2016 CAPANA / CJAR Conference

8-9 July 2016

Paper Session 3

Are Top Management Teams Compensated as Teams? A Structural Approach

By

Chen Li
The City University of New York

Are Top Management Teams Compensated as Teams? A Structural Approach

Chen Li*

Zicklin School of Business

Baruch College

The City University of New York

December 22, 2015

*This paper is based on chapter 2 of my dissertation at Carnegie Mellon University. I am indebted to my dissertation committee, George-Levi Gayle, Jonathan Glover, Pierre Jinghong Liang, and Robert Miller, for their invaluable guidance and support. I also thank Michael Ewens, Masako Darrough, Diane Denis, Karam Kang, Thomas Ruchti, Douglas Schroeder, Jack Stecher, Martin Wu, Anastasia Zakolyukina, Ran Zhao, Ronghuo Zheng, and workshop participants at Carnegie Mellon University, CUNY-Baruch College, Ohio State University, Stanford University, University of Chicago, and Baruch-SWUFE Conference 2014 for their comments. I acknowledge the financial support from a PSC-CUNY Award, jointly funded by the Professional Staff Congress and the City University of New York. All errors are my own. E-mail: Chen.Li@baruch.cuny.edu.

Are Top Management Teams Compensated as Teams? A Structural Approach

Using data from S&P1500 firms from 1993 to 2005, this paper structurally tests two competing models of the principal–multiagent moral hazard. Called the Team Model and Individual Model, respectively, they are intended to capture a crucial consideration in executive compensation design, that is, whether shareholders are concerned about top managers' unilateral deviation in effort and incentivize top managers as a team or as separate individuals. The Team Model can rationalize the observed relationship between executive compensation and stock returns. The Individual Model can be plausible only when it appeals to an extra assumption that managers have heterogeneous risk preferences across firm characteristics and industrial sectors; thus it is less robust. These findings suggest that isolating the CEO from other executives in research designs and regulatory considerations regarding executive compensation seems less preferred.

Keywords: Moral Hazard, Top Management Team, Executive Compensation, Structural Model

JEL Codes: D82, J33, M12, M52

1 Introduction

Top managers' activities are hardly observed by shareholders and may deviate from shareholders' objectives. This problem, called moral hazard, is one of the main theoretical reasons why shareholders base compensation on firm–denominated securities. Researchers in accounting, finance, and economics have been pervasively testing whether the observed executive compensation conforms to the optimal contract derived from a moral hazard model.[1] However, based on even a basic moral hazard model, understanding the efficiency of observed executive compensation and drawing inferences about compensation reforms depend on whether shareholders are motivating the managers as individuals or a team. This paper investigates this important yet underexplored question.

Specifically, linking each manager's compensation to overall firm performance, shareholders may or may not take advantage of interest alignment within top management teams. Depending on how shareholders perceive and respond to the moral hazard problem, theoretical models provide different empirical implications for the components in optimal compensation. If managers can only shirk together, shareholders may motivate the managers as a team by providing incentives to avoid joint shirking only. In the optimal compensation predicted by a team model, the shadow price of incentive compatibility constraint reflects the marginal cost of all managers shirking, and the likelihood ratio of stock returns captures that the informativeness of performance measure is affected by the distribution of stock returns conditional on all managers shirking and that distribution conditional on all managers working. By contrast, if one manager could shirk no matter whether other managers work, shareholders need to motivate them as individuals by providing incentives to avoid unilateral shirking in addition to joint shirking. In the optimal compensation predicted by an individual model, the shadow price reflects the marginal cost of unilateral shirking, and the likelihood ratio is affected by the distribution of stock returns conditional on one manager shirking and that

[1] Several surveys are provided by Murphy [1999, 2012], Bushman and Smith [2001], Core et al. [2003], and Edmans [2009].

distribution conditional on all managers working. Given the preceding theoretical distinctions and empirical predictions provided by the two types of models, a test of which type can better reflect the incentive nature underlying the observed relationship between executive compensation and stock returns is important not only for academic studies on the optimality of executive compensation but also for regulations that intend to improve corporate governance. This paper performs this test.

Previous studies using a nonstructural approach to infer the incentive nature have obtained few and mixed results, because their research designs face the challenge that the shadow price and likelihood ratio which determine the compensation shape in theory cannot be observed by researchers, in addition to other unobservable primitives in a typical moral hazard model such as effort costs and risk aversion, such that those studies have to rely on indirect tests and examine only the implications of theory.[2] These insufficiencies call for the structural approach used in this paper, which directly examines the restrictions that each model imposes on data as a whole.[3] Doing so can to a large extent mitigate the preceding concerns over indirectly testing models with theoretically important components left out of simple regressions, because these internally consistent restrictions explicitly capture the incentive nature of different contract types and discipline unmeasurable parameters together with observed compensation and stock returns within a unified framework.

Called the *Individual Model* and *Team Model*, respectively, the two models in this paper may both be used to rationalize observed executive compensation but depart in terms of a crucial consideration in compensation design, that is, whether shareholders are concerned about top managers' unilateral deviation in effort choice and, consequently, whether they

[2] Main et al. [1993] attribute the positive association between compensation variation within management teams and firm performance to tournaments (an individual incentive mechanism). Henderson and Fredrickson [2001] get ambiguous conclusions by showing that the relationship between the executive pay gap and ROA is affected by both coordination needs and a demand for individual incentives. Li [2014] attributes the negative relationship between future firm performance and current pay gap between the CEO and the second highest paid executive to mutual monitoring (team incentive). Bushman et al. [2016] find that the deleterious effect of deviating from the optimal contract (proxied by the dispersion of pay-performance-sensitivity in top management teams) is affected by factors related to managerial coordination.

[3] Ittner and Larcker [2002] and Gow et al. [2015] make a similar point.

incentivize the members of a top management team as separate individuals or as a team. In the *Individual Model*, given that unilateral shirking is possible, shareholders motivate both managers working as a Nash equilibrium in the managers' subgame and prefer both working to either or both managers shirking. In the *Team Model*, conditional on that managers can only shirk together and working is Pareto optimal, shareholders compare both working only with both shirking. The key theoretical feature of the *Team Model* is that a team incentive compatibility constraint replaces an individual incentive compatibility constraint. The key implication for testing the *Team Model* is that the likelihood ratio stays the same in the optimal compensation between both managers.

The intuition for my empirical strategy is as follows. Even though we do not know how shareholders design the optimal compensation, we do observe the compensation they offer, and managers generate the output in the form of market returns. The optimal compensation contract of one type can essentially be described by a distinct, well-defined theoretical model. It is standard in structural approach literature to assume that both shareholders and managers behave as the model predicts. Consequently, if the observed data pattern is statistically consistent with a group of restrictions imposed on data by that theoretical model, the consistency suggests that the observed compensation schemes feature the corresponding model. The purpose of the model specification tests is to find out which model, as a whole, can better statistically explain the joint distribution of compensation and stock returns, allowing the contract shape to vary with firm characteristics, industrial sectors, and macroeconomic fluctuations.

The data used in this paper cover S&P 1500 firms from 1993 to 2005. In addition to the total compensation from the ExecuComp database, the opportunity costs of holding firm stocks and stock options are included in managers' total compensation to comprehensively measure the incentive pay. This measurement follows Antle and Smith [1985].[4] Also, both the density of the gross abnormal return (performance measure in this paper) in equilibrium

[4] Hall and Liebman [1998], Margiotta and Miller [2000], Gayle and Miller [2009, 2015], Gayle et al. [2015a], and Gayle et al. [2015b] also follow the same approach.

and the optimal compensation scheme are nonparametrically estimated. The nonparametric method can exploit the information from data as much as possible and also avoid rejecting a model due to specific functional assumptions on contract form and distribution.

The result of model identification shows that, without imposing on data the restrictions implied by shareholders' profit maximization, the pattern of compensation and stock returns can be empirically consistent with either model. This result indicates that the descriptive properties of compensation—usually based on comparative statics derived from the subset of equilibrium conditions in shareholders' cost minimization problem—may not be sufficient to help us distinguish the two potential underlying models without considering other restrictions that those confounding unobservable parameters need to satisfy.

The results of model specification tests show that the *Team Model* is more robust than the *Individual Model* in rationalizing the correlation between the observed top executive compensation and stock returns. Specifically, under the least restrictive assumption that managers have heterogeneous risk preferences across firm types and industrial sectors, both the *Individual Model* and the *Team Model* cannot be rejected. However, under the most restrictive assumption that managers have homogeneous risk preference across firm types and industries, only the *Team Model* cannot be rejected. These results suggest that the *Team Model* tends to better capture the incentive nature of executive compensation.

This paper extends existing literature from a few aspects. First, this paper is closely related to a small and growing literature in which researchers use a structural approach to analyze executive compensation. The closest to the present study is Gayle and Miller [2015], who use the same semiparametric set identification technique to identify three single-agent moral hazard models and compare the empirical relevance of those models. This paper is distinguished from the preceding by identifying and testing multiagent models. A few other papers in this line of research instead focus on counterfactual estimation. Margiotta and Miller [2000] and Gayle and Miller [2009] quantify welfare costs derived from parametric models of a single agent to examine the importance of moral hazard in executive compensation. Gayle et al. [2015b] estimate the social welfare costs related to CEO compensation before and after the Sarbanes–Oxley Act (SOX) to investigate the consequences of regulatory intervention for private contracts. Gayle et al. [2015a] quantify different sources of pay gaps, including moral hazard, human capital accumulation, and career concerns. Taylor [2013] quantitatively analyzes how learning about a CEO's ability affects the pay level.

Second, this paper enriches empirical studies on executive compensation in accounting literature by, for the first time, adopting the structural approach to explicitly test the incentive nature of compensation for top management teams. Investigations of the compensation of multiple managers together are infrequent in the accounting literature,[5] partially because it is unclear which type of incentive models statistical analyses should be based on. The findings in this paper tend to call for more attention to the positive effects of managerial coordination and the implications generated from team-based models in future empirical research on executive compensation. For example, theoretical studies have suggested factors that affect incentive pay in team settings, such as reputation concern and group identity (Itoh [1990]), corporate culture (Kreps [1990]), and long-term relationships (Arya et al. [1997], Che and Yoo [2001]).

More broadly, this paper joins the very recently growing application of the structural approach in accounting research, as evidenced by Zakolyukina [2014] and Beyer et al. [2014] on earnings management; Bertomeu et al. [2015] on voluntary disclosure; and Gerakos and Syverson [2015] on audit market competition. These papers focus on estimating a given structural model rather than testing alternative models, which is the focus of this paper. Also, we examine different areas of accounting research and exploit different identification strategies.

Testing the popularly adopted incentive nature of executive compensation, as this paper does, also sheds light on regulation and practice. Fostering a stable and close network

[5]Recent exceptions include Li et al. [2014], Li [2012], and Bushman et al. [2016]. Although the management literature tends to conclude that "attention to executive groups, rather than to individuals, often yields better explanations of organizational outcomes" [(Hambrick [2007, p. 334]), its emphasis is on behavioral integration and collective cognition based on demographic characteristics.

within top management teams may be beneficial to a firm, but otherwise could be detrimental if the managers tend to collude against shareholders' interests.[6] The distinct potential consequences make it important for any regulatory reform on corporate governance to acknowledge the common properties of incentives provided in most executive compensation packages. Given the findings in this paper, more regulations that extend monitoring to non-CEOs may be expected in the future. For example, the recent SEC clawback proposal covers more officers than just CEOs and CFOs, who are the focus of the SOX which was enacted in 2002. For an individual firm, when coordination among managers has been taken advantage of by the majority of its peers, as this paper suggests, investment in human resources to facilitate cooperation seems necessary to maintain competitive advantage.

The rest of the paper is organized as follows. Section 2 lays out two theoretical models and the optimal contract derived from each model. Section 3 describes the data used in the empirical implementation. Section 4 establishes the identification. Section 5 discusses the hypothesis tests and the procedures of structural estimation. Section 6 reports the results of structural estimation and model specification tests. Section 7 concludes.

2 Models

This section lays out the two principal-multiagent models of moral hazard as the theoretical underpinning of the identification and hypothesis tests. The two structural models in this paper aim to sufficiently distinguish the incentive nature up to the extent that the primitive parameters can be recovered from the observed compensation and stock returns. These models, however, are not constructed to comprehensively explore the delicate strategic interactions between and within shareholders and managers in complex reality.

This section is structured as follows. After introducing the timeline, production technology, and managers' preferences for the two models, for each model, I first solve the share-

[6] Larcker and Tayan [2013] discuss the benefits and challenges in implementing a corporate governance system that relies on trust.

holders' cost minimization problem (as in the second step of Grossman and Hart [1983]) and then solve their profit maximization problem (as in the first step of Grossman and Hart [1983]). At the end of this section, I compare these two models with a few other models in the literature that study moral hazard problems of multiple agents.

2.1 Timeline

In each model, risk-neutral shareholders (principal) and two risk-averse managers (agents) interact as follows. At the beginning of a period, the shareholders propose a compensation scheme $w_i(x)$ for manager $i = 1, 2$; x is the joint output whose distribution is conditional on the effort choices of the two managers. Let V denote the firm value at the beginning of this period and $\widetilde{x}$ denote the abnormal stock return realized from this period; $\widetilde{x}$ is the idiosyncratic component of the firm's stock return, which is under the control of the managers. The performance measure x, called gross abnormal return, is the return to shareholders before compensating the managers and is given by

$$x = \widetilde{x} + \frac{w_1}{V} + \frac{w_2}{V}. \tag{1}$$

Facing the shareholders' offer, each manager decides whether to take the offer or reject. If one manager rejects the offer, he gets his outside option. It is assumed that neither manager can operate the firm by himself and that one manager has to wait for another manager to join the team to proceed together. This is realistic because modern firms are large such that they are rarely run by a single manager.

After accepting the shareholders' offer, each manager can choose between working and shirking. The interdisciplinary nature of managing large diversified firms requires that top managers work together to make better decisions. The frequent interaction in their routine work makes it possible for them to observe each other's effort, but it can be hard to describe to anyone outside the teams.[7] Both models assume that the two managers can observe each other's effort choice, but the shareholders cannot observe these choices. Such information asymmetry between the shareholders and managers creates a moral hazard problem, considering that more managerial effort can benefit the shareholders but is more costly to the managers. The moral hazard due to managers' unobservable efforts is the fundamental friction in single-agent models. In multiagent models, there is another potential friction called *free riding*. If one manager shirks, he can avoid his entire disutility of working but only has to partially bear the loss from the reduction in output as long as the other manager works. Thus each manager has an incentive to count on the other one and shirks. Let j_i denote manager i's effort choice. The three mutually exclusive choices are defined as follows:

$$j_i = \begin{cases} 0, & \text{if manager } i \text{ rejects the offer,} \\ 1, & \text{if manager } i \text{ accepts the contract but shirks later,} \\ 2, & \text{if manager } i \text{ accepts the contract and works later.} \end{cases} \tag{2}$$

At the end of the period, the joint output x is realized and manager i gets paid according to his compensation scheme $w_i(x)$. Conditional on the two managers' effort choice (j_1, j_2), x is a random draw from an independent and identical distribution across firms of the same type.

2.2 Production Technologies

The technologies that the two managers use to generate output are captured by the probability density function of the joint output x conditional on the two managers' effort choices. Denote $f(x)$ as the density of x conditional on both managers working (the effort pair on the equilibrium path). Throughout this paper, I use the symbol $E[\bullet]$ to represent the expectation taken over $f(x)$, or $\int \bullet \, f(x)dx$. When manager i chooses to shirk but the other manager chooses to work, the product $g_i(x)f(x)$ denotes the corresponding density of x;

[7] This assumption rules out the revelation mechanism as found in Ma [1988].

$g_i(x)$ is the likelihood ratio between the density of x conditional on manager i's unilateral shirking over the density of x conditional on the equilibrium effort pair. In the framework of single joint output, without specifying the individual contribution as an additive or a multiplicative technology, $g_1(x) \neq g_2(x)$ simply means that shareholders can provide individual incentive to each manager based on his distinct influence on the distribution of the gross abnormal return.[8]

It is noteworthy that the specification of the densities on and off the equilibrium path does not put any functional form restrictions on the distribution. The first advantage of relaxing functional assumptions is that the densities specified in the framework can accommodate discrete distributions used in multiple-agent models close to the present paper, as in Itoh [1993], Arya et al. [1997], and Che and Yoo [2001]. Therefore the framework can be applied to broad models. Many theoretical models do not rely on particular functional forms to guide empirical analysis (Holmstrom [1979, 1982], Grossman and Hart [1983]). Keeping consistent with theory in production, the technology assumption set up for minimal use of structures gives the tests in this paper more credibility.

The second advantage of not being tied to a specific distribution is that the flexibility of densities and thus that of the likelihood ratio frees the optimal compensation form from a rigid linear shape or similar, because the variation of compensation essentially comes from the variation of the likelihood ratio across various realizations of the performance measure. As Hemmer [2004] illustrates with relative performance evaluation studies, the lack of empirical support can be caused by overly relying on linear compensation regression, which is not required by the theoretical results (Holmstrom [1982]). This paper avoids this problem.

Also, the specifications are general enough to capture the type of performance evaluation that shareholders may adopt in reality. To illustrate, one manager may mainly take charge of the right-tail performance of the firm, for instance, the head of a research and development department whose primary task is to maintain high growth or a Chief Marketing Officer

[8]This setup is suggested by Margiotta and Miller [2000] in their discussion of extending their single-agent framework to a multiagent one.

who is responsible for continued market expansion. By contrast, the other manager may be someone who monitors the downside risk of the firm, for instance, a Chief Financial Officer who watches financial stress and bankruptcy risk or a Chief Executive Officer who is responsible for both tails of the gross abnormal return.

Denote $g_0(x)$ as the likelihood ratio of the density of x conditional on both managers shirking over that conditional on both managers working; $g_0(x) \equiv g_1(x)g_2(x)$ is implied by the primitive assumption that one manager's marginal influence on the density of x is unconditional on the other manager's effort choice. This assumption rules out the possibility that the two managers have exactly the same marginal influence on the distribution of the gross abnormal return when they unilaterally shirk. Also, this assumption can be consistent with the managers' efforts sharing substitutability, independence, or complementarity.

The likelihood ratio $g_k(x)$ ($k = 0, 1, 2$) has the following properties: (i) $g_k(x) \geq 0, \forall x$; (ii) $E[g_k(x)] = 1$; (iii) $\lim_{x \to \infty} g_k(x) = 0$, meaning that an extraordinary output can be realized only when no one shirks; (iv) $g_k(x)$ is bounded, which implies that the contract cannot achieve the first best allocation by using a signal that can be perfectly informative at extreme realizations of x (Mirreless [1975]); and (v) the shareholders and managers have conflicting interests in the sense that shareholders can benefit more if the managers work than if they shirk. These assumptions are standard in structural models of optimal contracting (Gayle and Miller [2009, 2015]).

2.3 Managers' Preferences

Each manager's preference can be expressed using a negative exponential utility function with multiplicatively separable preference on effort. The utility function is denoted by $-\widetilde{\alpha}_{ij_i} \exp[-\rho w_i(x)]$, for manager $i = 1, 2$. The two managers have the same coefficient of absolute risk aversion, denoted by ρ, but differ in the cost of effort. The cost is captured by the coefficient $\widetilde{\alpha}_{ij_i}$ ($i = 1, 2$, $j_i = 1, 2$) in the managers' utility functions, corresponding to manager i's effort choice j_i (specified in (2)). For manager i, it is assumed that $0 < \widetilde{\alpha}_{i1} < \widetilde{\alpha}_{i2}$, meaning that manager i would not choose to work if he faced fixed compensation but instead would prefer shirking. To interpret shirking, managers are not necessarily lazy, but instead they pursue their own benefits, which conflict with the shareholders' interests. Take empire building, for example. The managers may exert substantial labor input to pick up projects that maximize their own private perks but do not maximize the firm's value. Also, the outside option of the two managers is assumed to be the same and is denoted by $\widetilde{\alpha}_0$. The utility function is normalized by $\widetilde{\alpha}_0$, and thus the outside option becomes -1, and the effort disutility coefficient thereafter becomes $\alpha_{ij_i} \equiv \widetilde{\alpha}_{ij_i}/\widetilde{\alpha}_0$.

Manager i's compensation $w_i(x)$ is a function of the gross abnormal return x. The expected utility conditional on the distribution of x given the managers' effort pair (j_1, j_2) is

$$-\alpha_{ij_i} E\left[\exp\left(-\rho w_i(x)\right) \mid j_1, j_2\right], \; i = 1, 2. \tag{3}$$

I tabulate the expected utilities conditional on the four effort pairs in the table below. In each of the four cells, manager 1's (the row player) expected utility is in the bottom left corner, and manager 2's (the column player) is in the upper right corner. To simplify notation, $v_i(x) \equiv \exp\left(-\rho w_i(x)\right)$, $i = 1, 2$. Take, for example, $-\alpha_{11}E[v_1(x)g_1(x)]$. It represents the expected utility of manager 1 conditional on the distribution of x when he shirks but manager 2 works.

Effort Choices	Manager 2 Works	Manager 2 Shirks
Manager 1 Works	$-\alpha_{22}E[v_2(x)]$	$-\alpha_{21}E[v_2(x)g_2(x)]$
	$-\alpha_{12}E[v_1(x)]$	$-\alpha_{12}E[v_1(x)g_2(x)]$
Manager 1 Shirks	$-\alpha_{22}E[v_2(x)g_1(x)]$	$-\alpha_{21}E[v_2(x)g_0(x)]$
	$-\alpha_{11}E[v_1(x)g_1(x)]$	$-\alpha_{11}E[v_1(x)g_0(x)]$

2.4 Shareholder's Cost Minimization Problem

This paper assumes that shareholders prefer both managers working to other effort pairs. As in the second step in Grossman and Hart [1983], shareholders minimize the expected compensation required to motivate both managers to work. To describe shareholders' constrained optimization problem, the two models have the same objective function but apply different constraints.

2.4.1 Objective Function

The utility of risk-neutral shareholders is measured in monetary terms. The shareholders' benefit is the expected firm value growth conditional on both managers working, which is a constant when managers' effort choices are fixed. The shareholders' cost is the total compensation paid to the two managers, as a function of the output x. Therefore, to maximize net benefit ex ante, shareholders minimize the expected total compensation of the two managers, and the expectation is taken over the distribution of the gross abnormal return conditional on both managers working.

By definition, $v_i(x)$ is monotonically decreasing in $w_i(x)$, so the objective function of the cost-minimizing shareholders is equivalent to maximizing the following expected value:

$$E\left[\ln v_1(x) + \ln v_2(x)\right]. \tag{4}$$

This objective function in the shareholders' cost minimization problem is the same between the two models. However, depending on whether the shareholders motivate the managers as individuals or a team, shareholders face different constraints across the two models.

2.4.2 Participation Constraint

Shareholders design the optimal compensation contracts such that, at the beginning of the period when managers decide whether to accept or reject the job offer, each manager finds that accepting the offer and working diligently during the following period is weakly better than rejecting the shareholders' offers to instead pursue an outside option. This is the *participation constraint*, which is the same in both the *Individual Model* and the *Team Model* and is given by

$$-\alpha_{i2}E\left[v_i(x)\right] \geq -1, \text{ for manager } i \in \{1,2\}. \tag{5}$$

2.4.3 Incentive Compatibility Constraint

Given that shirking is more tempting to the managers ($\alpha_{i1} < \alpha_{i2}$), to induce both managers to work, the optimal compensation contracts need to provide the managers sufficient incentive not only to accept the offers but also to exert effort in line with the shareholders' interests. This is the *incentive compatibility constraint*, which distinguishes the two models depending on whether the managers can shirk unilaterally or only jointly.

In the *Individual Model*, shareholders design the optimal compensation to induce one manager to work as a best response to the other manager's working; that is, both managers working is a Nash equilibrium in the two managers' subgame. Therefore the incentive compatibility constraint is

$$-\alpha_{i2}E\left[v_i(x)\right] \geq -\alpha_{i1}E\left[v_i(x)g_i(x)\right], \text{ for manager } i \in \{1,2\}. \tag{6}$$

If both working is the unique Nash equilibrium, then one manager's shirking cannot be the best response to the other manager's shirking. This implies that

$$-\alpha_{i2}E\left[v_i(x)g_{-i}(x)\right] > -\alpha_{i1}E\left[v_i(x)g_1(x)g_2(x)\right], \text{ for manager } i \in \{1,2\}.$$

In the *Team Model*, working is preferred only as a Pareto optimal strategy rather than as a Nash equilibrium strategy. This is the whole idea of this model and makes it essentially different from the *Individual Model*. The following incentive compatibility constraint states that for manager i, the expected utility conditional on both working (on the left-hand side) is no less than the expected utility conditional on both shirking (on the right-hand side):

$$-\alpha_{i2}E\left[v_i(x)\right] \geq -\alpha_{i1}E\left[v_i(x)g_0(x)\right], \text{ for manager } i \in \{1,2\}. \tag{7}$$

If the free riding problem exists, then it implies that both shirking is the (unique) Nash equilibrium. Consequently, the following inequalities hold simultaneously:

$$-\alpha_{i1}E\left[v_i(x)g_0(x)\right] > -\alpha_{i2}E\left[v_i(x)g_{-i}(x)\right], \text{ for manager } i \in \{1,2\}.$$

2.5 Optimal Contract

In the *Individual Model*, shareholders minimize the objective function (4) subject to the participation constraints (5) and the incentive compatibility constraints (6). In the *Team Model*, shareholders minimize (4) subject to the participation constraints (5) and the incentive compatibility constraints (7). The following proposition gives the optimal contract for each model.[9]

Proposition 1 *Let μ_i^M be the shadow price associated with manager i's incentive compatibility constraint in model $M \in \{I$ (Individual Model), T (Team Model)$\}$; $w_{i,M}^*(x)$ is the optimal compensation paid to manager i in model M:*

$$\begin{aligned} w_{i,I}^*(x) &= \frac{1}{\rho}\ln\alpha_{i2} + \frac{1}{\rho}\ln\left[1+\mu_i^I - \mu_i^I\left(\frac{\alpha_{i1}}{\alpha_{i2}}\right)g_i(x)\right], \textit{ for manager } i \in \{1,2\} && (8)\\ w_{i,T}^*(x) &= \frac{1}{\rho}\ln\alpha_{i2} + \frac{1}{\rho}\ln\left[1+\mu_i^T - \mu_i^T\left(\frac{\alpha_{i1}}{\alpha_{i2}}\right)g_0(x)\right], \textit{ for manager } i \in \{1,2\}, && (9)\end{aligned}$$

where μ_i^I uniquely solves

$$E\left[\frac{\alpha_{i1}g_i(x)}{\alpha_{i2}+\mu_i^I\alpha_{i2}-\mu_i^I\alpha_{i1}g_i(x)}\right] = E\left[\frac{\alpha_{i2}}{\alpha_{i2}+\mu_i^I\alpha_{i2}-\mu_i^I\alpha_{i1}g_i(x)}\right], \textit{ for manager } i \in \{1,2\}$$

[9] All proofs in this paper are saved for Appendix A.

and μ_i^T *uniquely solves*

$$E\left[\frac{\alpha_{i1}g_0(x)}{\alpha_{i2}+\mu_i^T\alpha_{i2}-\mu_i^T\alpha_{i1}g_0(x)}\right]=E\left[\frac{\alpha_{i2}}{\alpha_{i2}+\mu_i^T\alpha_{i2}-\mu_i^T\alpha_{i1}g_0(x)}\right], \textit{ for manager } i\in\{1,2\}.$$

The intuition is as follows. In the *Individual Model*, the managers are incentivized to respond to their own influence on the distribution of the gross abnormal return, so that the optimal compensation accounts for the informativeness of the joint output differently between the two managers, that is, $g_1(x)$ and $g_2(x)$ enter the formula, respectively. In the *Team Model*, the optimal contract merely prevents the managers' simultaneous shirking and thus relies on the informativeness of the joint output drawn from the distribution conditional on both managers shirking, which is captured by $g_0(x)$. Also, in the *Team Model*, the incentive effect provided to one manager will be dampened by the other manager's effort choice. This implies that the marginal cost of motivating a single manager using the optimal contract suggested by the *Team model* will differ from that marginal cost in the *Individual Model*. The different values of the shadow price between the two models (i.e., μ_i^I and μ_i^T) capture such discrepancy.

Importantly, if the observed compensation and stock returns are generated from the equilibrium of a model, the managers' risk attitude (ρ), their effort tastes (α_{ij}), and the informativeness of the performance signal ($g_i(x)$ or $g_0(x)$) together explain the compensation of each manager. Relative features of the two managers' compensation schemes, for example, the pay gap or the difference of pay-performance-sensitivity between managers, can be rationalized by any of these two models, depending on the values of the preceding primitive parameters. This again confirms that the descriptive properties between the two managers' compensations are not sufficient to distinguish the two models.

Each manager gets her highest compensation, denoted by $w_i(\overline{x})$, when the informativeness of the corresponding output realization is highest, that is, $g_i(\overline{x})=0$ or $g_0(\overline{x})=0$. Furthermore, if the managers' efforts are observable to shareholders, $g_i(x)$ or $g_0(x)$ equals zero for any x. This is the first best scenario without information asymmetry on effort. Thus only the participation constraint is binding for each manager at his effort choice of working, and the shadow price of the incentive compatibility constraint drops. As a result, the optimal compensation equals $(1/\rho)\ln\alpha_{i2}$, which is the sufficient amount required to motivate manager i to work if his effort can be perfectly monitored by shareholders. Last, optimal compensation increases with the informativeness of the performance signal about working. While an output realization is more likely drawn from the distribution under which manager i works, that is, $g_i(x)$ or $g_0(x)$ is smaller, she gets higher compensation at that signal, keeping all other things constant.

It is important to emphasize the advantages of using a nonparametric approach placing no restrictions on the distributions of output. According to the optimal contract (8) and (9), strong assumptions on the functional form of the likelihood ratio $g_i(x)$ or $g_0(x)$ are required to legitimize a linear contract shape or, alternatively, an OLS analysis. For example, if the distributions belong to the exponential family, the contract shape can be approximated by a linear regression after log transformation. However, equity-based compensation including stock options is convex in stock returns in general, which violates the linearity assumption on contract shape.

2.6 Shareholders' Profit Maximization

As in the first step in Grossman and Hart [1983], shareholders maximize the expected net benefits by motivating both managers working rather than other effort pairs. Shareholders' benefit is the expected increase in the equity value of the firm in the contract period, which is calculated by multiplying the market value of the firm at the beginning of the period, as denoted by V in (1), with the gross abnormal return x and then taking the expectation over the distribution of x conditional on the two managers' effort choices in that period; that is, $E[Vx \mid j_1, j_2]$.

Shareholders' cost is the total compensation paid to the two managers. Denote w_i^s as the optimal fixed compensation paid to the manager i, ($i = 1, 2$), if shareholders merely wish to induce the manager to stay in the firm but allow him to shirk. The superscript s refers to shirking; w_i^s can be derived from an equation resembling a binding participation constraint at shirking. In that equation, on one side is the value of manager i's outside option, and on the other side is the manager i's expected CARA utility from a flat compensation w_i^s multiplied by his disutility coefficient of shirking (α_{i1}). Solving such an equation gives the optimal compensation to induce manager i to shirk as

$$w_i^s = \frac{1}{\rho} \ln \alpha_{i1}, \text{ for } i = 1, 2.$$

The optimal effort pair to be implemented in the two models is the same, that is, both managers work. However, the suboptimal benchmark effort pairs are different. In the *Individual Model*, the suboptimal effort pair is that no more than one manager works. By contrast, in the *Team Model*, there is only one benchmark effort pair, that is, both managers shirk.

2.7 Links to Theoretical Literature

The *Individual Model* is standard in theoretical literature on multiagent moral hazard and appears as the benchmark model in Itoh [1993], Arya et al. [1997], and Che and Yoo [2001], to contrast with their team models. It can be considered as a multiagent version of Holmstrom [1979], in the sense that one agent's participation constraint and incentive compatibility constraint, which are binding on their own, respectively, in Holmstrom [1979], now are conditional on the equilibrium effort choice of the other agent. In equilibrium, each agent takes working as the best response to the other working. The key difference in the *Team Model* is that the team incentive compatibility constraint replaces the individual incentive compatibility constraint. This feature is commonly shared by several theoretical models of team incentive. These models further develop various mechanisms to enforce the

Pareto optimal strategy, for example, the explicit side contracts without utility transfer in Itoh [1993], the finitely repeated game with implicit side contracts in Arya et al. [1997], and the infinitely repeated game with implicit side contracts in Che and Yoo [2001]. Though distinguishing these enforcement mechanisms is interesting, it is beyond the scope of a first-pass test, which this paper targets.[10] Instead, the *Team Model* in this paper assumes that managers choose Pareto optimality. Similarly in theoretical papers, simplification is made for the sake of research focus, for example, Friedman [2012] models the CEO and CFO as a syndicate without an explicit enforcement mechanism.

The setup in this paper distinguishes it from some other theoretical papers on multiagent moral hazard, due to the context and data availability. First, the single joint output assumption differs from mechanisms using agent-specific performance measures, such as relative performance evaluation (RPE; Holmstrom [1982]) and tournament (Lazear and Rosen [1981]), which stick to a restrictive contract form. These mechanisms encounter a sabotage problem. This counterproductive problem (Dye [1984]) may potentially explain why the RPE has not received much empirical support in the context of top management teams.[11] Besides, executive compensation is mostly equity based and tied to the overall firm performance, so that the single-output assumption seems to better capture top management team production. Second, the simultaneous move assumption differs from the sequential move assumed in Winter [2006, 2010]. A sequential move enlarges the information space and action space of players but seems to better describe stream-line workers rather than top managers who make joint decisions often. Third, the single-action assumption differs from the models of multiple tasks including reporting effort to upper levels in a hierarchy (Ma [1988]), reporting output (Friedman [2012], Indjejikian and Matejka [2009]), or helping

[10] Another type of team model allows agents to transfer utilities (Itoh [1993]). In an unreported test, I apply the same empirical procedure to a team model in which the setup is the same as the current one, except that the two agents can transfer utilities. This type of team model describes a "teamness" level higher than that of the current *Team Model* by treating managers as family members. Practically speaking, it may be less implementable in the context of top management teams. In my test, this type is rejected by the data.

[11] Gong et al. [2011] point out that the weak support of RPE since Antle and Smith [1986] may be due to data availability. However, their evidence from new information disclosed in proxy statements is limited to using RPE to incentivize CEOs across firms rather than top managers within one firm.

in addition to production (Itoh [1991], Ramakrishnan and Thakor [1991]). Testing these models apparently requests the information regarding externality between managers more than is available to this paper. For the same reason, I do not consider externality more than on output, for example, the synergy on effort cost modeled by Edmans et al. [2013].

3 Data

This section discusses the data source and the construction of key variables used in structural estimation. The sample covers S&P1500 firms from 1993 to 2005. The accounting data come from the COMPUSTAT North America database. The stock returns and market value are calculated from CRSP and Compustat PDE. The executive compensation comes from the ExecuComp database. I drop firm-year observations if the firm changed its fiscal year end such that all compensations and stock returns are 12-month based. A detailed description of variable construction is left to Appendix B.

3.1 Heterogeneity in the Data

This paper assumes that managers' preferences for effort and risk do not change after they accept the compensation contracts. However, managers with different preferences may sort into different types of firms. To account for the heterogeneity in the sample, firms are grouped by industrial sector, firm size, and capital structure.[12]

The whole sample is classified into three industrial sectors according to the Global Industry Classification Standard (GICS) code, denoting by S_{nt} the nth firm in year t. The primary sector ($S_{nt} = 1$) includes firms in energy, materials, industrials, and utilities. The consumer good sector ($S_{nt} = 2$) includes firms in consumer discretionary and consumer staples. The service sector ($S_{nt} = 3$) includes firms in health care, financial, and information technology and telecommunication services. In each industrial sector, firms are classified based on both

[12] The definition of firm types follows Gayle and Miller [2015] and Gayle et al. [2015b] in order to make the results more comparable.

the firm size, which is measured by the total assets on the balance sheet and denoted by A_{nt}, and the capital structure, which is measured by the debt-to-equity ratio and denoted by D/E_{nt}. Each of the two variables can have two values, that is, small (S) or large (L). If the beginning total assets of firm n in year t are below the median of total assets in its sector, $A_{nt} = S$; otherwise, $A_{nt} = L$. The same rule applies to D/E_{nt}. Firm type is denoted by $Z_{nt} = (A_{nt}, D/E_{nt})$ and includes four combinations of A_{nt} and D/E_{nt}.

Table 1 summarizes the firm characteristics cross-sectionally. As to the firm size, if compared based on book value (measured by the total assets on the balance sheet), firms in the consumer goods sector on average have significantly smaller book values than those in the other two sectors. If compared based on market value, the three sectors are not significantly different. As to capital structure measured by the debt-to-equity ratio, among the three sectiors that are significantly different from each other, the service sector has the highest mean and the biggest standard deviation.

3.2 Abnormal Stock Returns

For each firm in each fiscal year, I calculate a monthly compounded return adjusted for splitting and repurchasing and subtract the return to a value-weighted market portfolio (NYSE/NASDAQ/AMEX) from the compounded return to get the abnormal return for the corresponding fiscal year. The abnormal stock returns are summarized cross-sectionally in Table 2, conditional on firm size, capital structure, and industrial sector. They are all insignificantly different from zero, which is consistent with an underlying assumption that each type of firm faces a competitive market.

3.3 Compensation

Managers choose efforts based on overall wealth change implied by their compensation packages. Following Antle and Smith [1985], I construct a total compensation measure by adding wealth change from holding options and wealth change from holding stocks to all regular components provided by the ExecuComp database. These wealth changes can be interpreted as opportunity costs of holding firm-specific equity. This paper studies the two highest paid managers based on this total compensation measure.

Table 2 describes the two managers' compensation cross-sectionally. In all types of firms, the primary sector always provides the lowest compensation for both managers, and the service sector always provides the highest. In each sector, large firms offer higher compensation for both managers than small firms. As to the distribution of compensation conditional on capital structure, in the primary sector and the service sector, among firms of similar size, firms of high financial leverage (large debt-to-equity ratio) offer compensation no more than firms of low financial leverage. In the consumer goods sector, small firms have the same direction, but large firms go in the opposite direction. The variation across firm types together with the heterogeneity presented in Table 1 indicate that it seems necessary to test the robustness of a model by examining its explanatory power across different firm types.

Table 3 reports the position profiles of the two highest paid managers, who are classified into three categories:[13] (i) "Functional," if the manager is a CTO, CIO, COO, CFO, or CMO, but not any other; (ii) "General 1," if the manager is a chairman, president, CEO, or founder, but not any other; (iii) "General 2," if the manager is an executive vice-president, senior vice-president, vice-president, vice-chair, or other (defined in the database), but not any other. Also, there are four possible combinations if the manager holds multiple positions, including "Functional & General 1," "Functional & General 2," "General 1 & General 2," and "Functional & General 1 & General 2." The statistics reported in Table 3 show a similar distribution of positions across industrial sectors.

[13] The ExecuComp database provides the titles of executives. For missing values, a data set of executive positions used by Gayle et al. [2015a] is added.

4 Identification

This section establishes the identification of the two models laid out in Section 2. As is standard in the structural estimation literature, the identification is conducted under the assumption that, in reality, shareholders honor the contract and the two managers work, as the equilibrium predicts. Therefore the optimal compensation schemes and the distribution of the gross abnormal returns conditional on managers' equilibrium actions are assumed to be observable with only measurement errors and thus can be nonparametrically identified from the data. This paper assumes that the data of compensation and stock returns are cross-sectional, independent draws from the equilibrium of this model. The draws are repeated in each year $t = 1, ..., T$. The unobservable primitive elements to be identified include managers' preference parameters of risk and effort as well as the distribution of gross abnormal returns conditional on managers' off-equilibrium actions, which is pinned down to the likelihood ratio between the distribution of the gross abnormal returns off and on the equilibrium path because the on-equilibrium-path distribution can be identified from the data.

If we are only interested in estimating some sufficient statistics of a particular aspect of the economic model, for example, the pay-for-performance sensitivity given the primitive preference parameters fixed, a reduced form regression can accomplish this task. However, if we hope to test how well each entire model can rationalize the data of executive compensation and stock returns, to estimate the primitive parameters for future counterfactual analysis on contracting efficiency, or to arrive at policy implications that can only be made based on a particular model that fits reality, we need to go further to identify the models as a whole and estimate all the unobservable primitive elements (Matzkin [2007]).

4.1 Individual Model

The unobservable structural parameters in the *Individual Model* include each manager's effort preference over working and shirking relative to his outside option (denoted by α_{ij_i}, which is the time-independent effort disutility coefficient in manager i's utility functions when he chooses effort level j), the likelihood ratio of the distribution if manager i unilaterally shirks over that if both managers work in year t (denoted by $g_{it}(x_t)$, and the subscript t in x_t is dropped hereafter when it does not cause confusion), and the risk aversion parameter ρ. This paper shows that these unobservable parameters can be sequentially derived as mappings of the risk aversion parameter and the observables.

First, the disutility coefficients of working, that is, α_{i2} for $i = 1, 2$, can be identified from binding participation constraint. Shareholders design the optimal compensation such that, at the beginning of the period when managers decide whether to accept or reject the job offer, each manager is indifferent between rejecting the job to pursue an outside option and accepting the offer and working diligently during the following period. In the economic model, each manager's expected utility conditional on his subsequent effort choice (working) is equal to the value of his outside option.

Rearranging the terms of the participation constraints (5) when the equalities hold, we can find that only the disutility coefficients α_{i2} and the risk aversion parameter ρ are unknown. This indicates that if ρ can be identified, then α_{i2} can be expressed as a mapping of ρ and the observables. In this sense, α_{i2} are identified respectively for $i = 1, 2$ up to the risk aversion parameter as follows:

$$\alpha_{i2}(\rho) = E_t[v_{it}(x, \rho)]^{-1}. \tag{10}$$

Next, let's consider the likelihood ratios $g_{it}(x)$ for $i = 1, 2$. In the formula of optimal compensation (8), it is easy to check that the compensation reaches the highest value when the likelihood ratio equals zero. Consequently, assuming the data satisfy this restriction on the likelihood ratio, that is, $\lim_{x \to \infty} g_{it}(x) = 0$, then $\overline{w}_{it} \equiv w_{it}(\overline{x}_{it})$ satisfying $g_{it}(\overline{x}_{it}) = 0$ can be consistently estimated by the highest compensation. Now define the likelihood ratio $g_{it}(x, \rho)$ $(i = 1, 2)$ as a mapping of ρ and some quantities that can be calculated from the data-generating process:

$$g_{it}(x,\rho) = \frac{1/v_{it}(x,\rho) - 1/v_{it}(\overline{x}_i,\rho)}{E_t\left[1/v_{it}(x,\rho)\right] - 1/v_{it}(\overline{x}_i,\rho)}. \tag{11}$$

Note that the formula of $g_{it}(x,\rho)(i = 1,2)$ satisfies $E_t[g_{it}(x,\rho)] = 1$, which is required by the definition of the likelihood function, as well as $g_{it}(\overline{x}_i,\rho) = 0$, which is required by the model. Also, in the functional form of the likelihood ratios, the only unknown is the risk aversion parameter. This implies that these two ratios are identifiable up to the risk aversion parameter as well.

Then the disutility coefficients of shirking, that is, α_{i1} for $i = 1,2$, can be identified from the bind incentive compatibility constraint. The optimal compensation motivates each manager to exert effort in the shareholders' interests. In the economic model, the optimal compensation makes each manager's expected utility from working the same as his expected utility from shirking. In the econometric model, the data generated from this model satisfy the two equalities held in the incentive compatibility constraints (6) as well as the two equalities held in the participation constraints (5). These together help us derive the disutility coefficients $\alpha_{i1}(\rho)$ $(i = 1,2)$ as the mappings of the risk aversion parameter, as follows:

$$\alpha_{i1}(\rho) = E_t[v_{it}(x,\rho)g_{it}(x,\rho)]^{-1}. \tag{12}$$

In (12), for any known risk aversion parameter ρ, the shirking disutility coefficient α_{i1} is the only unknown in the equations, and thus it can be identified from the data along with the risk aversion parameter for $i = 1,2$.

Last, I consider the shadow price of each manager's incentive compatibility constraint in the Lagrangian formulation of the shareholders' cost minimization problem. Take manager 1, for example. I apply the property of the likelihood ratio $g_{1t}(\overline{x}_1) = 0$ in the formula of the optimal compensation $w^*_{1t}(x)$ in (8) and evaluate both sides at $\overline{x}_1$. Note that on the left-hand side of that formula, $w_{1t}(\overline{x}_1)$ can be identified and estimated by the highest compensation that manager 1 receives. That is, $w_{1t}(\overline{x}_1) = \max(w_{1t}(x_1))$. On the right-hand side, given that the disutility coefficients have been identified as previously and $g_{1t}(\overline{x}_1)$ drops off, only the shadow price μ^I_{1t} and the risk aversion parameter are left unknown. The same procedure applies to identifying the shadow price for manager 2 (μ^I_{2t}). Consequently, the two shadow prices can be expressed as the mappings of the risk aversion parameter, as follows:

$$\mu^I_{it}(\rho) = E_t\left[v_{it}(x,\rho)\right]/v_{it}(\overline{x},\rho) - 1. \tag{13}$$

Collectively, all parameters in the model can be recovered from the data-generating process, along with the risk aversion parameter.

Subsequently, I further explore other restrictions implied by the *Individual Model* to delimit the range of the risk aversion parameters. The first set of restrictions refers to shareholders' preferences on profit maximization. As assumed, the shareholders prefer motivating both managers to work to allowing either one or both of them to shirk. From the shareholders' viewpoint, the net profit of motivating a particular effort pair is the residual of the firm value growth deducted by the compensation cost. I calculate the shareholders' net benefit of motivating both managers to work and that of motivating no more than one manager to work, respectively. Those equilibrium restrictions imply that this difference should be nonnegative and constitute the following three inequalities in (14), (15), and (16). Λ_{1t} (Λ_{2t}) reflects that the shareholders' net benefit of motivating both managers to work is larger than that of having only manager 1 (2) shirk. By contrast, Λ_{3t} reflects that shareholders' net benefit is also larger than that of having both managers shirk:

$$\begin{aligned}
\Lambda_{1t}(\rho) &= E_t[V_t * x - w^*_{1t}(x) - w^*_{2t}(x)] - E_t[(V_t * x - w^s_{1t} - w^*_{2t}(x)) * g_{1t}(x,\rho)] \geq 0, && (14)\\
\Lambda_{2t}(\rho) &= E_t[V_t * x - w^*_{1t}(x) - w^*_{2t}(x)] - E_t[(V_t * x - w^*_{1t}(x) - w^s_{2t}) * g_{2t}(x,\rho)] \geq 0, && (15)\\
\Lambda_{3t}(\rho) &= E_t[V_t * x - w^*_{1t}(x) - w^*_{2t}(x)] - E_t[(V_t * x - w^s_{1t} - w^s_{2t}) * g_{1t}(x,\rho) * g_{2t}(x,\rho)] \\
&\geq 0, && (16)
\end{aligned}$$

where $w_{it}^{*}(x)$ is manager i's compensation if he works and is estimated from data and w_{it}^{s} is manager i's flat compensation to meet his outside option when shareholders prefer him shirking, that is,

$$w_{it}^{s} = \frac{1}{\rho} \ln \alpha_{i1}(\rho), \text{ for } i = 1, 2. \tag{17}$$

The second set of restrictions stems from the requirement that both managers working is the unique Nash equilibrium between the two managers. The incentive compatibility constraint has guaranteed that for each manager, shirking is not a best response to the other manager working such that the asymmetric effort pairs are ruled out from being a potential Nash equilibrium. To avoid "both managers shirk" being a Nash equilibrium in the subgame of the two managers, the optimal contract ensures that shirking is never a best response of one manager to the shirking of the other manager. In particular, manager 1's expected utility conditional on that he works but manager 2 shirks is higher than that conditional on both he and manager 2 shirking. The first term of the inequality in (18) ((19)) is manager 1 (2)'s expected utility conditional on that he works but manager 2 (1) shirks. The second term is manager 1 (2)'s expected utility conditional on both managers shirking. If the data are generated from this model, then the following two inequalities should hold:

$$\begin{aligned} \Lambda_{4t}(\rho) &= \{-\alpha_{12}(\rho) E_t[v_{1t}(x,\rho) g_{2t}(x,\rho)]\} \\ &- \{-\alpha_{11}(\rho) E_t\left[v_{1t}(x,\rho) g_{1t}(x,\rho) g_{2t}(x,\rho)\right]\} > 0, \end{aligned} \tag{18}$$

$$\begin{aligned} \Lambda_{5t}(\rho) &= \{-\alpha_{22}(\rho) E_t[v_{2t}(x,\rho) g_{1t}(x,\rho)]\} \\ &- \{-\alpha_{21}(\rho) E_t[v_{2t}(x,\rho) g_{1t}(x,\rho) g_{2t}(x,\rho)]\} > 0. \end{aligned} \tag{19}$$

The third source of equilibrium restrictions comes from the requirement that the likelihood ratio $g_{it}(x)$ be nonnegative. Recall the identification of $\overline{x}_i$, which is obtained by satisfying $g_{it}(\overline{x}_i) = 0$, $\forall\ x > \overline{x}_i$ $(i = 1, 2)$, where $g_{it}(x)$ in (11) is guaranteed to be nonnegative. However, the product $g_{1t}(x) g_{2t}(x)$ is another likelihood ratio such that the following restriction must be satisfied:

$$\Psi_{1t}(\rho) = E_t[g_{1t}(x, \rho) * g_{2t}(x, \rho)] = 1.$$

Collectively, the preceding restrictions implied by the *Individual Model* can be summarized by a function $Q_{\mathrm{I}}(\rho)$ as

$$Q_{\mathrm{I}}(\rho) \equiv \sum_{t=1}^{T} \left\{ \sum_{k=1}^{5} [\min(0, \Lambda_{kt}(\rho)]^2 + [\Psi_{1t}(\rho)]^2 \right\}.$$

Note that the $Q_{\mathrm{I}}(\rho)$ function has a distance-minimizing feature, which is the sum of two types of elements. The element corresponding to an equality restriction, that is, $\Psi_{1t}(\rho) = 0$, is the square of $\Psi_{1t}(\rho)$. The element corresponding to a nonnegative inequality restriction, that is, $\Lambda_{kt} > 0$, is the squared value of the minimum between Λ_{kt} and zero. As a result, $Q_{\mathrm{I}}(\rho)$ theoretically reaches zero if all restrictions implied by the model are satisfied. Thus, if a risk aversion parameter is admissible to the model, it belongs to the identified set[14] defined as

$$\Gamma_{\mathrm{I}} \equiv \{\rho > 0 : Q_{\mathrm{I}}(\rho) = 0\}. \tag{20}$$

4.2 Team Model

Compared with the same set of restrictions in the *Individual Model*, two key differences exist in the identification of the *Team Model*. First, shareholders are only concerned with symmetric deviation of managerial effort, so that each manager's compensation provides the same inference to back out the likelihood ratio $g_{0t}(x)$. Define $g_{0t}^{i}(x)$ $(i = 1, 2)$ as the likelihood ratio that can be identified from manager i's compensation scheme:

$$g_{0t}^{i}(x, \rho) = \frac{1/v_{it}(x, \rho) - 1/v_{it}(\overline{x}_i, \rho)}{E_t\left[1/v_{it}(x, \rho)\right] - 1/v_{it}(\overline{x}_i, \rho)}. \tag{21}$$

[14]The inclusion of inequality constraints prevents point identification.

The symmetric inference restriction implied by the *Team Model* leads to the following restriction, in which the two likelihood ratios are equal in unit mass:

$$\Psi_{2t}(\rho) = E_t[\mathbf{1}\{g_{0t}^1(x,\rho) = g_{0t}^2(x,\rho)\} - 1] \geq 0,$$

where $\mathbf{1}\{g_{0t}^1(x,\rho) = g_{0t}^2(x,\rho)\}$ is an index function equal to 1 if the condition is satisfied, and zero otherwise.[15]

Second, the suboptimal effort level to which the equilibrium effort is compared becomes both managers shirking. Shareholders prefer incentivizing both managers working to both shirking. This restriction can be reflected in the following inequality constraints:

$$\begin{aligned}
\Lambda_{6t}(\rho) &= E_t[V_t * x - w_{1t}^*(x) - w_{2t}^*(x)] - E_t[V_t * x * g_{0t}^1(x, w_1, w_2) - w_{1t}^s - w_{2t}^s] \geq 0, \\
\Lambda_{7t}(\rho) &= E_t[V_t * x - w_{1t}^*(x) - w_{2t}^*(x)] - E_t[V_t * x * g_{0t}^2(x, w_1, w_2) - w_{1t}^s - w_{2t}^s] \geq 0,
\end{aligned}$$

where the fixed compensation paid to both managers if the shareholders prefer them shirking (w_{it}^s) is the same as previously defined.

Collecting the restrictions implied by the *Team Model* as

$$Q_{\mathrm{T}}(\rho) \equiv \sum_{t=1}^{T} \left\{ \sum_{l=6}^{7} [\min(0, \Lambda_{lt}(\rho))]^2 + [\min(0, \Psi_{2t}(\rho))]^2 \right\},$$

I define Γ_{T}, a set of the risk aversion parameter admissible to this model, as

$$\Gamma_{\mathrm{T}} \equiv \{\rho > 0 : Q_{\mathrm{T}}(\rho) = 0\}. \tag{22}$$

Note that only binding constraints are used to develop the mappings from the risk aversion parameter to other primitive parameters. As a result, the likelihood ratio associated with unilateral shirking ($g_{it}(x)$, for $i = 1, 2$) cannot be identified. Inequality constraints can

[15] This function-wise restriction is constructed in a way similar to the nonnegative restriction on likelihood ratio imposed in Gayle and Miller [2015].

have empirical content only when all involved parameters can be identified. In this model, the binding constraints include the two participation constraints and the two incentive compatibility constraints, and both associate with symmetric effort choices only.

4.3 Summary of the Identification Results

In the previous subsections, the binding participation constraints and binding incentive compatibility constraints in each model helped us derive the mappings from the risk aversion parameter to the primitives in the model. The equilibrium restrictions customized to each model help us bound the risk aversion parameter with which the model can rationalize the data. The function $Q_{\mathrm{M}}(\rho)$ for each model M summarizes the equality and inequality restrictions in equilibrium, and it is a function of observables and the risk aversion parameter, which is the only unknown in the econometric model.

Given that all primitives introduced into the econometric model can be recovered from the data-generating process along with the risk aversion parameter, I denote the set of structural parameters for model $M \in \{$I (*Individual Model*), T (*Team Model*)$\}$ by

$$\begin{aligned}
\theta_{\mathrm{I}} &\equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_{1t}(x), g_{2t}(x)) \\
\theta_{\mathrm{T}} &\equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_{0t}(x));
\end{aligned}$$

then the following proposition formally states this result.

Proposition 2 *If the data are generated by one model M in the framework of this paper with true risk aversion parameter ρ^*, then θ_M^* can be identified from $(x_t, w_{it}, \overline{w}_{it})$ for $i = 1, 2$, that is, $\theta_M^* = \theta_M(\rho^*)$.*

5 Estimation and Tests

5.1 Hypotheses and Test Statistics

Intuitively, if the model $M \in \{\mathrm{I}\ (\textit{Individual Model}), \mathrm{T}\ (\textit{Team Model})\}$ can rationalize the data, there must exist some nonnegative values of the risk aversion parameter ρ such that the data restrictions summarized by the criterion function $Q_{\mathrm{M}}(\rho)$ are satisfied. In other words, the corresponding set Γ_M is nonempty. This intuition is the foundation of the estimation and tests in this section.

Recall that the $Q_{\mathrm{M}}(\rho)$ function has a distance-minimizing feature. Mathematically, if the data are generated by the model M, there must exist some nonnegative values of the risk aversion parameter ρ such that the population value $Q_{\mathrm{M}}(\rho)$ is zero. Following this logic, a subsampling algorithm is used to obtain a consistent estimate of the 95% confidence region of the risk aversion parameter that is admissible to the model. If the model is observationally equivalent to the data-generating process, this interval should not be empty. If the interval is empty, we can reject the null hypothesis that this model generates the data. Consequently, the estimated confidence region of the risk aversion parameter provides a criterion on whether the model is rejected. The estimation and the hypothesis test are accomplished at the same time.

For each model M, define the null hypothesis and alternative hypothesis as

$$
\begin{aligned}
H_0^M &: Q_{\mathrm{M}}(\rho) = 0 \text{ for some } \rho > 0 \text{, i.e., the model } M \text{ cannot be rejected} \\
H_A^M &: Q_{\mathrm{M}}(\rho) > 0 \text{ for all } \rho \text{, i.e., the model } M \text{ is rejected.}
\end{aligned}
$$

The sample analogue of $Q_{\mathrm{M}}(\rho)$ is calculated using data and the only unknown parameter (risk aversion) for each firm type Z in each sector S. The expectation in $Q_{\mathrm{M}}(\rho)$ is consistently estimated by an average weighted by the corresponding kernel densities.[16] Besides, in the

[16]The nonparametric estimation of optimal compensation and the density of gross abnormal returns can be found in Appendix C.

Individual Model, $v_{it}(\overline{x}_{it})$ is replaced with $exp\left(-\rho \overline{w}_{it}(\overline{x}_{it})\right)$, where $\overline{w}_{it} = \max\{w_{it}^1, ...w_{it}^{N_{ZS}}\}$. N_{ZS} is the number of firms of type Z in each sector S. In the *Team Model*, $v_{it}(\overline{x}_{it})$ is replaced with $exp\left(-\rho w_{it}(\overline{x}_t)\right)$, where $\overline{x}_t = \max\{\arg\max_x w_{1t}^n(x), \arg\max_x w_{2t}^n(x), \text{ for } n = 1, ..., N_{ZS}\}$. Formally, for each model $M \in \{\text{I } (\textit{Individual Model}), \text{ T } (\textit{Team Model})\}$,the sample analogue of $Q_{\text{M}}(\rho)$ is defined as

$$\begin{aligned} Q_{\text{I},ZS}^{(N)}(\rho) &\equiv \sum_{t=1993}^{2005} \left\{ \sum_{l=1}^{5} \left[\min(0, \Lambda_{it,ZS}^{(N)})\right]^2 + \left[\Psi_{1t,ZS}^{(N)}\right]^2 \right\} \\ Q_{\text{T},ZS}^{(N)}(\rho) &\equiv \sum_{t=1993}^{2005} \left\{ \sum_{l=6}^{7} \left[\min\left(0, \Lambda_{it,ZS}^{(N)}\right)\right]^2 + \left[\min\left(0, \Psi_{2t,ZS}^{(N)}\right)\right]^2 \right\}. \end{aligned}$$

Let us summarize the differences among the preceding two criterion functions. The suboptimal effort pair unfavorable to the shareholders is different between the *Individual Model* and the *Team Model* such that the restrictions corresponding to the shareholders' profit maximization are $\Lambda_{lt,ZS}^{(N)}$ $(l = 1, 2, 3)$ in the criterion function of the *Individual Model* but $\Lambda_{lt,ZS}^{(N)}$ $(l = 6, 7)$ in the *Team Model*. The restriction on the uniqueness of Nash equilibrium is only required by the *Individual Model*, so $Q_{I,ZS}^{(N)}(\rho)$ includes two unique terms $\Lambda_{lt,ZS}^{(N)}$ $(l = 4, 5)$. The restrictions on the likelihood ratios generate the term $\Psi_{1t,ZS}^{(N)}$ in the *Individual Model* to guarantee that the likelihood ratio associated with both managers shirking satisfies the integral-to-one property. In the *Team Model*, the symmetric inference of the likelihood ratio requires that the two likelihood ratios identified separately from the two managers' compensation schemes be equal with unit mass, which gives the last restriction, denoted by $\Psi_{2t,ZS}^{(N)}$.

The hypothesis test on each model M is based on the confidence region of the risk aversion parameter by which each model can be indexed. The intuition is that if the data are generated from a process observationally equivalent to one model with some values of the risk aversion parameter admissible to this model, then the corresponding criterion function $Q_{M,ZS}^{(N)}(\rho)$, which is evaluated by the observed data at a fixed risk aversion parameter belonging to the identified set, should be close enough to zero because of its distance-minimizing feature. By contrast, if that model cannot rationalize the data, then at least one of those restrictions summarized by the criterion function must be violated. Such violation makes the test statistic, that is, the criterion function multiplied by its asymptotic convergence rate, go to infinity as the sample size N goes to infinity. Consequently, if there do not exist positive values of the risk aversion parameter that, together with the observed data, can make the value of the test statistic small enough, the model should be rejected. Define the 95% confidence region of the identified set of the risk aversion parameter under model M in firm type Z and sector S as

$$\Gamma_{M,ZS}^{(N)} \equiv \{\rho > 0 : N_{ZS}^{a} * Q_{M,ZS}^{(N)}(\rho) \leq c_{95,ZS}^{M}\},$$

where N_{ZS}^{a} is the asymptotic convergence rate of $Q_{M,ZS}^{(N)}(\rho)$ with $a = 2/3$ and where $c_{95,ZS}^{M}$ is the 95% critical value of the test statistic; $c_{95,ZS}^{M}$ can be consistently estimated by the subsampling algorithm used in Gayle et al. [2015b]. Consequently, if the set $\Gamma_{M,ZS}^{(N)}$ is empty, I reject the model M for firm type Z in sector S. If it is not empty, I obtain the 95% confidence region of the risk aversion parameter set.

5.2 A Dynamic Variation

So far, managers' outside options have been assumed to be constant over time. However, managers' alternative career opportunities may fluctuate with the macroeconomy. Top managers may lose their jobs or receive shrunken compensation packages in recession years. Also, top managers studied by this paper are in late middle age on average, such that when they make effort choices, they may take into account consumption smoothing over the rest of their lives. Given these factors, a natural extension of the static models is a dynamic version that addresses the preceding two considerations.

To convert, the effort-dependent utility function defined in (3) now has a new expression:

$$-\alpha_{ij}{}^{\frac{1}{b_t-1}} E_t \left[\exp\left(\frac{-\rho w_{it}(x_t)}{b_{t+1}} \right) \mid j,k \right], \tag{23}$$

where b_t is the bond price in year t, which pays a unit of consumption per period forever.[17] Intuitively, now a manager consumes the interest of the bond purchased with the compensation in each period, that is, $w_{it}(x_t)/b_{t+1}$. This reflects her lifetime consumption smoothing. Also, the cash certainty equivalent of the nonpecuniary benefit of effort is deferred one more period to match the timing of compensation. It was $(1/\rho)\ln\alpha_{ij}$ in the static model as (17), but now it is $[b_{t+1}/\rho(b_t-1)]\ln\alpha_{ij}$ in the dynamic version.

The empirical implementation in this paper adopts the dynamic version of the two models to estimate the primitive parameters and to test each model. The utility function in participation constraints and incentive compatibility constraints in the static models are replaced with (23). The results are reported in the next section.

6 Results

6.1 Estimation of the Risk Aversion Parameter and Tests

Table 4 reports the estimates of the risk aversion parameter under each model by firm type and sector as well as its economic meaning in terms of a certainty equivalent value of a gamble.[18] The two panels in the table correspond to the two models. The column "Risk Aversion" reports the 95% confidence region of the identified set of the risk aversion parameter, where empty parentheses mean an empty set. The column "Certainty Equivalent" reports the amount that a manager would like to pay to avoid a gamble with equal chance to win or lose \$1 million given his coefficient of absolute risk aversion equal to the corresponding

[17] See the detailed construction of the bond prices in Gayle and Miller [2009, pp. 1748–1749].

[18] For a manager with risk aversion parameter ρ, the expected utility from a gamble with half chance to win or lose \$1 million is $EU = 0.5*\exp(-\rho*(-1/b)) + 0.5*\exp(-\rho*1/b)$, where b is the mean of the bond prices in the sample period (16.65). Thus the certainty equivalent to this gamble is $CE = \frac{-b}{\rho}\ln EU$.

value in the column “Risk Aversion."

A comparison of confidence regions between the two models shows that the level of the estimated risk aversion parameter is higher under the *Individual Model* than under the *Team Model*. Note that for the same industrial sector and firm type, whenever, between the *Individual Model* and the *Team Model*, the confidence regions are not perfectly overlapped, the *Team Model* always covers the lower range of the nonoverlapped interval, indicating that to rationalize the currently studied data of stock returns and executive compensation, this model has to go with less risk-averse managers.

To examine how sensitive the robustness of the model specification test is to the assumption on homogeneous risk preferences, I strengthen this assumption gradually. Take the *Individual Model* in panel A of Table 4 as an example. First, I assume managers' risk preferences can vary with capital structure but stay the same among firms of similar size. The column “Homogenous within Size" reports the confidence region overlapped among firms that fall into the same size category. In the primary sector, the common interval for small-size firms is (12.75, 16.25), which is the overlapped interval between (12.75, 26.38) of small-size and small-debt-to-equity-ratio firms and (0.89, 16.25) of small-size and large-debt-to-equity-ratio firms. Similar analysis applies to the large-size firms and to other sectors.

Then, I further strengthen the assumption on homogeneous risk preference by assuming that managers in the same sector have the same magnitude of risk aversion. This assumption makes it impossible to find an overlapped confidence region within either the primary or the consumer goods sector. This indicates a rejection of the model in these two sectors if managers' risk attitudes are not sensitive to firm-level characteristics. Only the service sector survives this level of homogeneity by presenting a common confidence region regardless of firm size and capital structure, which covers a range of (4.83, 7.85).

However, if managers' risk preferences cannot vary with industrial sector, firm size, or capital structure, then the last column, “Homogeneous across Sectors," shows that there is no common interval of the confidence regions of the risk aversion parameter, which means that the *Individual Model* would be rejected if such an amount of homogeneity in managers' risk preferences were to exist in the data. In panel B for the *Team Model*, I do the same analysis and report the common confidence regions subject to different levels of homogeneity of managers' risk preferences.

To sum up the main results, the *Individual Model* cannot be rejected in any type of firm if managers' risk preferences differ across firm types. This model can rationalize the data with managers who have heterogeneous risk preferences and are relatively more risk averse. If homogeneous risk preferences are assumed regardless of firm type, the *Individual Model* cannot be rejected only in the service sector, which accommodates many firms in the financial industry. However, if the homogeneity in risk preferences is assumed across industrial sectors, there is no common interval of the confidence regions of the risk aversion parameter. This means that the *Individual Model* is rejected if the managers are assumed to have homogeneous risk preferences.

The *Team Model* can rationalize the data in all types of firms with less risk-averse managers. What's more, when the homogeneous risk aversion assumption is put on data, this model survives up to the most restrictive case. There is a common confidence region sitting across all firm types and industrial sectors in the sample. As a result, the *Team Model* is more robust than the *Individual Model* in explaining the observed executive compensation of top management teams.

6.2 Discussion

Three issues are worth further discussion. The first issue is to compare the level of the estimated risk aversion between my two models. Note that the estimated risk aversion parameter under the *Individual Model* is in general higher than that under the *Team Model*. A higher parameter value corresponds to the managers who are more risk averse.

To understand this difference, a comparable result is found in the comparison of different models by Gayle and Miller [2015]. In their paper, the estimated risk aversion under a two-states pure moral hazard model is less than that under a hybrid moral hazard model in which the CEO has private information about the firm's states and shareholders pay a premium to induce truthful report. To explain their difference, the pure moral hazard model requires that managers' expected utilities be equalized across states (no premium for reporting truth), such that the estimated risk aversion needs to be low enough to mitigate the variation in CEOs' compensation between states.

A similar consequence of model fitting emerges in this paper. In the *Team Model*, the suboptimal effort choice in the binding incentive compatibility constraint is both shirking. This implies that the variation in the two managers' expected utilities is attributed to their different effort disutility coefficients only, rather than additionally to the differentiable informativeness of performance measure (captured by the likelihood ratios) as the *Individual Model* allows. As a result, to fit the data, the *Team Model* accommodates less risk-averse managers.

The second issue is to compare my estimates of risk aversion with those in previous papers on executive compensation covering similar sample periods. The papers are comparable in terms of the certainty equivalent amount of a gamble with half chance to win or lose $1 million. The estimated risk aversion under my *Individual Model* is close to what Gayle and Miller [2009] and Gayle et al. [2015a] find using a more parameterized approach. The point estimate of the risk aversion parameter in Gayle and Miller [2009] generates that certainty equivalent amount at $248,620, and Gayle et al. [2015a] gives $255,199. My *Individual Model* covers these two values in 8 out of 12 cases in Table 4.

The estimated risk aversion under the *Team Model* is less, but it is still at a reasonable level close to the estimates in Gayle and Miller [2015], whose approach this paper follows. Their paper reports the 95% confidence region of that certainty equivalent amount at [$8,849, $92,390] under their pure moral hazard model and as [$160,870, $181,710] under their hybrid moral hazard model. My estimated common region of the risk aversion parameter under the *Team Model* gives [$7,000, $75,000].

The third issue is to contrast my estimates based on structural models whose primitives are unknown against the predictions from theoretical models built on fixed primitives (presumably known). A potential confusion could arise as follows. At a fixed level of risk aversion, the incentive provided by a team contract is muted owing to a less restrictive incentive compatibility constraint. Theoretically, this implies that the compensation scheme under a team contract tends to be flatter. If the *Team Model* and the *Individual Model* both have nonempty confidence regions (assuming heterogeneous risk preference), shareholders seem to have adopted a compensation scheme unnecessarily steeper than they are supposed to use under a team contract. This tends to suggest a rejection of the *Team Model*.

However, in the structual estimation, managers' risk aversion is not known and may actually be lower than the level that the *Individual Model* requires to explain the data. Then, as theory predicts, there is more demand of incentive in a team contract, and thus a steeper compensation scheme is needed for less risk-averse managers. The structural estimation of the *Team Model* fits this scenario, and thus the *Team Model* can rationalize the data.[19]

7 Conclusion

This paper identifies and tests two competing structural models that are explicitly based on theoretical models of principal-multiagent moral hazard. The two models are intended to capture a crucial consideration in shareholders' optimal compensation design, that is, whether they incentivize the members of a top management team indeed as a team or only as separate individuals.

For each model, the equilibrium restrictions are exploited to delimit the identified set of the risk aversion parameters to which all other primitive parameters in the same model can be indexed. The confidence region of the identified set is the basis of the model specification test. The set identification method enables this paper to examine a richer set of equilibrium

[19] Appendix D sets up a binary output example to illustrate how the risk aversion parameter (ρ) and the information structure ($f(x)$ and $h(x)$) interact in the estimation to reconcile with the curvature of the compensation schemes.

restrictions by incorporating both equality and inequality moment conditions into the criterion functions of the tests. Besides, the nonparametric technique can, to a certain extent, alleviate concern about overusing auxiliary assumptions. This concern generally applies to structural modeling papers.

To analyze the results of the hypothesis tests and draw conclusions, this paper needs to delve into a discussion of the assumption of homogeneity of managers' risk preferences. Under the *Individual Model*, the identified sets are not empty, but they do not overlap across firm types and industrial sectors. To reconcile this model with the data, it is necessary to assume that managers' risk preferences vary with firm size, capital structure, and industrial sector. Although it is likely that top managers in general have a different risk attitude from other groups of people, it is unclear to what extent they are distinguishable among themselves in terms of risk aversion based on the characteristics of their employers. In contrast, the *Team Model* predicts a common range of risk aversion across all firms. This model cannot be rejected even with the most stringent assumption that the managers have homogenous risk preferences across all types of firms and industrial sectors. Therefore this model has more robust explanatory power for the observed correlation between the executive compensation of top management teams and stock returns.

A Proofs

Proof of Proposition 1. In the *Individual Model*, shareholders' cost minimization problem can be summarized into the Lagrangian as

$$\begin{aligned}
L &= E\left[\ln v_1(x) + \ln v_2(x)\right] \\
&\quad -\mu_1\left[\alpha_{12}E\left[v_1(x)\right] - \alpha_{11}E\left[v_1(x)g_1(x)\right]\right] \\
&\quad -\mu_2\left[\alpha_{22}E\left[v_2(x)\right] - \alpha_{21}E\left[v_2(x)g_2(x)\right]\right] \\
&\quad -\mu_3\left[\alpha_{12}E\left[v_1(x)\right] - 1\right] \\
&\quad -\mu_4\left[\alpha_{22}E\left[v_2(x)\right] - 1\right].
\end{aligned} \tag{24}$$

Take the derivative of the proceeding Lagrangian with respect to (w.r.t., hereafter) $v_i(x)$. We get the first-order condition (FOC) w.r.t. $v_1(x)$ as

$$1/v_1(x) = (\mu_1^I + \mu_3^I)\alpha_{12} - \mu_1^I\alpha_{11}g_1(x). \tag{25}$$

Multiply both sides with $v_1(x)$, take expectation over $f(x)$, and, using the binding participation constraint and incentive compatibility constraint for manager 1, we get

$$\mu_3^I = 1.$$

Since $v_i(x) \equiv \exp\left(-\rho w_i(x)\right)$, $i = 1, 2$, rearranging items in the proceeding FOC gives

$$\begin{aligned}
1/v_1(x) &= \exp\left(\rho w_1(x)\right) = \alpha_{12}\left[1 + \mu_1^I - \mu_1^I\frac{\alpha_{11}}{\alpha_{12}}g_1(x)\right] \\
w_1^*(x) &= \frac{1}{\rho}\ln\alpha_{12} + \frac{1}{\rho}\ln\left[1 + \mu_1^I - \mu_1^I\left(\frac{\alpha_{11}}{\alpha_{12}}\right)g_1(x)\right]
\end{aligned}$$

where μ_1^I satisfies the binding incentive compatibility constraint, such that it uniquely solves

$$\begin{aligned}
\alpha_{12}E\left[v_1(x)\right] &= \alpha_{11}E\left[v_1(x)g_1(x)\right] \\
\alpha_{12}E\left[\frac{1}{\alpha_{12}\left[1 + \mu_1^I - \mu_1^I\frac{\alpha_{11}}{\alpha_{12}}g_1(x)\right]}\right] &= \alpha_{11}E\left[\frac{g_1(x)}{\alpha_{12}\left[1 + \mu_1^I - \mu_1^I\frac{\alpha_{11}}{\alpha_{12}}g_1(x)\right]}\right] \\
E\left[\frac{\alpha_{12}}{1 + \mu_1^I - \mu_1^I\frac{\alpha_{11}}{\alpha_{12}}g_1(x)}\right] &= E\left[\frac{\alpha_{11}g_1(x)}{1 + \mu_1^I - \mu_1^I\frac{\alpha_{11}}{\alpha_{12}}g_1(x)}\right]
\end{aligned}$$

Similarly for manager 2,

$$\begin{aligned}
\mu_4^I &= 1 \\
w_2^*(x) &= \frac{1}{\rho}\ln\alpha_{22} + \frac{1}{\rho}\ln\left[1 + \mu_2^I - \mu_2^I\left(\frac{\alpha_{21}}{\alpha_{22}}\right)g_2(x)\right]
\end{aligned}$$

where μ_2^I uniquely solves

$$E\left[\frac{\alpha_{22}}{1+\mu_2^I-\mu_2^I\frac{\alpha_{21}}{\alpha_{22}}g_2(x)}\right]=E\left[\frac{\alpha_{21}g_2(x)}{1+\mu_2^I-\mu_2^I\frac{\alpha_{21}}{\alpha_{22}}g_2(x)}\right]$$

In the *Team Model*, shareholders' cost minimization problem can be summarized into the Lagrangian as

$$\begin{aligned} L &= E\left[\ln v_1(x)+\ln v_2(x)\right] \\ &\quad -\mu_1\left[\alpha_{12}E\left[v_1(x)\right]-\alpha_{11}E\left[v_1(x)g_0(x)\right]\right] \\ &\quad -\mu_2\left[\alpha_{22}E\left[v_2(x)\right]-\alpha_{21}E\left[v_2(x)g_0(x)\right]\right] \\ &\quad -\mu_3\left[\alpha_{12}E\left[v_1(x)\right]-1\right] \\ &\quad -\mu_4\left[\alpha_{22}E\left[v_2(x)\right]-1\right]. \end{aligned} \tag{26}$$

Note that the only difference from the *Individual Model* is that the likelihood ratios in the incentive compatibility constraints are replaced with $g_0(x)$. As a result, the optimal compensation in the *Team Model* differs from that in the *Individual Model* by the likelihood ratio and shadow price μ_i^T, which uniquely solves

$$E\left[\frac{\alpha_{i2}}{1+\mu_i^T-\mu_i^T\frac{\alpha_{i1}}{\alpha i}g_0(x)}\right]=E\left[\frac{\alpha_{i1}g_0(x)}{1+\mu_i^T-\mu_i^T\frac{\alpha_{i1}}{\alpha_{i2}}g_0(x)}\right].$$

■

Proof of Proposition 2.

Individual Model

We want to show that $\theta^*=\theta(\rho^*)$. Suppose ρ is known. Write down the Lagrangian as

$$\begin{aligned} L &= E\left[\ln v_{1t}(x)+\ln v_{2t}(x)\right] \\ &\quad -\mu_1\left[\alpha_{12}E_t\left[v_{1t}(x)\right]-\alpha_{11}E_t\left[v_{1t}(x)g_{1t}(x)\right]\right] \\ &\quad -\mu_2\left[\alpha_{22}E_t\left[v_{2t}(x)\right]-\alpha_{21}E_t\left[v_{2t}(x)g_{2t}(x)\right]\right] \\ &\quad -\mu_3\left[\alpha_{12}E_t\left[v_{1t}(x)\right]-1\right] \\ &\quad -\mu_4\left[\alpha_{22}E_t\left[v_{2t}(x)\right]-1\right]. \end{aligned} \tag{27}$$

Take the derivative of the proceeding Lagrangian w.r.t. $v_{it}(x)$. We get the FOC w.r.t. $v_{1t}(x)$ as

$$1/v_{1t}(x)=(\mu_1+\mu_3)\alpha_{12}-\mu_1\alpha_{11}g_{1t}(x). \tag{28}$$

Similarly, the FOC w.r.t. $v_{2t}(x)$ is

$$1/v_{2t}(x)=(\mu_2+\mu_4)\alpha_{22}-\mu_2\alpha_{21}g_{2t}(x). \tag{29}$$

Evaluate the FOCs at the threshold values of shirking distribution, that is, $g_{1t}(\overline{x}_1)=$

$g_{2t}(\overline{x}_2) = 0$. Respectively, we get

$$\begin{aligned} 1/v_{1t}(\overline{x}_1) &= (\mu_1 + \mu_3)\alpha_{12}, & (30) \\ 1/v_{2t}(\overline{x}_2) &= (\mu_2 + \mu_4)\alpha_{22}. & (31) \end{aligned}$$

Take the expectation of the FOCs over the distribution with both diligent managers to get

$$\begin{aligned} E_t\left[1/v_{1t}(x)\right] &= (\mu_1 + \mu_3)\alpha_{12} - \mu_1\alpha_{11}, & (32) \\ E_t\left[1/v_{2t}(x)\right] &= (\mu_2 + \mu_4)\alpha_{22} - \mu_2\alpha_{21}. & (33) \end{aligned}$$

The binding participation constraint for each manager gives

$$\alpha_{12}^* = E_t[v_{1t}(x)]^{-1}, \tag{34}$$

$$\alpha_{22}^* = E_t[v_{2t}(x)]^{-1}. \tag{35}$$

The binding incentive compatibility constraint gives

$$\alpha_{11}E_t\left[v_{1t}(x)g_{1t}(x)\right] = \alpha_{21}E_t\left[v_{2t}(x)g_{2t}(x)\right] = 1. \tag{36}$$

Multiply both sides of (28) by $v_{1t}(x)$ and integrate over $f(x)$; it follows that

$$1 = (\mu_1 + \mu_3)\alpha_{12}E_t\left[v_{1t}(x)\right] - \mu_1\alpha_{11}E_t\left[v_{1t}(x)g_{1t}(x)\right],$$

and plugging (34) and (36) into the preceding, it follows that

$$\mu_3^* = 1.$$

Multiply both sides of (29) by $v_{2t}(x)$ and integrate over $f(x)$; it follows that

$$1 = (\mu_2 + \mu_4)\alpha_{22}E_t\left[v_{2t}(x)\right] - \mu_2\alpha_{21}E_t\left[v_{2t}(x)g_{2t}(x)\right],$$

and plugging (35) and (36) into the preceding, it follows that

$$\mu_4^* = 1.$$

Multiplying (30) by $E_t\left[v_{1t}(x)\right]$ and using $\mu_3 = 1$ and (34), it follows that

$$\begin{aligned} E_t\left[v_{1t}(x)\right]/v_{1t}(\overline{x}_1) &= \mu_1 + \mu_3, \\ \text{thus } \mu_1^* &= E_t\left[v_{1t}(x)\right]/v_{1t}(\overline{x}_1) - 1. \end{aligned}$$

Similarly, multiplying (31) by $E_t\left[v_{2t}(x)\right]$ and using $\mu_4 = 1$ and (35) it follows that

$$\begin{aligned} E_t\left[v_{2t}(x)\right]/v_{2t}(\overline{x}_2) &= \mu_2 + \mu_4, \\ \text{thus } \mu_2^* &= E_t\left[v_{2t}(x)\right]/v_{2t}(\overline{x}_2) - 1. \end{aligned}$$

Equations (30) and (32) together give

$$1/v_{1t}(\overline{x}_1) - E\left[1/v_{1t}(x)\right] = \mu_1\alpha_{11}.$$

Equations (28) and (30) together give

$$1/v_{1t}(\overline{x}_1) - 1/v_{1t}(x) = \mu_1\alpha_{11}g_{1t}(x).$$

Such that $g_{1t}(x)$ is obtained as

$$g_{1t}^*(x) = \frac{1/v_{1t}(x) - 1/v_{1t}(\overline{x}_1)}{E\left[1/v_{1t}(x)\right] - 1/v_{1t}(\overline{x}_1)}.$$

Similarly, we get $g_{2t}(x)$ as

$$g_{2t}^*(x) = \frac{1/v_{2t}(x) - 1/v_{2t}(\overline{x}_2)}{E\left[1/v_{2t}(x)\right] - 1/v_{2t}(\overline{x}_2)}.$$

Plugging into (36), it follows that

$$\begin{aligned}
\alpha_{11}^* &= \left[\frac{E_t[v_{1t}(x)] - v_{1t}(\overline{x}_1)}{1 - v_{1t}(\overline{x}_1)E[1/v_{1t}(x)]}\right]^{-1}, \\
\alpha_{21}^* &= \left[\frac{E_t[v_{2t}(x)] - v_{2t}(\overline{x}_2)}{1 - v_{2t}(\overline{x}_2)E[1/v_{2t}(x)]}\right]^{-1}.
\end{aligned}$$

Team Model

See the proof for the *Individual Model*. The only difference is that the likelihood ratios with respect to each manager are the same, that is, $g_{0t}^1(x) = g_{0t}^2(x)$, which introduces an additional restriction on data.

$$\begin{aligned}
g_{0t}^1(x) &= \frac{1/v_{1t}(x) - 1/v_{1t}(\overline{x}_1)}{E\left[1/v_{1t}(x)\right] - 1/v_{1t}(\overline{x}_1)}, \\
g_{0t}^2(x) &= \frac{1/v_{2t}(x) - 1/v_{2t}(\overline{x}_2)}{E\left[1/v_{2t}(x)\right] - 1/v_{2t}(\overline{x}_2)}.
\end{aligned}$$

■

B Data construction details

Firm heterogeneity: Firm heterogeneity is described by industrial sector and firm characteristics for each firm in each year. First, the whole sample is classified into three industrial sectors according to the Global Industry Classification Standard (GICS) code obtained from COMPUSTAT. The primary sector includes firms in energy (GICS: 1010), materials (GICS: 1510), industrials (GICS: 2010, 2020, 2030), or utilities (GICS: 5510). The consumer goods sector includes firms in consumer discretionary (GICS: 2510, 2520, 2530, 2540, 2550) or consumer staples (GICS: 3010, 3020, 3030). The service sector includes firms in health care (GICS: 3510, 3520), financial (GICS: 4010, 4020, 4030, 4040), or information technology and telecommunication services (GICS: 4510, 4520, 5010). Firms that appear in different sectors over the sample period are classified into the sector in which they appear most frequently.

Second, firm characteristics include two aspects (firm size and capital structure). The firm size is measured by the total assets on the balance sheet (AT, COMPUSTAT variable name in bracket hereafter). The capital structure is captured by the debt-to-equity ratio. The numerator of the ratio is the total liabilities (LT) and the denominator is the total common equity (CEQ). The market value equals the stock price at the end of fiscal year (PRCCF) times the number of outstanding common shares (SHRSOUT) plus the liquidating value of preferred stock (PSTKL, if available).The book values of asset, liability, and equity and the market values are deflated to the base year 2006. The deflator is calculated using Consumer Price Index for All Urban Consumers (all items with seasonal adjustment), which was obtained from the website of the U.S. Department of Labor (Bureau of Labor Statistics).

Abnormal Stock Returns: I collect raw stock prices and adjustment factors from the CRSP and COMPUSTAT/PDE. For each firm in the sample, I calculate monthly compounded returns adjusted for splitting and repurchasing for each fiscal year and subtract return to a value-weighted market portfolio (NYSE/NASDAQ/AMEX) from this raw return to get the raw abnormal return for its corresponding fiscal year. I drop firm-year observations if the firm changed its fiscal year end, such that all compensations and stock returns are 12-month based and consequently comparable with each other.

Compensation: Following the concept of income equivalent total compensation adopted by Antle and Smith [1985, 1986], Hall and Liebman [1998], and Margiotta and Miller [2000], the total compensation is constructed by adding change in wealth from options held and change in wealth from stocks held to the other components of compensation included in COMPUSTAT ExecComp.

Owing to the data availability, for each sample year, I cannot observe all the inputs of the Black–Scholes formula for grants carried from years before 1993, the beginning year of our sample. COMPUSTAT ExecComp only provides the valuation information for those options newly granted after 1993, including the number of underlying stock shares, exercise prices, expiration dates, and issue dates. However, I need to know these Black–Scholes inputs for options granted before year 1993 to completely value the wealth change of managers by estimating the value of unexercised options and updating it each year. To facilitate the calculation, I make the following assumptions: (i) all options are not exercised until expiration dates, (ii) stock options granted before 1993 are exercised in a FIFO fashion, and (iii) each manager holds his own stock options granted before 1993 for a period of the average length of holding across all years when he is in the sample. Consequently, I can back out the issue dates and exercised prices for options granted before 1993 for each manager. The same routines apply to those nonzero options granted before the manager entered the sample. Then I apply the dividend-adjusted Black–Scholes formula to reevaluate the managers' call options for each manager in each year. The dividend-adjusted Black–Scholes formula used is as follows. Let c denotes the call option value, K the exercise price, T_m the time to maturity (in years), S the underlying security price, q the dividend yield, r the risk–free rate, and σ the implied volatility. Let $N(\cdot)$ denote the standard normal cumulative distribution function. Then the call option value is given by

$$c = Se^{-qT_m}N(d_1) - Ke^{-rT_m}N(d_2), \tag{37}$$

$$d_1 = \frac{\ln(S/K) + (r - q + \sigma^2/2)T_m}{\sigma\sqrt{T_m}}, \tag{38}$$

and

$$d_2 = d_1 - \sigma\sqrt{T_m}. \tag{39}$$

Bond prices In the dynamic version, b_t is denoted as 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. The price of this bond was derived using the method described in Gayle and Miller [2009, pp. 1748–1749]. The term structure of interest rates underlying the bond price series was constructed from data on Treasury bills of varying maturities and were obtained from the Board of Governors website (http://www.bog.frb.fed.us/).

C Nonparametric Estimation of Compensation and the Probability Density Function of Gross Abnormal Returns in Equilibrium

I calculate the gross abnormal returns by

$$x_{nt} \equiv \widetilde{x}_{nt} + \frac{\widetilde{w}_{1nt}}{V_{n,t-1}} + \frac{\widetilde{w}_{2nt}}{V_{n,t-1}},$$

where $V_{n,t-1}$ is the market value of firm n at the end of year $t-1$, $\widetilde{x}_{nt}$ represents the observed abnormal returns, and $\widetilde{w}_{imt}$ is manager i's total compensation observed in firm n in year t.

(Z_n, S_n) are firm type variables for firm n based on firm size, capital structure, and industrial sector, as defined in the main text. For each firm type Z (four types in total) and each sector S (three sectors in total), I nonparametrically estimate the optimal compensation conditional on a set of grid points of bond price b_τ using the following kernel regression (Pagan and Ullah [1999])[20]:

$$\begin{aligned} w_{in\tau} &\equiv E_\tau[\widetilde{w}_{int}|x_{nt}, b_\tau] \\ &= \frac{\sum_{m=1}^{N} \widetilde{w}_{imt} K\left(\frac{x_{mt}-x_{nt}}{h_x}, \frac{b_{mt}-b_\tau}{h_b}\right)}{\sum_{m=1}^{N} K\left(\frac{x_{mt}-x_{nt}}{h_x}, \frac{b_{mt}-b_\tau}{h_b}\right)}, \end{aligned}$$

where N is the number of observations belonging to firm type Z and sector S. Then the density of gross abnormal return x_{nt} is nonparametrically estimated by a kernel estimator:

$$f(x_{nt}|Z,S) = \frac{\sum_{m=1}^{N} K\left(\frac{x_{mt}-x_{nt}}{h_x}\right)}{N}.$$

D A Binary Example

I use a binary output example to illustrate how the risk aversion parameter (ρ) and the information structure ($f(x)$ and $h(x)$) interact in the estimation to reconcile with the curvature of the compensation schemes. Each manager $i = 1, 2$ has two effort options $j \in \{1 = shirk, 2 = work\}$ and two outputs, either high or low, $x \in \{x_H, x_L\}$. The pay schedule is defined as $w(x_k)$ for $k = H, L$. The following table gives the conditional probability $prob(x|j)$, that is, $f(x)$ or $f(x)h(x)$ in the continuous case. In particular, $p \equiv prob(x|work)$ and $q \equiv prob(x|shirk)$; subscripts correspond to *Individual Model* (I) or *Team Model* (T).

Model	Both Models	Individual Model	Team Model
effort pair	i work, $-i$ work	i shirk, $-i$ **work**	i shirk, $-i$ **shirk**
x_H	p	q_I $(< p)$	q_T $(< p)$
x_L	$1-p$	$1-q_I$	$1-q_T$

The CARA utility function of manager i is specified as $-\alpha_{i1}e^{-\rho w(x)}$ if manager i shirks and as $-\alpha_{i2}e^{-\rho w(x)}$ if manager i works, for $x \in \{x_H, x_L\}$; ρ is the risk aversion parameter, and α_{ij} are the effort disutility coefficients, defined as before. Note $0 < \alpha_{i1} < \alpha_{i2}$.

The incentive compatibility constraint implies that for a given $q \in \{q_I, q_T\}$ and $\{\alpha_{ij}\}_{i=1,2,j=1,2}$, the optimal compensation scheme for manager i satisfies the following in-

[20] $K(\cdot)$ is a multivariate standard normal kernel density function:

$$K(\frac{x_{mt}-x_{nt}}{h_x}, \frac{b_{mt}-b_\tau}{h_b}) = \exp\left\{-0.5 * \left(\frac{x_{mt}-x_{nt}}{h_x}\right)^2\right\} * \exp\left\{-0.5 * \left(\frac{b_{mt}-b_\tau}{h_b}\right)^2\right\},$$

where $h_x = 0.96 * sd_x * N^{-1/6}$, $h_b = 0.96 * sd_b * N^{-1/6}$ are a cross-validation bandwidth and $sd_x(sd_b)$ is the standard deviation of gross abnormal return (bond price).

equality:

$$
\begin{aligned}
& p * [-\alpha_{i2} e^{-\rho w_i(x_H)}] + (1-p) * [-\alpha_{i2} e^{-\rho w_i(x_L)}] \\
\geq \quad & q * [-\alpha_{i1} e^{-\rho w_i(x_H)}] + (1-q) * [-\alpha_{i1} e^{-\rho w_i(x_L)}] \\
\Longrightarrow \quad & (\alpha_{i2} p - \alpha_{i1} q) e^{-\rho w_i(x_H)} \leq (\alpha_{i1} - \alpha_{i2} + \alpha_{i2} p - \alpha_{i1} q) e^{-\rho w_i(x_L)} \\
\Longrightarrow \quad & e^{-\rho [w_i(x_H) - w_i(x_L)]} \leq \frac{\alpha_{i1} - \alpha_{i2}}{\alpha_{i2} p - \alpha_{i1} q} + 1. \qquad (40)
\end{aligned}
$$

Note that the left hand-side of the last line decreases in the wage spread between two output levels for agent i because $\rho > 0$, and the right-hand side of the last line decreases in q because $\alpha_{i1} < \alpha_{i2}$.

From the shareholders' perspective, if manager i's preference of risk and effort costs are fixed, the compensation spread $w_i(x_H) - w_i(x_L)$ increases in q. From the researcher's perspective, the data tell about the spread (> 0) and p, which are both fixed. The *Team Model* has $q_T < q_I$ because the incentive compatibility constraint is relaxed and thus the suboptimal effort pair is both shirking. Given the binding incentive compatibility constraint (equality held in (40)) and fixed wage spread, ρ is expected to be smaller in the *Team Model*, which rationalizes the same data as the *Individual Model* does.

References

[1] ANTLE, R., and A. SMITH. 'Measuring Executive Compensation: Methods and an Application.' Journal of Accounting Research 23 (1985): 296–325.

[2] ANTLE, R., and A. SMITH. 'An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives.' Journal of Accounting Research 24 (1986): 1–39.

[3] ARYA, A., J. FELLINGHAM, and J. GLOVER. 'Teams, Repeated Tasks, and Implicit Incentives.' Journal of Accounting and Economics 23 (1997): 7–30.

[4] BERTOMEU, J., P. MA, and I. MARINOVIC. 'How Often Do Managers Withhold Information?,' Working paper, 2015. Available at https://www.krannert.purdue.edu/ academics/Accounting/bkd_speakers/papers/Bertomeu.pdf

[5] BEYER, A., I. GUTTMAN, and I. MARINOVIC. 'Earnings Management and Earnings Quality: Theory and Evidence,' Working paper, 2014. Available at http://ssrn.com/abstract=2516538.

[6] BUSHMAN, R. M., and A. J. SMITH. 'Financial Accounting Information and Corporate Governance.' Journal of Accounting and Economics 32 (2001): 237–333.

[7] BUSHMAN, R., Z. DAI, and W. ZHANG. 'Management Team Incentive Dispersion and Firm Performance.' The Accounting Review Forthcoming (2016).

[8] CHE, Y.-K., and S.-W. YOO. 'Optimal Incentives for Teams.' American Economic Review 91 (2001): 525–541.

[9] CORE, J. E., W. GUAY, and D. F. LARCKER. 'Executive Equity Compensation and Incentives: A Survey.' Federal Reserve Bank of New York Economic Policy Review April (2003): 27–50.

[10] DYE, R. A.. 'The Trouble with Tournaments.' Economic Inquiry 22 (1984): 147–149.

[11] EDMANS, A., and X. GABAIX. 'Is CEO Pay Really Inefficient? A Survey of New Optimal Contracting Theories.' European Financial Management 15 (2009): 486–496.

[12] EDMANS, A., I. GOLDSTEIN, and J. ZHU. 'Contracting with Synergies,' Working paper, 2013.

[13] FRIEDMAN, H. L. 'The Role and Responsibilities of the CFO: Implications for Reporting Quality, Firm Value, and Incentive Compensation,' Working paper, 2012. Available at

[14] GAYLE, G.-L., and R. A. MILLER. 'Has Moral Hazard Become a More Important Factor in Managerial Compensation?' American Economic Review 99 (2009): 1740–1769.

[15] GAYLE, G.-L., and R. A. MILLER. 'Identifying and Testing Models of Managerial Compensation.' Review of Economic Studies 82 (2015): 1074–1118.

[16] GAYLE, G.-L., L. GOLAN, and R. A. MILLER. 'Promotion, Turnover and Compensation in the Executive Labor Market.' Econometrica 83 (2015a): 2293–2369.

[17] GAYLE, G.-L., C. LI, and R. A. MILLER. 'Was Sarbanes-Oxley Costly? Evidence from Optimal Contracting on CEO Compensation,' Working paper, 2015b. Available at http://ssrn.com/abstract=2652122.

[18] GERAKOS, J., and C. SYVERSON. 'Competition in the Audit Market: Policy Implications.' Journal of Accounting Research 53 (2015): 725–775.

[19] GONG, G., L. Y. LI, and J. Y. SHIN. 'Relative Performance Evaluation and Related Peer Groups in Executive Compensation Contracts.' The Accounting Review 86 (2011): 1007–1043.

[20] GOW, I. D., D. F. LARCKER, and P. C. REISS. 'Causal Inference in Accounting Research,' Working paper, 2015. Available at http://research.chicagobooth.edu/ ~/media/90CF65A6E20D44C6A70C579937A9778C.pdf.

[21] GROSSMAN, S. J., and O. D. HART. 'An Analysis of the Principal-Agent Problem.' Econometrica 51 (1983): 7–45.

[22] HALL, B. J., and J. B. LIEBMAN. 'Are CEOs Really Paid Like Bureaucrats?' The Quarterly Journal of Economics 103 (1998): 653–691.

[23] HAMBRICK, D. C.. 'Upper Echelons Theory: An Update.' Academy of Management Review 32 (2007): 334–343.

[24] HEMMER, T. 'Lessons Lost in Linearity: A Critical Assessment of the General Usefulness of LEN Models in Compensation Research.' Journal of Management Accounting Research 16 (2004): 149–162.

[25] HENDERSON, A. D., and J. W. FREDRICKSON. 'Top Management Team Coordination Needs and the CEO Pay Gap: A Competitive Test of Economic and Behavioral Views.' The Academy of Management Journal 44 (2001): 96–117.

[26] HOLMSTROM, B. 'Moral Hazard and Observability.' The Bell Journal of Economics 10 (1979): 74–91.

[27] HOLMSTROM, B. 'Moral Hazard in Teams.' The Bell Journal of Economics 13 (1982): 324–340.

[28] INDJEJIKIAN, R., and M. MATEJKA. 'CFO Fiduciary Responsibilities and Annual Bonus Incentives.' Journal of Accounting Research 47 (2009): 1061–1093.

[29] ITOH, H. 'Coalitions, Incentives, and Risk Sharing,' Unpublished manuscript, 1990. Available at http://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/37905/1/10.pdf.

[30] ITOH, H. 'Incentives to Help in Multi-agent Situations.' Econometrica 59 (1991): 611–636.

[31] ITOH, H. 'Coalitions, Incentives, and Risk Sharing.' Journal of Economic Theory 60 (1993): 410–427.

[32] ITTNER, C., and D. LARCKER. 'Empirical Managerial Accounting Research: Are We Just Describing Management Consulting Practice?' European Accounting Review 11 (2002): 787–794.

[33] KREPS, D. 'Corporate Culture and Economic Theory.' In Perspectives on Positive Political Economy. Cambridge, UK: Cambridge University Press, 1990.

[34] LARCKER, D. F., and B. TAYAN. 'Trust: The Unwritten Contract in Corporate Governance.' In Rock Center for Corporate Governance at Stanford University Closer Look Series: Topics, Issues and Controversies in Corporate Governance and Leadership. No. CGRP-34. Stanford, CA: Stanford University, 2013.

[35] LAZEAR, E. P., and S. ROSEN. 'Rank-Order Tournaments as Optimum Labor Contracts.' The Journal of Political Economy 89 (1981): 841–864.

[36] LI, F., M. MINNIS, V. NAGAR, and M. RAJAN. 'Knowledge, Compensation, and Firm Value: An Empirical Analysis of Firm Communication,' Journal of Accounting and Economics 58 (2014): 96-116.

[37] LI, M. Moral Hazard and Internal Discipline: Theory and Evidence. Chicago: University of Chicago Booth School of Business, 2012.

[38] LI, Z. F. 'Mutual Monitoring and Corporate Governance.' Journal of Banking and Finance 45 (2014): 255–269.

[39] MA, C.-T. 'Unique Implementation of Incentive Contracts with Many Agents.' The Review of Economic Studies 55 (1988): 555–572.

[40] MAIN, B. G. M., C. A. O'REILLY III, and J. WADE. 'Top Executive Pay: Tournament or Teamwork?' Journal of Labor Economics 11 (1993): 606–628.

[41] MARGIOTTA, M. M., and R. A. MILLER. 'Managerial Compensation and the Cost of Moral Hazard.' International Economic Review 41 (2000): 669–719.

[42] MATZKIN, R. L. 'Nonparametric Identification.' In Handbook of Econometrics. Vol. 6, Part B. New York: Elsevier, 2007.

[43] MIRRLEES, J. A. 'The Theory of Moral Hazard and Unobservable Behaviour,' Mimeo, Oxford University, 1975.

[44] MURPHY, K. J. 'Executive Compensation.' In Handbook of Labor Economics, ed. O. C. Ashenfelter and D. Card. Vol. 3, Part B. New York: Elsevier, 1999.

[45] MURPHY, K. J. 'Executive Compensation: Where We Are, and How We Got There.' In Handbook of the Economics of Finance: Corporate Finance, ed. R. M. S. George, M. Constantinides, and M. Harris. 1st ed. Elsevier Science North Holland: Amsterdam, 2013.

[46] RAMAKRISHNAN, R. T. S., and A. V. THAKOR. 'Cooperation versus Competition in Agency.' Journal of Law, Economics, and Organization 7 (1991): 248–286.

[47] TAYLOR, L. A. 'CEO Wage Dynamics: Estimates from a Learning Model.' Journal of Financial Economics 108 (2013): 79–98.

[48] WINTER, E. 'Optimal Incentives for Sequential Production Processes.' The RAND Journal of Economics 37 (2006): 376–390.

[49] WINTER, E. 'Transparency and Incentives among Peers.' The RAND Journal of Economics 41 (2010): 504–523.

[50] ZAKOLYUKINA, A. A. 'Measuring Intentional GAAP Violations: A Structural Approach,' Working paper, 2014. Available at http://ssrn.com/abstract=2242251.

Table 1: Cross-Sectional Summary on Firm Characteristics

This table reports the cross-sectional statistics of three firm characteristics by industrial sector. Mean is reported and standard deviation is in the parentheses below. Assets are measured by the total assets on the balance sheet (AT). For debt/equity ratios, the numerator is the total liabilities (LT) and the denominator is the total common equity (CEQ). Market value equals the stock price at fiscal year end (PRCCF) times the number of outstanding shares (SHROUT) plus the liquidating value of preferred stock (PSTKL), if it is available. The book values of asset, liability, and equity and market values are deflated to the base year 2006, using a deflator constructed from the Consumer Price Index for All Urban Consumers (all items with seasonal adjustment), which was obtained from the website of the U.S. Department of Labor (Bureau of Labor Statistics). Both assets and market value are measured in millions of 2006 US$.

Sector	Primary	Consumer Goods	Service
Assets	4704	3059	4688
	(7423)	(5035)	(8307)
Debt/equity	1.84	1.52	2.56
	(1.40)	(1.41)	(3.36)
Market value	3285	3440	3417
	(4808)	(5181)	(5059)
Observations	6583	5004	8023

Table 2: Cross-Sectional Summary on Abnormal Stock Returns and Total Compensation

Total compensation is measured in thousands of 2006 US$. Mean is reported and standard deviation is in the parentheses below. In the first three columns, the third row for each type of firm reports the number of observations. In each industrial sector and each year, firms are classified based on the firm size, which is measured by the total assets on the balance sheet and denoted by A, and the capital structure, which is measured by the debt-to-equity ratio and denoted by D/E. Each of the two variables can have two values, that is, small (S) or large (L). If the value of total assets of firm n in year t is below the median of total assets in its sector, $A_{nt} = S$; otherwise, $A_{nt} = L$. The same rule applies to D/E_{nt}. I collect raw stock prices and adjustment factors from the CRSP and COMPUSTAT/PDE. For each firm in the sample, I calculate monthly compounded returns adjusted for splitting and repurchasing for each fiscal year and subtract the return to a value-weighted market portfolio (NYSE/NASDAQ/AMEX) from this raw return to get the raw abnormal return for its corresponding fiscal year. The total compensation is constructed by adding change in wealth from options held and change in wealth stocks held to the other components of compensation included in COMPUSTAT ExecuComp. See Appendix B for a detailed calculation. The ranking of compensation is based on the total compensation.

	Abnormal Stock Returns			Highest Compensation			Second Highest Compensation		
$[A, D/E]$	Primary	Consumer Goods	Service	Primary	Consumer Goods	Service	Primary	Consumer Goods	Service
$[S, S]$	-0.020	-0.030	-0.026	4746	6294	6877	1761	2042	2518
	(0.317)	(0.339)	(0.366)	(8297)	(11048)	(10577)	(2727)	(3561)	(3529)
	2284	1707	3079						
$[S, L]$	-0.005	-0.037	-0.009	3798	5039	6484	1523	1879	2552
	(0.325)	(0.354)	(0.335)	(6020)	(8379)	(10064)	(2221)	(3018)	(3424)
	1004	791	928						
$[L, S]$	-0.021	-0.028	-0.027	8409	10224	13994	3488	4050	5754
	(0.277)	(0.276)	(0.325)	(10887)	(13612)	(15765)	(4319)	(4807)	(5927)
	1003	791	928						
$[L, L]$	-0.017	-0.024	0.030	7501	11388	11483	3158	4506	4705
	(0.264)	(0.296)	(0.276)	(10311)	(13647)	(13868)	(3875)	(4877)	(5308)
	2292	1715	3088						

Table 3: Distribution of Positions Held by the Two Highest Paid Managers

"1st" is the highest paid manager and "2nd" is the second highest paid based on total compensation. For each level of managers, I count the frequency of holding certain types of positions as follows: "Functional" for managers holding one of the positions CTO, CIO, COO, CFO, or CMO, but not any others; "General 1" for managers holding one of the positions chairman, president, CEO, or founder, but not any others; "General 2" for managers holding one of the positions executive vice-president, senior vice-president, vice-president, vice-chair, or other (defined in the database), but not any others; "Functional & General 1" for managers holding at least one position from each of the Functional category and the General 1 category, respectively, but none from the General 2 category. The same rule applies to "Functional & General 2" and to "General 1 & General 2." "Functional & General 1 & General 2" is for managers holding at least one position from each of the three categories.

	Primary		Consumer Goods		Service	
Compensation Rank	1st	2nd	1st	2nd	1st	2nd
Functional	1%	1%	1%	1%	1%	2%
General 1	21%	8%	17%	8%	21%	9%
General 2	18%	51%	23%	50%	23%	51%
Functional & General 1	4%	10%	6%	11%	4%	8%
Functional & General 2	5%	18%	7%	15%	6%	15%
General 1 & General 2	50%	12%	45%	14%	44%	13%
Functional & General 1 & 2	1%	1%	1%	1%	1%	2%
Total	100%	100%	100%	100%	100%	100%
Number of observations	6583		5004		8023	

Table 4: The Risk Aversion Parameter's 95% Confidence Regions

Column $[A,\ D/E]$ defines the firm type, which is based on firm size (total assets, A) and capital structure (debt-to-equity ratio, D/E). S (L) means the corresponding element is below (above) its sector median. The confidence region is estimated by a subsampling procedure using 300 replications of subsamples with size equal to 15% of the full sample. The certainty equivalent for a given risk aversion value is the amount paid to avoid a gamble with equal probability to win and lose US$1 million and is measured in US$ million with the median of the bond price in the sample period (16.65).

A: *Individual Model*—different likelihood ratio + different shadow price of incentive compatibility constraint

Sector	$[A,\ D/E]$	Risk Aversion	Certainty Equivalent	Homogeneous within Size	Homogeneous within Sector	Homogeneous across Sectors
Primary	$[S,\ S]$	(12.75, 26.38)	(0.350, 0.589)			
	$[S,\ L]$	(0.89, 16.25)	(0.027, 0.426)	(12.75, 16.25)		
	$[L,\ S]$	(6.16, 33.62)	(0.181, 0.665)			
	$[L,\ L]$	(0.89, 2.34)	(0.027, 0.070)	(,)	(,)	
Consumer Goods	$[S,\ S]$	(0.26, 3.79)	(0.008, 0.113)			
	$[S,\ L]$	(1.83, 33.62)	(0.055, 0.665)	(1.83, 3.79)		
	$[L,\ S]$	(0.34, 1.13)	(0.010, 0.034)			
	$[L,\ L]$	(0.70, 2.34)	(0.021, 0.070)	(0.70, 1.13)	(,)	
Service	$[S,\ S]$	(4.83, 26.38)	(0.143, 0.589)			
	$[S,\ L]$	(0.55, 12.75)	(0.016, 0.350)	(4.83, 12.75)		
	$[L,\ S]$	(1.44, 7.85)	(0.043, 0.228)			
	$[L,\ L]$	(1.44, 20.7)	(0.043, 0.507)	(1.44, 7.85)	(4.83, 7.85)	(,)

B: *Team Model*—same likelihood ratio + different shadow price of incentive compatibility constraint

Sector	$[A,\ D/E]$	Risk Aversion	Certainty Equivalent	Homogeneous within Size	Homogeneous within Sector	Homogeneous across Sectors
Primary	$[S,\ S]$	(0.10, 20.70)	(0.003, 0.529)			
	$[S,\ L]$	(0.16, 12.75)	(0.005, 0.370)	(0.16, 12.75)		
	$[L,\ S]$	(0.05, 10.00)	(0.002, 0.301)			
	$[L,\ L]$	(0.08, 1.83)	(0.003, 0.059)	(0.08, 1.83)	(0.16, 1.83)	
Consumer Goods	$[S,\ S]$	(0.05, 2.98)	(0.002, 0.095)			
	$[S,\ L]$	(0.21, 20.70)	(0.007, 0.529)	(0.21, 2.98)		
	$[L,\ S]$	(0.02, 0.89)	(0.001, 0.028)			
	$[L,\ L]$	(0.03, 2.34)	(0.001, 0.075)	(0.03, 0.89)	(0.21, 0.89)	
Service	$[S,\ S]$	(2E-9, 33.62)	(2E-9, 0.685)			
	$[S,\ L]$	(0.04, 16.25)	(0.001, 0.447)	(0.04, 16.25)		
	$[L,\ S]$	(0.02, 4.83)	(0.001, 0.153)			
	$[L,\ L]$	(0.05, 33.62)	(0.002, 0.685)	(0.05, 4.83)	(0.05, 4.83)	(0.21, 0.89)