^*). \end{aligned}$$

Therefore from (24) the cost minimizing contract for inducing work given by $w(x)$ is profit maximizing for the parameterization of the model given by $(\gamma^*, \alpha_1^*, \alpha_2^*, g^*(x))$.

- (4). We conclude that if $Q_0(\gamma^*) \geq 0$, then γ^* indexes a PMH model in which it is optimal for the principal to induce work by offering the agent $w(x)$ that jointly with $f(x)$ generates the random variable (X, W) . The parameterization $(\gamma^*, \alpha_1^*, \alpha_2^*, g^*(x))$ is therefore observationally equivalent to $(\gamma, \alpha_1, \alpha_2, g(x))$.

The proof is completed by proving the second step. Let $v(x, \gamma^*) \equiv \exp(-\gamma^* w)$ and set $\bar{v}(\gamma^*) \equiv \exp(-\gamma^* \bar{w})$. Since the objective function in (A.3) 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 (A.3) and that the complementary-slackness conditions are satisfied when the Kuhn Tucker multipliers, denoted $\theta_0(\gamma^*)$ and $\theta_1(\gamma^*)$, are defined as

$$\theta_0^* \equiv \{\alpha_2^* E[v(x, \gamma^*)]\}^{-1} \quad (\text{A.11})$$

and

$$\theta_1^* \equiv (\alpha_1^*)^{-1} \left\{ \bar{v}(\gamma^*)^{-1} - E[v(x, \gamma^*)^{-1}] \right\}. \quad (\text{A.12})$$

From their respective definitions, both θ_0^* and θ_1^* are strictly positive as we have already shown in the body of the article that both $\alpha_1(\gamma^*)$ and $\alpha_2(\gamma^*)$ are positive and $\bar{v}(\gamma^*)^{-1} > E[v(x, \gamma^*)^{-1}]$.

- (a). Appealing to the definitions of $\alpha_1(\gamma^*)$, $\alpha_2(\gamma^*)$ and $g(x, \gamma^*)$ given in (13) through (15),

$$\begin{aligned} g^*(x) - \frac{\alpha_2^*}{\alpha_1^*} &= \left\{ \frac{\bar{v}(\gamma^*)^{-1} - v(x, \gamma^*)^{-1}}{\bar{v}(\gamma^*)^{-1} - E[v(x, \gamma^*)^{-1}]} \right\} - \left\{ \frac{\bar{v}(\gamma^*)^{-1} - E[v(x, \gamma^*)^{-1}]}{\bar{v}(\gamma^*)^{-1} - E[v(x, \gamma^*)^{-1}]} \right\} \\ &= \frac{E[v(x, \gamma^*)^{-1}] - v(x, \gamma^*)^{-1}}{\bar{v}(\gamma^*)^{-1} - E[v(x, \gamma^*)^{-1}]} \\ &= \frac{\alpha_2^* \theta_0^* - v(x, \gamma^*)^{-1}}{\alpha_1^* \theta_1^*}, \end{aligned}$$

where the third line follows from the definitions of θ_0^* and θ_1^* defined in (A.11) and (A.12). Rearranging we obtain

$$v(x, \gamma^*)^{-1} = \theta_0^* \alpha_2^* - \theta_1^* \alpha_2^* + \theta_1^* \alpha_1^* g^*(x), \tag{A.13}$$

which is the first-order condition of the cost minimizing contract PMH model for working given in (A.4).

- (b). From the definition of α_2^* implied by (15), the participation constraint in (A.1) is met with equality. From (A.11), θ_0^* is positive. Therefore the complementary-slackness condition for participation is satisfied. Noting from (A.12) that θ_1^* is positive, it follows from its definition, and the expression for $g^*(x) - \alpha_2^*/\alpha_1^*$ given above, that

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

since by (14) $\alpha_1^* > 0$. Thus, the complementary-slackness condition for incentive compatibility is satisfied.

||

Proof of Corollary 2.1. There are three steps. First we show that, for all $\gamma > 0$,

$$E[w] > \gamma^{-1} \ln \left\{ \frac{1 - E[e^{\gamma w} - \gamma \bar{w}]}{E[e^{-\gamma w}] - e^{-\gamma \bar{w}}} \right\}. \tag{A.14}$$

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

$$E[x] < E \left[x \frac{e^{\gamma \bar{w}} - e^{\gamma w}}{e^{\gamma \bar{w}} - E[e^{\gamma w}]} \right]. \tag{A.15}$$

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 (25).

- (1). Since e^w is convex in w , Jensen's inequality implies

$$E[e^{-\gamma w}] > e^{-E[\gamma w]}.$$

Taking the logarithm of each side, dividing through by γ and rearranging yields

$$E[w] + \gamma^{-1} \ln E[e^{-\gamma w}] > 0. \tag{A.16}$$

But from (15) and the discussion following (16)

$$\begin{aligned} E[w] + \gamma^{-1} \ln E[e^{-\gamma w}] &= E[w] - \gamma^{-1} \ln \alpha_2(\gamma) \\ &< E[w] - \gamma^{-1} \ln \alpha_1(\gamma). \end{aligned} \tag{A.17}$$

Combining inequalities (A.16) and (A.17), we obtain (26) upon substituting in the expression for $\alpha_1(\gamma)$ given by (14).

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

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

- (3). Suppose $E[x]=0$, the probability density function for x is symmetric, that is $f(x) = f(-x)$, and let $w(x)$ be monotone decreasing in x , so that $w(-x) > w(x)$ for all $x > 0$. Then,

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

||

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

$$1 = \eta_0 \left[\sum_{s=1}^2 \int_{\underline{x}}^{\infty} \varphi_s v_s(x) f'_s(x) dx \right] \equiv \eta_0 E[v_s(x)],$$

where we use 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 η_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

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

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

Proof of Theorem 3.2. 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 γ^* for γ ,

- (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 η_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 subtracting

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

from

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

yielding

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

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 η_2 above into the first-order condition for the second state evaluated at the limit $x \rightarrow \infty$ to obtain

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

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 α_1 the subject of the equation

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

- (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

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

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

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

as 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

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

Upon collecting terms,

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

solving for η_4 we now have

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

- (6). $\eta_3 = \eta_3(\gamma^*)$ 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

$$\eta_1(\alpha_2/\alpha_1) = \bar{v}_1^{-1} - E[v_s(x)]^{-1} + \eta_3 \bar{h}. \tag{A.18}$$

Making η_1 the subject of the equation 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 η_1 , we establish $g_1(x) = g_1(x, \gamma^*)$.

||

Proof of Theorem 3.2. The proof follows the same steps as the proof of Theorem 2.2. First, we define some candidate values for the Kuhn Tucker multipliers as functions of γ and establish they are positive. Then we show that if $\gamma \in \Gamma_H$, the first-order conditions for the optimization problem in (34) are satisfied in both states. Next, we demonstrate that 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. Finally, we derive the formula for $\Psi_1(\gamma)$ and show that $\Psi_1(\gamma^*) = 0$.

- (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 (40), (41) and (42), these equations define candidate values for the five Kuhn Tucker multipliers in the γ parameterization of the hybrid model. 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 Γ_H .

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

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

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 (35) 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 that the participation constraint is met with equality, and hence the complementary-slackness condition for participation is satisfied. The complementary-slackness conditions for the truth-telling and sincerity constraints are directly imposed by virtue of $\gamma \in \Gamma H$. 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 pure 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

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

proving from (29) that the complementary-slackness condition for incentive compatibility in the second state holds.

Multiplying the first line of (35), 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

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

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 (29), it now follows that the complementary-slackness condition for incentive compatibility in the first state also holds.

(4). To show that $\Psi_1(\gamma^*) = 0$, we note that α_1 appears in both incentive compatibility constraints defined in equation (29). The complementary slackness conditions associated with equation (29) are:

$$\eta_1(\gamma^*) E_1 \left\{ \left[1 - (\alpha_1(\gamma^*)/\alpha_2(\gamma^*)) g_1(x, \gamma^*) \right] v_1(x, \gamma^*) \right\} = 0 \quad (\text{A.19})$$

$$\eta_2(\gamma^*) E_2 \left\{ \left[1 - (\alpha_1(\gamma^*)/\alpha_2(\gamma^*)) g_2(x, \gamma^*) \right] v_2(x, \gamma^*) \right\} = 0. \quad (\text{A.20})$$

Rearranging equations (A.19) and (A.20) and making $\alpha_2(\gamma)/\alpha_1(\gamma)$ the subject gives:

$$\alpha_2(\gamma^*)/\alpha_1(\gamma^*) = E_1[g_1(x, \gamma^*) v_1(x, \gamma^*)] E_1[v_1(x, \gamma^*)]^{-1} = E_2[g_2(x, \gamma^*) v_2(x, \gamma^*)] E_2[v_2(x, \gamma^*)]^{-1}.$$

Since $\eta_1(\gamma^*) > 0$ then we define $\Psi_1(\gamma)$ as:

$$\Psi_1(\gamma) \equiv \eta_1(\gamma) \{ E_2 [g_2(x, \gamma) v_2(x, \gamma)] E_2 [v_2(x, \gamma)]^{-1} - E_1 [g_1(x, \gamma) v_1(x, \gamma)] E_1 [v_1(x, \gamma)]^{-1} \}. \quad (A.21)$$

First we simplify the first term in equation (A.21) by making $\widehat{\alpha}_2(\gamma)/\widehat{\alpha}_1(\gamma)$ the subject of equation (36) and rearranging to get:

$$\begin{aligned} \widehat{\alpha}_2(\gamma)/\widehat{\alpha}_1(\gamma) &= \frac{\bar{v}_2(\gamma)^{-1} E_2 [v_2(x, \gamma)] - 1}{\bar{v}_2(\gamma)^{-1} E_2 [v_2(x, \gamma)] - E_2 [v_2(x, \gamma)^{-1}] E_2 [v_2(x, \gamma)]} \\ &= E_2 [g_2(x, \gamma) v_2(x, \gamma)] E_2 [v_2(x, \gamma)]^{-1}, \end{aligned}$$

where the last equality holds by substituting for $g_2(x, \gamma)$ from equation (38). Therefore using the formula for $\eta_1(\gamma)$ in equation (42) implies:

$$\eta_1(\gamma) E_2 [g_2(x, \gamma) v_2(x, \gamma)] E_2 [v_2(x, \gamma)]^{-1} = \bar{v}_1(\gamma)^{-1} - E [v_s(x, \gamma)]^{-1} + \eta_3(\gamma) \bar{h}. \quad (A.22)$$

Turning to the second term in equation (A.21) and substituting for $g_1(x, \gamma)$ from equation (39) gives:

$$\begin{aligned} \eta_1(\gamma) E_1 [g_1(x, \gamma) v_1(x, \gamma)] E_1 [v_1(x, \gamma)]^{-1} &= \bar{v}_1(\gamma)^{-1} - E_1 [v_1(x, \gamma)]^{-1} + \eta_3(\gamma) \bar{h} \\ - \eta_3(\gamma) E_1 [h(x) v_1(x, \gamma)] E_1 [v_1(x, \gamma)]^{-1} &- \eta_4(\gamma) \frac{\widehat{\alpha}_1(\gamma)}{\widehat{\alpha}_2(\gamma)} E_1 [g_2(x, \gamma) h(x) v_1(x, \gamma)] E_1 [v_1(x, \gamma)]^{-1}. \end{aligned} \quad (A.23)$$

Substituting (A.22) and (A.23) into (A.21) gives the formula in equation (44).

||

Proof of Corollary 3.1. Suppose $w_1(x) = w_2(x)$. This implies (49) holds, which implies $E_2 [e^{-\gamma w_2}]^{-1} = E [e^{-\gamma w_s}]^{-1}$. Substituting these equalities into the numerator of (40) implies $\eta_4(\gamma) = 0$ for all $\gamma > 0$, and thence, using (41) $\eta_3(\gamma) = 0$ too. Comparing (36) with (15), and (37) with (14), it immediately follows from (49) that $\alpha_2(\gamma) = \widehat{\alpha}_2(\gamma)$ and $\alpha_1(\gamma) = \widehat{\alpha}_1(\gamma)$. Also setting $\eta_3(\gamma) = \eta_4(\gamma) = 0$ implies $\eta_1(\gamma) = \eta_2(\gamma)$, and hence from (38) and (39) that $g_1(x, \gamma) = g_2(x, \gamma) = g(x, \gamma)$, the last equality following from (13). This proves that if $w_1(x) = w_2(x)$ then $\Gamma_H = \Gamma$. The second sentence in the corollary now follows from Corollary 2.1. ||

Proof of Corollary 3.2. By inspection $\Gamma_H \not\subseteq \Gamma_2$ because Γ_H does not impose the restriction (49). To complete the proof by showing $\Gamma_2 \not\subseteq \Gamma_H$, we first determine Γ_2 and show it does not impose the truth telling and sincerity constraints.

Following the comparison of the PMH2 model with the HMH model defined in (34), $E_2 [e^{-\gamma w_2}] = E [e^{-\gamma w_s}]$ in PMH2, which implies $E_2 [e^{-\gamma w_2}]^{-1} = E [e^{-\gamma w_s}]^{-1}$. Substituting these equalities into the numerator of (40) implies $\eta_4(\gamma) = 0$ for all $\gamma > 0$, and thence, using (41) $\eta_3(\gamma) = 0$ too. Comparing (36) with (15), and (37) with (14), it immediately follows from (49) that $\alpha_2(\gamma) = \widehat{\alpha}_2(\gamma)$ and $\alpha_1(\gamma) = \widehat{\alpha}_1(\gamma)$. Also setting $\eta_3(\gamma) = \eta_4(\gamma) = 0$ implies from (38) and (39) that $g_s(x, \gamma) \equiv (e^{\gamma w_s} - e^{\gamma w_s(x)}) / (e^{\gamma w_s} - E [e^{-\gamma w_s(x)}])$ for $w_s \equiv w_s(x)$. Then, analogous to (25), we obtain for $s \in \{1, 2\}$

$$\Upsilon_s(\gamma) \equiv E [x] - E [w_s] - E \left[x \frac{e^{\gamma w_s} - e^{\gamma w_s}}{e^{\gamma w_s} - E [e^{\gamma w_s}]} \right] + \gamma^{-1} \ln \left(\frac{1 - E [e^{\gamma w_s - \gamma w_s}]}{E [e^{-\gamma w_s}] - e^{-\gamma w_s}} \right). \quad (A.24)$$

Combining the restrictions for PMH2, the observationally equivalent set for that model is

$$\Gamma_2 \equiv \{ \gamma > 0; \Upsilon_s(\gamma) \geq 0 \text{ for } s \in \{1, 2\} \text{ and } E_1 [e^{-\gamma w_1}] = E_2 [e^{-\gamma w_2}] \}.$$

To show $\Gamma_2 \not\subseteq \Gamma_H$ we impose the additional restriction (49) on Γ_H and then deduce, following the arguments in previous paragraph that $\Upsilon_s(\gamma) \geq 0$ for $s \in \{1, 2\}$. Therefore the three restrictions defining Γ_2 for PMH2 also apply to HMH when (49) holds. But in addition, all elements in Γ_H satisfy $\Psi_3(\gamma) \geq 0$, $\Psi_4(\gamma) \geq 0$, and $\Psi_3(\gamma) \Psi_4(\gamma) = 0$. (The discussion following (43) refers.) Therefore $\Gamma_2 \not\subseteq \Gamma_H$. ||

Acknowledgments. We thank two editors of this journal, four referees, Chen Li, Maher Said, Melissa Tartari, and Steve Tadelis for their comments. We have also benefited from presentations at the Econometric Society 2007 Summer Meetings, SITE Theory-Based Micro-Econometric Modeling 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.

Supplementary Data

Supplementary materials are available at *Review of Economic Studies* online.

REFERENCES

- ABOWD, J. M. and KAPLAN, D. S. (1999), "Executive Compensation: Six Questions That Need Answering", *The Journal of Economic Perspectives*, **13**, 145–168.
- ACKERBERG, D. A. and BOTTICINI, M. (2002), "Endogenous Matching and the Empirical Determinants of Contract Form", *Journal of Political Economy*, **110**, 564–591.
- AGGARWAL, R. K. and SAMWICK, A. A. (1999), "The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation", *Journal of Political Economy*, **107**, 65–105.
- ANDREWS, D. W. and BARWICK, P. J. (2012), "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure", *Econometrica*, **80**, 2805–2826.
- ANDREWS, D. W. and GUGGENBERGER, P. (2009), "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities", *Econometric Theory*, **25**, 669.
- ANDREWS, D. W. and SOARES, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", *Econometrica*, **78**, 119–157.
- ANTLE, R. and SMITH, A. (1985), "Measuring Executive Compensation: Methods and an Application", *Journal of Accounting Research*, **23**, 296–325.
- BAKER, G. P. and HALL, B. J. (2004), "CEO Incentives and Firm Size", *Journal of Labor Economics*, **22**, 767–798.
- BENMELECH, E., KANDEL, E. and VERONESI, P. (2010), "Stock-Based Compensation and CEO (Dis) Incentives", *The Quarterly Journal of Economics*, **125**, 1769–1820.
- BIAIS, B., BISIÈRE, C. and DÉCAMPS, J. P. (2000), "A Structural Econometric Investigation of the Agency Theory of Financial Structure", *In Proceedings of the American Finance Association Meeting*, Boston.
- BOLTON, P. and DEWATRIPONT, M. (2005), *Contract Theory* (Cambridge: MIT Press).
- BRUNK, H. D. (1958), "On the Estimation of Parameters Restricted by Inequalities", *Annals of Mathematical Statistics*, **29**, 437–454.
- CAMPO, S., GUERRE, E., PERRIGNE, I. and VUONG, Q. (2011), "Semiparametric Estimation of First-Price Auctions with Risk-Averse Bidders", *The Review of Economic Studies*, **78**, 112–147.
- CARDON, J. H. and HENDEL, I. (2001), "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey", *RAND Journal of Economics*, **32**, 408–427.
- CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243–1284.
- COPELAND, A. and MONNET, C. (2009), "The Welfare Effects of Incentive Schemes", *The Review of Economic Studies*, **76**, 93–113.
- CROCKER, K. J. and MORGAN, J. (1998), "Is Honesty the Best Policy? Curtailing Insurance Fraud Through Optimal Incentive Contracts", *Journal of Political Economy*, **106**, 355–375.
- CROCKER, K. J. and SLEMROD, J. (2007), "The Economics of Earnings Manipulation and Managerial Compensation", *The RAND Journal of Economics*, **38**, 698–713.
- DUBOIS, P. and VUKINA, T. (2004), "Grower Risk Aversion and the Cost of Moral Hazard in Livestock Production Contracts", *American Journal of Agricultural Economics*, **86**, 835–841.
- DYE, R. A. and SRIDHAR, S. S. (2005), "Moral Hazard Severity and Contract Design", *RAND Journal of Economics*, **36**, 78–92.
- EDMANS, A., GABAIX, X., SADZIK, T. and SANNIKOV, Y. (2012), "Dynamic CEO Compensation", *The Journal of Finance*, **67**, 1603–1647.
- FERRALL, C. and SHEARER, B. (1999), "Incentives and Transactions Costs within the Firm: Estimating an Agency Model using Payroll Records", *The Review of Economic Studies*, **66**, 309–338.
- FINNERTY, J. E. (1976), "Insiders and Market Efficiency", *Journal of Business*, **31**, 1141–1148.
- FRISCH, R. (1934), "Statistical Confluence Analysis by Means of Complete Regression Systems (Vol. 5)" (Oslo, Norway: Universitetets Økonomiske Institutt).
- FUDENBERG, D., HOLMSTROM, B. and MILGROM, P. (1990), "Short-Term Contracts and Long-Term Agency Relationships", *Journal of Economic Theory*, **50**, 1–31.
- GAYLE, G.-L. and MILLER, R. A. (2009a), "Insider Information and Performance Pay", *CESifo Economic Studies*, **55**, 515–541.
- GAYLE, G.-L. and MILLER, R. A. (2009b), "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review*, **99**, 1740–1769.
- GAYLE, G.-L., GOLAN, L. and MILLER, R. A. (2014), "Promotion, Turnover and Compensation in the Executive Labor Market", (Working Paper, Tepper School of Business, Carnegie Mellon University).
- GROSSMAN, S. and HART, O. (1983), "An Analysis of the Principal-Agent Problem", *Econometrica*, **51**, 7–46.
- GUERRE, E., PERRIGNE, I. and VUONG, Q. (2009), "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions", *Econometrica*, **77**, 1193–1227.
- HALL, B. J. and LIEBMAN, J. B. (1998), "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, **113**, 653–680.
- HAUBRICH, J. G. (1994), "Risk Aversion, Performance Pay, and the Principal-Agent Problem", *Journal of Political Economy*, **102**, 258–276.
- HAYES, R. M. and SCHAEFER, S. (2000), "Implicit Contracts and the Explanatory Power of Top Executive Compensation for Future Performance", *RAND Journal of Economics*, **31**, 273–293.
- JAFFE, J. F. (1974), "Special Information and Insider Trading", *Journal of Business*, **47**, 410–428.

- JENSEN, M. C. and MURPHY, K. J. (1990), "Performance Pay and Top-Management Incentives", *The Journal of Political Economy*, **98**, 225–264.
- KARTIK, N. (2009), "Strategic Communication with Lying Costs", *The Review of Economic Studies*, **76**, 1359–1395.
- LACKER, J. M. and WEINBERG, J. A. (1989), "Optimal Contracts Under Costly State Falsification", *Journal of Political Economy*, **97**, 1345–1363.
- LAFFONT, J. J. and MATOUSSI, M. S. (1995), "Moral Hazard, Financial Constraints and Sharecropping in El Oulja", *The Review of Economic Studies*, **62**, 381–399.
- LAFFONT, J.-J. and MARTIMORT, D. (2002), *The Theory of Incentives* (Princeton: Princeton University Press).
- LAFFONT, J. J. and TIROLE, J. (1986), "Using Cost Observation to Regulate Firms", *The Journal of Political Economy*, **94**, 614–641.
- LAZEAR, E. P. (2000), "The Power of Incentives", *American Economic Review*, **90**, 410–414.
- LI, C. (2013), "Mutual Monitoring within Top Management Teams: A Structural Modeling Investigation", (Working paper, Baruch College, City University of New York).
- LORIE, J. H. and NIEDERHOFFER, V. (1968), "Predictive and Statistical Properties of Insider Trading", *Journal of Law and Economics*, **11**, 35–51.
- MAGGI, G. and RODRIGUEZ-CLARE, A. (1995), "Costly Distortion of Information in Agency Problems", *RAND Journal of Economics*, **26**, 675–689.
- MALCOMSON, J. M. and SPINNEWYN, F. (1988), "The Multiperiod Principal–Agent Problem", *Review of Economic Studies*, **55**, 391–407.
- MARGIOTTA, M. and MILLER, R. A. (2000), "Managerial Compensation and the Cost of Moral Hazard", *International Economic Review*, **41**, 669–719.
- MAS-COLELL, A., WHINSTON, M. D. and GREEN, J. R. (1995), *Microeconomic Theory* (Vol. 1), (New York: Oxford University Press).
- MIRPLEES, J. A. (1975), "The Theory of Moral Hazard and Unobservable Behavior, Part 1", (Working Paper, Oxford University).
- MYERSON, R. (1982), "Optimal Coordination Mechanisms in Generalized Principal–Agent Problem", *Journal of Mathematical Economics*, **10**, 67–81.
- NEWBY, W. K. and MCFADDEN, D. (1994), "Estimation and Hypothesis Testing Large Sample", In *Handbook of Econometrics*, 4, Engle, R. F. and McFadden, D. L. (eds.), (Amsterdam: Elsevier), 2111–2245.
- OYER, P. (2004), "Why Do Firms Use Incentives That Have No Incentive Effects?" *Journal of Finance*, **59**, 1619–1650.
- OYER, P. and SCHAEFER, S. (2011), "Personnel Economics: Hiring and Incentives", *Handbook of Labor Economics*, **4**, 1769–1823.
- PAARSCH, H. J. and SHEARER, B. (2000), "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records", *International Economic Review*, **41**, 59–92.
- PAKES, A., PORTER, J., HO, K. and ISHII, J. (2006), "Moment Inequalities and Their Application" (CEMMAP Working Paper).
- PARSONS, V. L. (1978), "The Limiting Distribution of the Isotonic Estimator at a Point", (Unpublished Manuscript).
- PENG, L. and RÖELL, A. (2008), "Manipulation and Equity-Based Compensation", *The American Economic Review*, **98**, 285–290.
- PERRIGNE, I. and VUONG, Q. (2011), "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard", *Econometrica*, **79**, 1499–1539.
- REY, P. and SALANIE, B. (1990), "Long-Term, Short-Term and Renegotiation: On the Value of Commitment in Contracting", *Econometrica*, **58**, 597–619.
- ROMANO, J. P. and SHAIKH, A. M. (2010), "Inference for the Identified Set in Partially Identified Econometric Models", *Econometrica*, **78**, 169–211.
- ROSEN, A. M. (2008), "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities", *Journal of Econometrics*, **146**, 107–117.
- SEYHUN, H. N. (1986), "Insiders' Profits, Cost of Trading, and Market Efficiency", *Journal of Financial Economics*, **16**, 189–212.
- SEYHUN, H. N. (1992a), "The Effectiveness of Insider Trading Sanctions", *Journal of Law and Economics*, **35**, 149–182.
- SEYHUN, H. N. (1992b), "Why Does Aggregate Insider Trading Predict Future Stock Returns?" *Quarterly Journal of Economics*, **107**, 1303–1331.
- SHEARER, B. (1996), "Piece-Rates, Principal-Agent Models, and Productivity Profiles: Parametric and Semi-Parametric Evidence from Payroll Records", *Journal of Human Resources*, **31**, 275–303.
- SHEARER, B. (2004), "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment", *The Review of Economic Studies*, **71**, 513–534.
- VERA-HERNANDEZ, M. (2003), "Structural Estimation of a Principal-Agent Model: Moral Hazard in Medical Insurance", *RAND Journal of Economics*, **34**, 670–693.
- WRIGHT, F. T. (1981), "The Asymptotic Behavior of Monotone Regression Estimates", *Annals of Statistics*, **9**, 443–448.