

Mutual Monitoring within Top Management Teams: A Structural Modeling Investigation

(JOB MARKET PAPER)

Chen Li*

2012-12-05

Abstract

Mutual monitoring as a solution to moral hazard has been extensively studied by theorists. However, do shareholders actually take advantage of mutual monitoring among executives to design individual compensation? This paper non-parametrically identifies and tests three structural models of principal-multiagent moral hazard using the data of S&P1500 firms from 1993 to 2005. The Mutual Monitoring with Individual Utility Maximization Model is the most consistent with the data. The No Mutual Monitoring Model is also plausible but relies on the assumption that managers have heterogeneous risk preferences across firm characteristics. The Mutual Monitoring with Total Utility Maximization Model is rejected by the data. This paper indicates that shareholders seem to recognize and exploit complementary incentive mechanisms, such as mutual monitoring among self-interested top executives, to design compensation.

Keywords: Mutual Monitoring, Moral Hazard, Top Management Team, Executive Compensation, Nonparametric Identification, Structural Estimation, Moment Inequalities.

*I am indebted to my dissertation committee—George-Levi Gayle, Jonathan Glover (co-chair), Pierre Jinghong Liang, and Robert Miller (co-chair)—for their invaluable guidance and support. I also thank Michael Ewens, Jack Stecher, Ran Zhao, and seminar participants at Carnegie Mellon for their comments. All errors are my own. Li is a Ph.D. candidate in accounting at Carnegie Mellon University. E-mail: cli1@andrew.cmu.edu.

1 Introduction

Shareholders design optimal compensation to mitigate the moral hazard of hidden effort and free riding in top management teams. In a seminal paper, Fama (1980) points out that "each manager has a stake in the performance of the managers above and below him and, as a consequence, undertakes some amount of monitoring in both directions."¹ Although theoretical models have extensively explored how mutual monitoring is intertwined with individual compensation in the optimal contract responding to moral hazard (Bolton and Dewatripont 2005; Glover 2012), empirical studies mainly examine individual incentives to understand top executive compensation (MacLeod 1995; Murphy 1999, 2012; Core et al. 2003). In general, overlooking the effect of mutual monitoring as a self-policing vehicle may lead to incomplete or even misleading evaluations of the severity of the moral hazard problem and, thus, of the efficiency of executive compensation. At the heart of this gap in the literature is a question about the empirical relevance of mutual monitoring models: do shareholders actually take advantage of mutual monitoring in optimal compensation design?

The research challenge is that mutual monitoring among top executives is rarely codified in their contracts or observed by outsiders. So far, a few indirect tests have produced only mixed results by studying the association between firm performance and the top executives' cooperation/monitoring incentives proxied by relative properties of compensation.² However, the optimal compensation is usually derived from primitive parameters³ which also determine the optimal effort and output that shareholders prefer in equilibrium, creating an endogeneity problem acknowledged by empiricists (Prendergast 1999; Core et al. 2003).

Taking a more direct approach, the empirical investigation in this paper identifies and tests three competing structural models that are explicitly based on theoretical models of principal-multiagent moral hazard. I set up my models with one joint output (stock return), one risk-neutral principal (shareholders), and two risk-averse agents (the two highest paid managers), who have the same absolute risk aversion coefficient but differ in their costs of effort. The three models differ in terms of how the shareholders provide managers with incentives to participate and incentives to work rather than shirk. These differences depend on whether and how the managers monitor each other, as follows.

If shareholders believe the managers cannot effectively side contract to monitor each

¹A recent paper (Landier et al. 2012) provides evidence of bottom-up monitoring of CEOs by top executives who joined the firm before the current CEO.

²Evidence in support of cooperation/monitoring can be found in Li (2011) and Bushman et al. (2012). Unsupportive evidence is provided by Main et al. (1993), Henderson and Fredrickson (2001), and Bushman et al. (2012).

³For example, these deeper parameters can be managers' risk preferences, costs of effort, and the relative informativeness of a performance measure on the equilibrium path versus off the equilibrium path.

other, they have to provide the managers with individual incentives through the compensation contract. The first model, called *no mutual monitoring*, describes this case and serves as a benchmark. Without mutual monitoring, the shareholders are concerned about managers' unilateral shirking and design the optimal compensation such that both managers working (the optimal effort pair throughout this paper) is a Nash equilibrium in the managers' subgame. Alternatively, if shareholders believe managers can side contract on mutually observable efforts, they will take advantage of the mutual monitoring in contract design (Holmstrom and Milgrom 1990; Varian 1990; Ramakrishnan and Thakor 1991; Itoh 1993, among others). The managers cooperate both to choose working as a Pareto-dominant equilibrium and to agree on equal expected utility due to their equal bargaining power in the private coordination process. Furthermore, if shareholders think the managers engage in mutual monitoring to pursue group interests, the second model, called *mutual monitoring with total utility maximization*, describes this case. In this model, the shareholders provide the two managers with incentives only based on their total expected utility.⁴ By contrast, if the managers pursue self-interest, the third model, called *mutual monitoring with individual utility maximization*, describes this case. Because each manager chooses working based on individual rationality, shareholders need to tailor each of those two incentives to each manager's preference over his own expected utility maximization.⁵

The intuition for my empirical strategy is as follows. Even though we do not know how shareholders design the incentives of the optimal contract in their minds, we do observe the compensation they offer and the output the managers generate. Traditionally, we test comparative statics, such as the relation between pay and performance, to infer what the optimal contract may look like, for example, whether internal monitors are motivated to monitor and enhance firm value (Armstrong et al. 2010) or whether relative performance evaluation is adopted (Antle and Smith 1986). Instead of focusing on the consequences of the optimal contract, this paper directly examines the data restrictions required by an optimal contract to discipline parameters so that the observed compensation and stock returns can be consistently understood within a unified framework. Theory helps here because the optimal contract can essentially be described by a well-defined theoretical model. If share-

⁴This model has the essence of the *mutual monitoring with utility transfer model* in Itoh (1993, page 416). To make the current model less restrictive on the data, I drop Itoh's assumption that the two managers can transfer payments to share risk ex post. This assumption seems unrealistic among top executives and would be rejected by the data. I retain only Itoh's assumption on transferable utility in my model.

⁵This model essentially says that a Pareto-dominant strategy is played in equilibrium without utility transfer even though free-riding is optimal from the viewpoint of individual incentives. There are a few mechanisms that can be empirically consistent with this model, 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), the infinitely repeated game with implicit side contracts in Che and Yoo (2001), leadership by setting example in Hermalin (1998), and the peer pressure in Kandel and Lazear (1992), among others.

holders honor their compensation arrangements with managers and managers exert optimal effort to generate stock returns as expected, then the observed compensation and stock returns are random draws from the equilibrium of a theoretical model that characterizes that optimal contract in shareholders' minds, after controlling for the heterogeneity in the data. Intuitively, if the data restrictions implied by the equilibrium of the theoretical model are statistically consistent with the observed data pattern, this consistency suggests that the observed compensation schemes have the flavor of that model. In this paper, the "flavor" refers to whether shareholders exploit mutual monitoring and how managers are engaged. The purpose of the tests is to find out which type of model (contract) can explain the entire data best, allowing the contract shape to vary with firm characteristics, industrial sectors, and macroeconomic fluctuations.

First I show that, without imposing on data the restrictions from shareholders' profit maximization over the alternative effort pairs of managers, the pattern of compensation and stock returns can be empirically consistent with a model with or without mutual monitoring. An important implication is that the descriptive properties of compensation, which are usually based on comparative statics derived from the subset of equilibrium conditions, may not be sufficient to help us distinguish the two types of models without considering other restrictions that those confounding parameters need to satisfy. This partially helps to illustrate why different research designs can lead to opposite results in the literature.

Then I exploit other equilibrium restrictions implied by this model, for example, shareholders' preferences over all possible effort pairs and managers' time-invariant preferences over risk, to govern the identified set of the risk aversion parameter to which all other primitive parameters in the same model are indexed. These restrictions are summarized by a criterion function that has a distance-minimizing property. If the model can explain the data, there must exist some reasonable values of the risk aversion parameter in the identified set such that the criterion function reaches its lower bound.

Next, I bring the theoretical restrictions to the data I investigate. The measurement of total compensation follows Antle and Smith (1985) by incorporating opportunity costs of holding firm stocks and stock options into managers' wealth.⁶ There are two noteworthy features of the panel data I investigate, which cover S&P 1500 firms from 1993 to 2005. First, the two managers studied in this paper earn the highest total compensation for a given firm-year, and their compensation contracts are intensively equity based. This indicates not only that they have significant influence on the stock returns due to their occupational seniority but also that they can substantially benefit from the improvement of this joint output. This

⁶Among followers are Hall and Liebman (1998), Margiotta and Miller (2000), Gayle and Miller (2009, 2012), and Gayle et al. (2012).

tight interest alignment provides a channel and an incentive of sanction that favor the two models in which shareholders take advantage of mutual monitoring (Kandel and Lazear 1992). Second, for 94 percent of the sample firm-years, the two managers either hold a functional position (CTO, CIO, COO, CFO, CMO)⁷ or sit on the top rank, including the positions of president, chairman, CEO, and founder. These two types of positions are hardly substitutable. As a result, it is reasonable to assume that shareholders prefer both managers working to allowing either one to shirk.

To account for the measurement errors in the compensation and to acknowledge the flexibility of shareholders' contract designs, this paper nonparametrically estimates both the optimal compensation scheme as a function of the gross abnormal return and the density of the gross abnormal return in equilibrium. To reduce the concern of overusing structures, the nonparametric method in this paper enables exploiting the information from data as much as possible and also avoids rejecting a model due to specific model assumptions on contract form and distribution. This method shortens "the distance between those roads to the point where now some econometric models are specified with no more restrictions than those that a theorist would impose" (Matzkin 2007, page 5311).

Last, I calculate the criterion function with the data for each model, such that I can construct a hypothesis test for the model based on the confidence region of the identified set of the risk aversion parameter. I use a similar testing strategy developed for the single-agent model of moral hazard and hidden information by Gayle and Miller (2012), who investigate the role of accounting information in CEOs' compensation contracts and are followed by Gayle et al. (2012), who explore the consequences of the Sarbanes-Oxley Act on CEOs' compensation. If the confidence region is empty or only contains unreasonable values, the model is rejected.

The main results emerge from the preceding steps, as follows. The mutual monitoring with total utility maximization model is rejected, even under the least restrictive assumption that managers have heterogeneous risk preferences across firm types and industrial sectors. The confidence region is empty in large firms of the primary sector and in small firms with high financial leverage of the service sector. The nonempty confidence regions cover values close to zero in all other firms, indicating that to be reconciled with the data, this model requires almost risk-neutral managers. Such near-risk neutrality contradicts the setup of this model, which assumes that the managers are risk averse. This contradiction essentially rejects this model.

Under the same heterogeneity assumption of risk aversion, both the no mutual monitor-

⁷CTO: Chief Technology Officer, CIO: Chief Information Officer, COO: Chief Operation Officer, CFO: Chief Financial Officer, CMO: Chief Marketing Officer.

ing model and the mutual monitoring with individual utility maximization model cannot be rejected. However, under the most restrictive assumption that managers have homogeneous risk preference across firm types and industries, only the mutual monitoring with individual utility maximization model cannot be rejected. In this sense, the mutual monitoring with individual utility maximization model is the most robust among the three models to rationalize the correlation between the observed top executive compensation and stock returns. This result implies that we may need to account for the cross-sectional variation of mutual monitoring in trying to understand the incentives embedded in executive compensation. Intuitively, enforceable mutual monitoring among top managers can help shareholders partially save compensation cost. In turn, a large equity-based component in compensation aligns the interests of a group of managers through a joint output that provides the channel and the incentive for mutual punishment and reward.

Furthermore, I examine how shareholders perceive managers engaging in mutual monitoring, which has not been tested previously in the literature. I find that shareholders consider that the managers monitor each other to pursue self-interest rather than to pursue their collective interests. This result has implications for how to account for the effect of mutual monitoring on compensation in empirical research. If shareholders take into account the utility transfer that is implicitly assumed for total utility maximization, the shape of the optimal compensation is more similar between managers than individual utility maximization predicts. Previous studies using the closeness of managers' compensation schemes to detect team incentives, for example, the pay disparity (Main et al. 1993) and the dispersion of pay-performance-sensitivity (Bushman et al. 2012), do not support a dominant effect of cooperation/monitoring. The results in this paper suggest that moderate closeness can be consistent with the model of mutual monitoring if managers are not identical and only care about their own payoffs. Consequently, this result implies that the proxy choice should account for the underlying incentive and enforcement mechanism of mutual monitoring, which was ignored in previous studies.

The preceding more direct answers have the potential to advance our understanding of how shareholders respond to the moral hazard in top management teams and how managers are engaged in mutual monitoring. This enriched understanding can extend structural modeling studies by suggesting that the mutual monitoring may be incorporated as a baseline in rationalizing the curvature of executive compensation. This paper also sheds light on studies that investigate the determinants and consequences of executive compensation by calling attention to appropriate control for the implicit incentive effect of mutual monitoring in addition to traditional corporate governance factors, which rely on explicit provisions of incentives. Instead of focusing on the similarity of compensation shape, researchers may

want to consider factors that affect the enforcement of mutual monitoring 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) suggested by theoretical studies, and the team duration used by the empirical paper of Bushman et al. (2012).

The remaining is arranged as follows. In Section 2, I compare the static versions of the three models. To incorporate dynamic considerations, I estimate and test the dynamic versions of these models in later sections.⁸ Section 3 discusses the data and the nonparametric estimation. Section 4 establishes the identification. Section 5 introduces the estimation and hypothesis tests. Section 6 reports and discusses the results. Section 7 discusses feasible extensions, and Section 8 concludes.

2 Models

This section lays out the three principal-multiagent models of moral hazard as the theoretical underpinning of the structural model identification and the hypothesis tests. These models aim to sufficiently distinguish the shareholders' perception on mutual monitoring up to the extent that the primitive parameters can be recovered from the observed compensation and abnormal stock returns. These models are not constructed to comprehensively explore the delicate strategic interactions between shareholders and managers in complex reality. However, as I gradually introduce the three models, I will discuss how these general models can be empirically consistent with some well-established models in the theoretical literature of multiagent moral hazard.

I model the shareholders' decision-making process following the two-step procedure in Grossman and Hart (1983). I start from their second step by formulating the shareholders' cost minimization problem. I assume throughout this paper that shareholders prefer motivating both managers to work. In the following, I first introduce the three models' common setups, including the timeline, technologies, managers' preferences, and shareholders' objective function. Then I discuss their differences in terms of whether and how shareholders take into account managers' mutual monitoring at the optimal contract design. If shareholders take advantage of managers' mutual monitoring, they contrast implementing the optimal effort pair (both managers working) with the suboptimal effort pair (both managers shirking); otherwise, they are concerned about each manager's unilateral shirking. If managers can transfer utility, shareholders provide incentives based on managers' total utilities. Otherwise, the incentive is consistent with each manager's utility maximization.

⁸The dynamic version falls into the principal-agent moral hazard framework of Margiotta and Miller (2000), as descended from Grossman and Hart (1983) and Fudenberg et al. (1990).

At the end of this section, I discuss the first step of Grossman and Hart (1983) after the optimal contracts are derived. In this step, shareholders compare their net benefit from implementing a given effort pair of the two managers and select the optimal effort that gives the largest net benefit among all possible effort pairs.

2.1 Timeline

In a static model, the timeline of the interaction between the risk-neutral shareholders and the two risk-averse managers⁹ is as follows. At the beginning of a period, the shareholders propose a compensation scheme $w_i(x)$ for manager i ; 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 \tilde{x} denote the abnormal stock return realized from this period; \tilde{x} is the idiosyncratic component of the firm's stock return, which is under the control of the managers. To be consistent with the tradition of agency models, I construct the performance measure variable x , called gross abnormal return, as

$$x = \tilde{x} + \frac{w_1}{V} + \frac{w_2}{V}.$$

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. I assume neither manager can operate the firm by himself. This is realistic because modern firms are large such that they are rarely run by a single manager. As a result, one manager has to wait for another manager to join the team and proceed together.

After accepting the shareholders' offer, each manager can choose between two effort levels, namely, working and shirking. The interdisciplinary knowledge set of managing large diversified firms requires that top managers work closely 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 teams¹⁰. I assume in all models 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 of hidden action is the fundamental friction in single-agent models. In the multiagent models of this paper, there is another friction called *free riding*. If one manager shirks, he can avoid his entire disutility

⁹It might be interesting to explore the coordination among more than two managers, for example, embedding a coalition stability problem into the principal-agent setting. However, this is not the focus of this paper and is thus left for future studies.

¹⁰This assumption rules out the revelation mechanism like Ma (1988).

of working but only has to partially bear the loss from the reduction in output if the other manager works. Thus each manager has an incentive to count on the other one and shirks. To account for the unilateral shirking, it is necessary to specify the effort choice for each manager. Let j denote manager 1's effort choice and k denote manager 2's. To sum up, I define the three mutually exclusive choices as

$$j(k) = \begin{cases} 0, & \text{if manager 1(2) rejects the offer} \\ 1, & \text{if manager 1(2) accepts the contract but shirks later} \\ 2, & \text{if manager 1(2) accepts the contract and works later.} \end{cases}$$

At the end of the period, the joint output x is realized and manager i gets paid according to his compensation scheme $w_i(x)$. Conditioning on the managers' effort choice (j, k) , x is a random draw from an independent and identical distribution across firms in this static model (or across both firms and periods in a dynamic model), after controlling for the heterogeneity in the data.

2.2 Technologies

The technologies are captured by the probability density function (PDF) of the joint output x conditional on the two managers' effort choices. I denote $f(x)$ as the PDF of x conditional on both managers working, that is, 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 $\int \bullet f(x)dx$.

As to the PDFs of x conditional on managers' effort pairs off the equilibrium path, I introduce likelihood ratios to distinguish between managers' unilateral shirking and simultaneous shirking. To be specific, when manager i chooses to shirk but the other manager chooses to work, the product $g_i(x)f(x)$ denotes the corresponding PDF of x ; $g_i(x)$ is the likelihood ratio between the PDF of x conditional on manager i 's unilateral shirking over the PDF of x conditional on the equilibrium effort pair. In the single output framework, 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.¹¹

This specification is general enough to capture the 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 who is

¹¹This setup is suggested by Margiotta and Miller (2000) in their discussion on extending their single-agent framework to a multiagent one.

responsible for continuous 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.

Assuming that one manager's marginal influence on the PDF of x is unconditional on the other manager's effort choice, the product $g_1(x)g_2(x)f(x)$ is the PDF of x when both managers choose to shirk. This can be proved in the following Lemma.¹² Denote $g(x)$ as the likelihood ratio of the PDF of x conditional on both managers shirking over that conditional on both managers working.

Lemma 1

$$E[g(x)] \equiv \int g_1(x)g_2(x)f(x)dx = 1.$$

Two points are noteworthy. First, the unconditional density 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. Mathematically, the stochastic nature of the likelihood ratio makes $g_1(x) \neq g_2(x)$, because otherwise, $E[g_i(x)] = E[g_i^2(x)] = 1$ implies that $g_i(x)$ turns out to be a constant. Second, this unconditional density assumption can be consistent with the production of substitutability, independence, or complementarity. The stochastic property of production is captured by the difference in expected output, as follows: if the increment in expected output due to manager 1 switching from shirking to working conditional on manager 2 working is larger than that increment conditional on manager 2 shirking, then the production has a complementarity property; if the former increment is smaller than the latter, the two managers are substituted in production; if the two increments are the same, the production is considered as independent. Formally,

$$\begin{aligned} & \{E[x \mid j = 2, k = 2] - E[x \mid j = 1, k = 2]\} - \{E[x \mid j = 2, k = 1] - E[x \mid j = 1, k = 1]\} \\ = & \left\{ \int xf(x)dx - \int xg_1(x)f(x)dx \right\} - \left\{ \int xg_2(x)f(x)dx - \int xg_1(x)g_2(x)f(x)dx \right\} \\ = & \int x [1 - g_1(x)] [1 - g_2(x)]f(x)dx \\ & \begin{cases} > 0, \text{ complementary in production} \\ = 0, \text{ independent in production} \\ < 0, \text{ substitute in production.} \end{cases} \end{aligned}$$

Subsequently, I discuss four properties of the likelihood ratios. I denote in general the PDF associated with a suboptimal effort pair by the product $h(x)f(x)$ and $h(x) \in$

¹²All proofs are in Appendix A.

$\{g_1(x), g_2(x), g(x)\}$. First, by the definition of the likelihood ratio, $h(x)$ is nonnegative for any x , that is, $h(x) \geq 0, \forall x$, and also it satisfies

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

Second, I assume that an extraordinary output can be realized only when no one shirks. To put it mathematically, $h(x)$ satisfies

$$\lim_{x \rightarrow \infty} h(x) = 0.$$

Third, I assume $h(x)$ 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). Fourth, the shareholders and managers have conflicting interests in the sense that shareholders can benefit more if the managers work than if they shirk. To reflect such a conflict, I assume that the expected gross abnormal return increases with the number of working managers, namely,

$$\int xf(x)g(x)dx < \int xf(x)g_i(x)dx < \int xf(x)dx.$$

2.3 Managers' Preferences

Each manager's preference can be expressed using a negative exponential utility function with multiplicatively separable preference on effort.¹³ 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 $\tilde{\alpha}_{ij(k)}$ ($i = 1, 2, j(k) = 1, 2$) in the managers' utility functions as (1) and (2), defined later; $\tilde{\alpha}_{1j}$ ($\tilde{\alpha}_{2k}$) corresponds to manager 1(2)'s effort choice $j(k)$. For manager i , I assume $0 < \tilde{\alpha}_{i1} < \tilde{\alpha}_{i2}$, 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'. Take empire building, for example. The managers may exert substantial labor input to pick up projects that maximize their own private perks but not maximize the firm's value.

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

¹³The CARA utility function has obvious merit for tractability and is widely used in theoretical research, for example, the LEN model in agency theory.

Formally,

$$\text{Manager 1's expected utility} \equiv -\tilde{\alpha}_{1j}E[\exp(-\rho w_1(x)) | j, k], \quad (1)$$

$$\text{Manager 2's expected utility} \equiv -\tilde{\alpha}_{2k}E[\exp(-\rho w_2(x)) | j, k]. \quad (2)$$

In particular, on the equilibrium path, manager i gets his expected utility from compensation under the distribution of x conditional on both managers working adjusted by manager i 's effort cost coefficient with respect to working ($\tilde{\alpha}_{i2}$): $-\tilde{\alpha}_{i2} \int v_i(x)f(x)dx$.

As to the off-equilibrium path efforts, if manager i shirks but the other manager does not, manager i 's expected utility is modified by replacing his disutility coefficient with the one corresponding to shirking and replacing the distribution with that under manager i unilaterally shirking: $-\tilde{\alpha}_{i1} \int v_i(x)g_i(x)f(x)dx$.

If both managers shirk, the disutility coefficient remains $\tilde{\alpha}_{i1}$, but the distribution is replaced with that conditional on both managers shirking. Manager i 's expected utility is represented by: $-\tilde{\alpha}_{i1} \int v_i(x)g_1(x)g_2(x)f(x)dx$ or $-\tilde{\alpha}_{i1} \int v_i(x)g(x)f(x)dx$.

2.4 Shareholder's Cost Minimization Problem

2.4.1 Objective Function

For now, I assume that the shareholders prefer both managers working. The shareholders are assumed to be risk neutral, and thus their utility is measured in monetary terms, including a cost and a benefit. The shareholders' cost is the total compensation paid to the two managers, which needs to be delicately tied to the gross abnormal return x . The shareholders' benefit is the expected firm value growth conditional on both managers working, which is a constant when managers' effort choices are fixed. Consequently, the shareholders' optimization problem is to minimize the expected total compensation of the two managers. Furthermore, the expectation is taken over the distribution of the gross abnormal return conditional on both managers working. To simplify notation, I define the negative of manager i 's utility from compensation as

$$v_i(x) \equiv \exp(-\rho w_i(x)), \quad i = 1, 2.$$

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:

$$\int [\ln v_1(x) + \ln v_2(x)] f(x)dx. \quad (3)$$

This objective function in the shareholders' cost minimization problem is the same between the three models. However, depending on whether the shareholders believe that the managers can monitor each other and whether the shareholders perceive that the mutual monitoring can be implemented by the managers' private agreement on utility transfer, shareholders face different constraints across the three models. These differences become clearer in the following subsections.

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 denoted by $-\tilde{\alpha}_0$.¹⁴ Such a restriction is called the *participation constraint*, which places a bound on the set of feasible compensation schemes that shareholders can use to minimize the cost. Because the managers' preferences can be preserved for an increasing transformation, I normalize the utility function by dividing it with $\tilde{\alpha}_0$, and thus the outside option is normalized to -1 . Consequently, the effort disutility coefficient hereafter is the ratio of that coefficient over the outside option, that is,

$$\alpha_{ij} \equiv \frac{\tilde{\alpha}_{ij}}{\tilde{\alpha}_0}.$$

In both the no mutual monitoring model and the mutual monitoring with individual utility maximization model, managers make effort choices to maximize each manager's own expected utility such that the participation constraint is individualized to each manager's incentive. Formally, in (4) and (5), on the left-hand side of the top (bottom) line is manager 1 (2)'s expected utility, which consists of a CARA utility from compensation conditional on the distribution of the joint output if both managers work and a multiplicative disutility coefficient associated with manager 1 (2) working. The expectation is taken over the distribution of x conditional on both managers working. On the right-hand side is manager 1 (2)'s outside option normalized to -1 . The following weak inequalities reflect managers' preference over the two options:

$$-\alpha_{12} \int v_1(x) f(x) dx \geq -1, \tag{4}$$

$$-\alpha_{22} \int v_2(x) f(x) dx \geq -1. \tag{5}$$

¹⁴The outside option does not vary with the gross abnormal return, but this does not imply that the reservation compensation is zero.

In contrast, in the mutual monitoring with total utility maximization model, the two managers coordinate efforts through utility transfer in side contracts. Even though monetary transfer between top executives is hardly seen and probably prohibited in many firms¹⁵, and thus not allowed in my model, there are other channels for executives to punish or reward each other. For example, the two managers might use a side contract to split perquisites. The total utility maximization model can be seen as incorporating their nonmonetary transfers using a quasi-linear utility function that allows for transferable utility. My purpose is not to defend the transferable utility assumption but instead to include a model that allows for a richer set of side contracts, in the spirit of Itoh (1993).

The shareholders treat the two managers as a unitary decision maker, and thus the contract is based merely on the managers' total utility. The *group participation constraint* says that the two managers can be collectively better off by taking the shareholders' offer and subsequently working than by rejecting the offer. The following inequality reflects such a restriction. The left-hand side is the sum of the two managers' expected utilities conditional on both working, and the right-hand side is the total value of their outside options; that is,

$$-\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx \geq -2. \quad (6)$$

Note that the summation of the two managers' utilities puts the same weight on each. This implies an extra constraint in the mutual monitoring with total utility maximization model, called the *equal sharing rule*. I assume that the two managers agree to equalize expected utilities for any effort pair. This rule may reflect that the managers have equal bargaining power in the top management team or that it is necessary to keep fairness to reach an agreement on effort coordination.

Taking into account the possibility of managers' effort coordination in a side contract based on such a sharing rule, shareholders provide equal expected utility to the two managers in the optimal contract, when they both work and when they both shirk. As a result, in equilibrium there is no utility transfer between the two managers. On the left-hand (right-hand) side of equation (7) is the expected utility of manager 1 (2) given both managers shirking. On the left-hand (right-hand) side of equation (8) is the expected utility of manager 1 (2) given both managers working:

$$-\alpha_{11} \int v_1(x)f(x)g(x)dx = -\alpha_{21} \int v_2(x)f(x)g(x)dx, \quad (7)$$

$$-\alpha_{12} \int v_1(x)f(x)dx = -\alpha_{22} \int v_2(x)f(x)dx. \quad (8)$$

¹⁵Tirole (1992) points out that repeated interactions are the more plausible enforcement of side contracts.

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. Such a restriction on the shareholders' cost minimization problem is called the *incentive compatibility constraint*. It is helpful to tabulate the expected utilities conditional on the four effort pairs, shown in the table following. 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.

		Manager 2	
		Work	Shirk
Manager 1	Work	$-\alpha_{22}E[v_2(x)]$ $-\alpha_{12}E[v_1(x)]$	$-\alpha_{21}E[v_2(x)g_2(x)]$ $-\alpha_{12}E[v_1(x)g_2(x)]$
	Shirk	$-\alpha_{22}E[v_2(x)g_1(x)]$ $-\alpha_{11}E[v_1(x)g_1(x)]$	$-\alpha_{21}E[v_2(x)g(x)]$ $-\alpha_{11}E[v_1(x)g(x)]$

In the no mutual monitoring model, shareholders only use monetary incentive to avoid managers shirking. The informativeness of the gross abnormal return at each realization differs between the two managers. 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. The following two inequalities reflect this constraint.

In (9), the left-hand side is manager 1's expected utility if both managers work, which holds the same expression as previously defined in the participation constraint corresponding to manager 1. The right-hand side is manager 1's expected utility if manager 1 unilaterally shirks. It is calculated by multiplying his shirking disutility coefficient (α_{11}) by the utility from monetary compensation. And the expectation is taken over the distribution of the gross abnormal return conditional on that manager 1 unitarily shirks. The inequality (10) applies

the same constraint, which provides working incentive to manager 2:

$$-\alpha_{12} \int v_1(x)f(x)dx \geq -\alpha_{11} \int v_1(x)f(x)g_1(x)dx, \quad (9)$$

$$-\alpha_{22} \int v_2(x)f(x)dx \geq -\alpha_{21} \int v_2(x)f(x)g_2(x)dx. \quad (10)$$

In the mutual monitoring with total utility maximization model, the *group incentive compatibility constraint*, as it is called, is again based on total utility, as in the participation constraint, saying that both working is collectively preferred by the two managers to both shirking. Mathematically, the total expected utility from both working is weakly larger than that from both shirking, that is,

$$-\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx, \quad (11)$$

$$\geq -\alpha_{11} \int v_1(x)f(x)g(x)dx - \alpha_{21} \int v_2(x)f(x)g(x)dx. \quad (12)$$

A caveat is that in this model, I implicitly assume that both working strictly Pareto dominates unilateral shirking¹⁶. In principle, the optimal compensation schemes also need to satisfy the other two inequality constraints such that both working Pareto dominates either one shirking. The intuition is that the optimal compensation needs to prevent a shirker from bribing the worker with a perquisite transfer. This implies that shareholders offer compensation such that the shirker's utility after perquisite transfer, which equals half of the total utility when he unilaterally shirks, should be no more than what he can get from working, that is, half of the total utility when both managers work. This intuition applies to both managers.¹⁷

Note that the empirical optimal contracting approach of this paper assumes that the

¹⁶If the incentive compatibility constraints associated with unilateral shirking are binding, the identification of the current model will not change as long as the incentive compatibility constraint in (11) remains binding as assumed in the optimal contract in this paper. Otherwise, the binding constraints of unilateral shirking and the non-binding constraint of both shirking would constitute another structural model essentially different from the one studied in this paper, which might give different predictions on data.

¹⁷Formally, to guarantee that both working is Pareto dominant over either manager unilaterally shirking, the current compensation scheme needs to satisfy the following inequalities:

$$-\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx > -\alpha_{11} \int v_1(x)f(x)g_1(x)dx - \alpha_{22} \int v_2(x)f(x)g_1(x)dx$$

$$-\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx > -\alpha_{12} \int v_1(x)f(x)g_2(x)dx - \alpha_{21} \int v_2(x)f(x)g_2(x)dx.$$

If the two managers are identical in both effort cost and productivity, these two inequalities will be automatically satisfied when the compensation has strategic complementarity.

compensation must have already satisfied these restrictions and that the researcher's task is to identify the primitive parameters, for example, the costs of effort, from the data. In Section 4, I show that the parameters introduced so far in the mutual monitoring with total utility maximization model can be identified as mappings of the risk aversion parameter and quantities from the data-generating process; that is, extra constraints do not help identify the parameters used earlier.¹⁸ Even though these two extra constraints would provide more restrictions on the risk aversion parameter and might help us further shrink the set of the identified risk aversion parameter, assuming these two extra constraints are satisfied would not be a concern unless this model cannot be rejected, which is not found in this paper.

In the mutual monitoring with individual utility maximization model, the two separate incentive compatibility constraints state for each manager that 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). Equation (13) is the incentive compatibility constraint for manager 1, and (14) is for manager 2:

$$-\alpha_{12} \int v_1(x)f(x)dx \geq -\alpha_{11} \int v_1(x)f(x)g(x)dx, \quad (13)$$

$$-\alpha_{22} \int v_2(x)f(x)dx \geq -\alpha_{21} \int v_2(x)f(x)g(x)dx. \quad (14)$$

Maximizing individual utility implies that the two managers cannot transfer utility. As a result, compared with both working, unilateral shirking makes at least one manager worse off such that asymmetric effort strategy cannot be sustained in the equilibrium of this model. Consequently, shareholders are concerned only about the collusion in which both managers shirk.

In this model, the two participation constraints and the two incentive compatibility constraints are binding in equilibrium and make working a Pareto-dominant strategy for each manager. As a result, the Pareto frontier meets at the outside option. Note that both shirking is a Nash equilibrium in the managers' subgame due to the free rider problem. However, the payoff of shirking is no more than working in the coalition such that neither manager has an incentives to leave the coalition. Because the two managers cannot transfer utility, they will not deviate from the point they can reach under the current contract with a specific Pareto allocation weight on the managers' expected utilities. Note that the equal sharing rule/bargaining power applies here too; that is, the weight of the two managers' expected utility is the same.

¹⁸If exploiting these two extra constraints may change the prediction on the parameter value in the current model, it indicates another model rather than a model nested into the current one. That would suggest testing a new model, which is a task independent of what is done in this paper.

Again, all this mutual monitoring with individual utility maximization model describes is that no manager shirks even though there is a free rider opportunity and that working is preferred only as a Pareto-dominant strategy rather than as a Nash equilibrium strategy. Theoretical literature provides different mechanisms of mutual monitoring which guarantees that Pareto dominance is played in equilibrium. Though they appeal to different equilibrium concepts, they can be empirically consistent with the mutual monitoring with individual utility maximization model set up here, 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), the infinitely repeated game with implicit side contracts in Che and Yoo (2001), leadership by setting examples in Hermalin (1998), and Kandel and Lazear (1992) who model peer pressure, among others. Ideally, if there is sufficient data, we may be able to distinguish between those incentive mechanisms; however, doing so is neither possible given the data available to this paper nor the focus here. In the Extension Section, I discuss in detail to what extent an alternative model can be identified, which is empirically consistent with the mutual monitoring with individual utility maximization model, and features a trigger strategy in repeated play with the rent of stay.

2.5 Optimal Contracts

The shareholders' cost minimization problem subject to the participation constraints and the incentive compatibility constraints has a Lagrangian formulation. Thus the optimal contract can be derived by solving the first-order conditions of the shareholders' constrained optimization problem. The following proposition gives the optimal contract under each model. Note that α_{ij} and $g_i(x)$ are the same as previously defined, μ_1 is the shadow price associated with manager 1's incentive compatibility constraint and μ_2 with manager 2's, $w_i^*(x)$ is the optimal compensation paid to manager i .

Proposition 2

$$w_1^*(x) = \frac{1}{\rho} \ln \alpha_{12} + \frac{1}{\rho} \ln \left[1 + \mu_1 - \mu_1 \left(\frac{\alpha_{11}}{\alpha_{12}} \right) g_1(x) \right], \quad (15)$$

$$w_2^*(x) = \frac{1}{\rho} \ln \alpha_{22} + \frac{1}{\rho} \ln \left[1 + \mu_2 - \mu_2 \left(\frac{\alpha_{21}}{\alpha_{22}} \right) g_2(x) \right]. \quad (16)$$

In the no mutual monitoring model, $w_i^(x)$ has exactly the same expression in (15) and (16). In the mutual monitoring with total utility maximization model, $\mu_1 = \mu_2 \equiv \mu$, and $g_1(x)$ and $g_2(x)$ are replaced by $g(x)$. In the mutual monitoring with individual utility maximization model, only $g_1(x)$ and $g_2(x)$ are replaced by $g(x)$.*

The intuition is as follows. In the no mutual monitoring model, the incentives are based on each manager's 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 other two models of mutual monitoring, the optimal contract merely prevents 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 $g(x)$. Furthermore, in the mutual monitoring with total utility maximization model, μ_1 and μ_2 are equal because of the group incentive compatibility constraint. In the mutual monitoring with individual utility maximization model, μ_1 and μ_2 are not the same because the incentive compatibility constraint is individually specified.

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 $g(x)$) together explain the compensation shape of each manager. Relative features of the two managers' compensation schemes can be rationalized by any of these three models, depending on the values of the preceding primitive parameters. This again confirms that the relative properties between the two managers' compensations are not sufficient to distinguish the three models, which are sharply distinct in terms of whether and how shareholders consider the mutual monitoring at optimal compensation design.

Three more points can help us understand the form of the optimal contracts. First, each manager gets his highest compensation denoted by $w_i(\bar{x})$ when the informativeness of corresponding output realization is highest, i.e. $g_i(\bar{x}) = 0$ or $g(\bar{x}) = 0$, given that the shadow price and disutility coefficients are all positive. Second, if the managers' efforts are observable to shareholders, $g_i(x)$ or $g(x)$ 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 their effort choice of working, and the shadow price of 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. Third, the 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 $g(x)$ is smaller, he gets higher compensation at that signal, keeping all other things constant.

2.6 Shareholder's Profit Maximization

Shareholders also need to compare the expected net benefits among different effort pairs and guarantee that motivating both managers to work is indeed better than motivating other effort pairs. This is the first step in the analysis of Grossman and Hart (1983). From shareholders' viewpoint, the 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 previously denoted by V , with the gross abnormal return x and then taking expectation over the distribution of x conditional on the two managers' effort choices in that period; that is, $E[V * x \mid j, k]$.

Shareholders' cost is the total compensation paid to the two managers. Denote w_i^s as the optimal fixed compensation paid to 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 normalized to -1 , and on the other side is 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

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

Shareholders pay the two managers to deliver efforts and benefit from the growth in firm value. Consequently, the net profit of motivating a particular effort pair is the expected residual of the firm value growth deducted by the compensation cost. The expectation is conditional on the managers' effort choice. The optimal effort pair to be implemented in the three models is the same, that is, both managers work. However, the suboptimal benchmark effort pairs are different. In the no mutual monitoring model, the suboptimal effort pair is that no more than one manager works. Thus motivating both managers to work is preferred by the shareholders if and only if

$$E[V * x - w_1^*(x) - w_2^*(x)] \geq \max \left\{ \begin{aligned} &E[(V * x - w_1^s - w_2^*(x)) * g_1(x)], \\ &E[(V * x - w_1^*(x) - w_2^s) * g_2(x)], \\ &E[(V * x - w_1^s - w_2^s) * g_1(x) * g_2(x)] \end{aligned} \right\}. \quad (17)$$

On the right-hand side of the preceding inequality, the first (second) term reflects the shareholders' net benefit of having only manager 1 (2) shirking. The third item is the shareholders' net benefit of having both managers shirking.

By contrast, In the mutual monitoring with total utility maximization model and the mutual monitoring with individual utility maximization model, there is only one benchmark effort pair, that is, both managers shirk. As a result, shareholders prefer motivating both managers to work if and only if the net benefit is higher by doing so than by taking the alternative, that is,

$$E[V * x - w_1^*(x) - w_2^*(x)] \geq E[(V * x - w_1^s - w_2^s) * g(x)].$$

2.7 Summarizing the Three Models

Before moving to the empirical implementation, I summarize the key differences between the three models. This comparison will guide the identification procedure and the model specification test in later sections. Depending on whether shareholders exploit mutual monitoring in the optimal compensation design and whether the two managers monitor each other as a unitary decision maker or as individual decision makers, the three models differ in the participation constraint, the incentive compatibility constraint, and the suboptimal benchmark in the shareholders' profit maximization problem.

If shareholders do not take advantage of the mutual monitoring between the two managers, the no mutual monitoring model characterizes this case. In this model, the participation constraint is specified for each manager, depending on each manager's differentiated marginal influence on the distribution of gross abnormal return. The incentive compatibility constraint is separately specified for each manager as well. The two managers choose working in a Nash equilibrium. The likelihood ratio associated with each manager's suboptimal effort is differentiated between the two managers. Also, the shadow price of each manager's incentive compatibility constraint is distinct. To maximize the net profit, shareholders compare between both managers working against at least one manager shirking.

If shareholders take advantage of mutual monitoring that managers can enforce through side contracts, the other two models fit this class. Shareholders are only concerned about both managers shirking. Furthermore, if the two managers choose efforts collectively, the mutual monitoring with total utility maximization model characterizes this case. Both the participation constraint and the incentive compatibility constraint are based on the total utility of the two managers. This model requires both the likelihood ratio and the shadow price of the incentive compatibility constraint to be symmetric between the two managers. Otherwise, if the two managers only pursue self-interest, the mutual monitoring with individual utility maximization model characterizes this case. The participation constraint and incentive compatibility constraint are specified for each manager. Shareholders again only

have to prevent the managers from both shirking. As a result, this model does not require the shadow price to be equal but requires the likelihood ratio to be symmetric.

3 Data

This section discusses the data source and the construction of key variables in the empirical implementation of this paper. The sample period covers 1993 to 2005. The firm characteristic data come from the COMPUSTAT North America database. The stock returns are from CRSP and Compustat PDE. The top executive compensation data come from the ExecComp database.

3.1 Heterogeneity in the Data

In my framework, 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.

Following are the detailed procedures to categorize observations. First, I classify the whole sample into three industrial sectors according to the Global Industry Classification Standard (GICS) code, denoting by S_{nt} the n th firm in year t . The primary sector ($S_{nt} = 1$) includes firms in energy (GICS: 1010), materials (GICS: 1510), industrials (GICS: 2010, 2020, 2030), and utilities (GICS: 5510). The consumer good sector ($S_{nt} = 2$) includes firms in consumer discretionary (GICS: 2510, 2520, 2530, 2540, 2550) and consumer staples (GICS: 3010, 3020, 3030). The service sector ($S_{nt} = 3$) includes firms in health care (GICS: 3510, 3520), financial (GICS: 4010, 4020, 4030, 4040), and information technology and telecommunication services (GICS: 4510, 4520, 5010). Next, in each industrial sector, I classify the firms based on 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 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} . I denote firm type as $Z_{nt} = (A_{nt}, D/E_{nt})$, which has four combinations of A_{nt} and D/E_{nt} .

In Table 1, I summarize 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 smaller book values than those in the primary or service sector. If compared based on market value, the three sectors have close market

values. The debt-to-equity ratio reflects the firms' capital structure. It has the highest value in the service sector and the biggest standard deviation as well. The yearly abnormal return of a firm is calculated by subtracting a market portfolio return from the firm's monthly compounded return for a given fiscal year. The abnormal return is not significantly different from zero in any sector.

[INSERT TABLE 1 HERE]

3.2 Key Variables in the Optimal Contracts

3.2.1 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. I drop firm-year observations if the firm changed its fiscal year end such that all compensations and stock returns are 12-month based.

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.

[INSERT TABLE 2 HERE]

3.2.2 Compensation

When managers make effort decisions, they care about their overall wealth change implied by their compensation packages. In the ExecComp database, available are salary, bonus, other annual compensation not properly categorized as salary and bonus, restricted stock granted during the year, aggregate value of stock options granted during the year as valued using S&P's Black-Scholes methodology, amount paid under the company's long-term incentive plan, and all other compensation. However, managers' wealth varies with their holdings in firm-specific equity as well. They can always offset the aggregate risks imposed in their compensation package by adjusting with a market portfolio but cannot avoid being exposed to nondiversifiable risks of holding firm stocks and options. As a result, managers' wealth changes in holding firm-specific equity are incorporated into total compensation given that they cannot diversify those idiosyncratic risks. Following the concept of wealth change initiated by Antle and Smith (1985, 1986),¹⁹ I construct the total compensation by adding

¹⁹Among followers are Hall and Liebman (1998), Margiotta and Miller (2000), Gayle and Miller (2009, 2012), and Gayle et al. (2011).

wealth change from holding options and wealth change from holding stocks into all regular components provided in the database. These wealth changes can be interpreted as opportunity costs of holding firm-specific equity. Consequently, the wealth change from holding stocks is equal to the beginning shares of held stocks multiplied by the abnormal returns. By holding the options from existing grants rather than disposing of this part of wealth into a market portfolio, the manager obtains the difference between the ending option value and the beginning option value multiplied by the market portfolio return.

The two managers studied in this paper are the two highest paid executives based on the total compensation. Table 2 describes their compensation cross-sectionally. In all types of firms (classified by firm size and capital structure), 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 (either small or large), 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.

Table 3 summarizes the time-series properties of the key components of the total compensation. A few things stand out. First, the compensation is heavily equity based for both managers. The sum of the four equity-based components, that is, the values of restricted stocks, values of granted options, changes in wealth from stocks held, and changes in wealth from options held, on average accounts for more than 80 percent of the total compensation. Second, the opportunity costs of holding firm-specific equity are significantly positive and similarly high for both managers. This indicates that the potential nonpecuniary or non-contractible benefits of holding the stocks or options from the current firm are large for the two highest paid managers. Third, the variation of the total compensation across years is not negligible for either manager. This suggests that it is necessary to take into account the effect of the macroeconomic fluctuation on the compensation schemes.

[INSERT TABLE 3 HERE]

Table 4 reports the position profiles of the two managers. I classify the positions held by the two highest paid managers into three categories. I count the frequency of holding positions of certain categories as follows. "Functional" = 1 if the manager holds the position of CTO, CIO, COO, CFO, or CMO, but not any other; otherwise, "Functional" = 0. "General 1" = 1 if the manager holds the position of chairman, president, CEO, or founder, but not any other; otherwise, "General 1" = 0. "General 2" = 1 if the manager holds the position of executive vice-president, senior vice-president, vice-president, vice-chair, or other

(defined in the database), but not any other; otherwise, "General 2" = 0. "Functional & General 1" = 1 if the manager simultaneously holds at least one position from each of the Functional category and the General 1 category but none from the General 2 category; otherwise, "Functional & General 1" = 0. The same rule applies to "Functional & General 2" and "General 1 & General 2." "Functional & General 1 & General 2" = 1 if the manager holds at least one position from each of the three categories; otherwise, the indicator equals zero.

[INSERT TABLE 4 HERE]

I first analyze the primary sector. The top three rows of Table 4 describe for each manager the frequency of holding positions of only one category. Both of the two managers rarely hold only the functional position. The highest paid managers have a larger chance to sit on the top rank of the general position (General 1), and by contrast, the second highest paid managers have a larger chance to sit on the low rank (General 2).

The three rows in the middle describe the two managers' title distributions when each manager holds a position from only two categories in the same year. Comparing the top two rows of the middle three with the row of "Functional" on the very top suggests that the chance of managers to obtain high compensation from holding one more general position in addition to the functional position is larger for the second highest paid managers than for the highest paid managers. In contrast, the bottom row of the three shows that the highest paid managers are more likely those who hold two general positions. In other words, holding a general position helps managers more to get higher compensation.

The very bottom row in Table 4 presents a very similar distribution feature as what is shown in the very top row for holding a functional position only. Here both managers rarely hold positions from all of the three types. The consumer goods sector and the service sector have exactly the same pattern as what was discussed previously for the primary sector.

3.2.3 Measurement Error

To be consistent with the theoretical implication of the performance measure and payment, the abnormal returns and total compensation need further adjustment. First, the performance measure in the optimal contract should be closely tied to managers' effort but eliminate the stochastic disturbances that are out of managers' control. Second, the performance measure should reflect the notion of output sharing between shareholders and managers and thus needs to incorporate compensation payments. Taking into account these two points, I construct the performance measure, or the *gross abnormal return*, as I call it, in the following steps. First, I subtract market portfolio return from the annual return to a firm stock in the same corresponding fiscal year and thus get the residual that captures the idiosyncratic

components in stock returns. This nondiversifiable portion generates working incentives. Given that either the gross abnormal return or the optimal compensation cannot be directly observed from the data, I construct their consistent estimators as discussed later. Here \tilde{x}_{nt} is the abnormal return and \tilde{w}_{int} is manager i 's total compensation from firm n in year t . (Z_{nt}, S_{nt}) are firm type variables, defined previously. I nonparametrically estimate the optimal compensation $w_{int}(x_{nt}|Z, S)$ using a kernel regression (see Appendix B for details):

$$w_{int}(x_{nt}|Z, S) = E_t[\tilde{w}_{int}|\tilde{x}_{nt}, V_{n,t-1}, Z_n, S_n], \quad i = 1, 2,$$

where $V_{n,t-1}$ is the market value of firm n at the end of year $t - 1$. Then I calculate the gross abnormal returns as

$$x_{nt} \equiv \tilde{x}_{nt} + \frac{w_{1nt}}{V_{n,t-1}} + \frac{w_{2nt}}{V_{n,t-1}}.$$

Then the PDF of gross abnormal return x_{nt} , that is, $f(x_{nt}|Z, S)$, is nonparametrically estimated as well by a kernel estimator.

3.3 Bond Prices and a Dynamic Consideration

In the static models, managers' outside options are 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.

The effort-dependent utility function defined in (1) and (2) now has a new expression:

$$-\alpha_{ij}^{\frac{1}{b_t-1}} E_t \left[\exp \left(\frac{-\rho w_{it}(x_t)}{b_{t+1}} \right) \mid j, k \right], \quad (18)$$

where b_t is the bond price in year t , which pays a unit of consumption per period forever.²⁰ 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 his life-time 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, but now it is $[b_{t+1}/\rho(b_t - 1)] \ln \alpha_{ij}$ in the dynamic version. I update the participation constraints and incentive compatibility constraints in the static models with the new utility function. This

²⁰See the detailed construction of the bond prices in Gayle and Miller (2009, page 1748-1749).

reinterpretation makes the models fit the framework of Margiotta and Miller (2000).²¹ The same treatment is used by Gayle and Miller (2012, page 26). In the following sections, I adopt the dynamic version of the three models to develop the identification and hypothesis tests.

4 Identification

This section establishes the identification for each model laid out in Section 2. I first briefly recap what variables have been introduced into the three models of principal–multiagent moral hazard, and then I classify these elements in the models into observables and unobservables from the perspective of researchers rather than the players in the models.

First, I introduce the technologies that are captured by the distribution of the gross abnormal returns, respectively, when both managers choose equilibrium actions and when they deviate from the equilibrium path. Then I specify the information asymmetry between shareholders and the two managers, that is, managers’ efforts are unobservable to shareholders but observable between the managers. Second, I specify managers’ preferences by parameterizing two CARA utility functions with a common risk aversion parameter and different disutility coefficients of effort. I specify the shareholders’ preferences by embedding a constrained cost-minimization problem into their selection of managers’ effort pairs to maximize the net profit. Given these primitive preferences and distributions parameterized as stated, we can perfectly predict the endogenous decisions that are made within the model by shareholders (compensation design that specifies the managers’ compensation as a function of the gross abnormal return) and by managers (choice among rejecting the job offer, shirking, and working).

Before classifying the observables and unobservables, I make an assumption on the players’ behavior in equilibrium for identification purposes; that is, shareholders are assumed to prefer both managers working, and the two managers are assumed to indeed work, as the optimal contracts intend to implement. These assumptions are natural. Because overall firms have been ongoing, it seems unlikely that the top executives shirked in general. Also, the top managers’ compensation is heavily tied to the stock returns and thus not flat, which would contradict the prediction if shareholders prefer managers shirking and simply pay them with constant wages, provided the moral hazard exists.

Given the above assumption on behavior, the optimal compensation schemes and the distribution of the gross abnormal returns conditional on managers’ equilibrium actions are assumed to be observable with measurement errors and thus can be nonparametrically

²¹This paper is descended from Grossman and Hart (1983) and Fudenberg et al. (1990).

identified from the data. The unobservable primitive elements left for researchers to identify 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.

Along with the behavioral assumption earlier made and some regularity conditions, the equilibrium restrictions, for example, the first-order conditions derived in the Lagrangian formulation of the shareholders' optimization problem (corresponding to Grossman and Hart's (1983) second step) and restrictions implied by shareholders' preferences over the optimal effort level (corresponding to their first step), can be used to derive the mappings from the joint distribution of the observables to the distribution of a random quantity that is the function of unobservable primitive elements. Such mappings bridge between observables and unobservables and thus essentially help us identify the model.

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 I hope to test how well each entire model can rationalize the data of executive compensation and abnormal 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, I need to go further to identify and estimate all the unobservable primitive elements (Matzkin 2007). To fulfill the first task, this paper takes three steps for each model, as follows.

In step 1, for one model, I assume that the risk aversion parameter is known and then show that all other primitive parameters in that model can be identified. Given the behavioral assumption I make, managers play the equilibrium strategies (both work) as shareholders desire. If the data of compensation and stock returns are generated by a model, the density of gross abnormal returns conditional on optimal effort choice $f(x)$ and the equilibrium compensation scheme $w_i(x)$ of manager i can be nonparametrically identified directly from the empirical distribution of the data. The optimal contract implies that both participation constraints and incentive compatibility constraints are binding. The first-order conditions in the Lagrangian formulation of the shareholders' cost minimization problem together with some regularity conditions on the likelihood ratios allow me to derive each structural parameter as a mapping of the risk aversion parameter and some quantities from the data generating process.

In step 2, I exploit other restrictions implied by the model to bound the risk aversion

parameter. These restrictions include the shareholders' preferences over managers' efforts (in inequalities) and other restrictions (in inequalities or equalities) tailored to each model. The mix of equality and inequality restrictions prevents the risk aversion parameter from point identification. Instead, I use these restrictions to delimit the identified set of this parameter. These extra restrictions, along with the mappings derived in the first step, characterize the identified set of the risk aversion parameters.

These equilibrium restrictions constitute a function $Q(\rho, x, w)$ in which the risk aversion parameter is the only unknown that is left to be identified and estimated. The $Q(\rho, x, w)$ function has a distance-minimizing feature; that is, if the data are generated from a process that can be rationalized by the model and by the true value of the risk aversion parameter ρ^* , we should have $Q(\rho^*, x, w) = 0$. To identify the model and estimate the risk aversion parameter, I search for a range of the risk aversion parameter that asymptotically satisfies this equation.

In step 3, I construct a hypothesis test for the model based on the identified set of the risk aversion parameter that indexes each model. Using a subsampling algorithm, I obtain a consistent estimate of the 95 percent 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. Otherwise, we can reject the null hypothesis that this model generates the data. Consequently, the confidence region of the risk aversion parameter provides a criterion on whether the model is rejected. Thus the estimation and the hypothesis test are accomplished at the same time. In the following, I discuss the detailed identification and test for each model.

4.1 No Mutual Monitoring Model

The unobservable structural parameters in the no mutual monitoring model include each manager's effort preference over working and shirking relative to his outside option (denoted by α_{ij} , which is the effort disutility coefficient in manager i 's utility functions when he chooses effort level j), the likelihood ratio of the distribution if manager i shirks over that if both managers work (denoted by $g_i(x)$, and the subscript t in x_t is dropped hereafter when it does not cause confusion), and the risk aversion parameter ρ . I assume the data of compensation and stock returns are repeatedly cross-sectional independent draws from the equilibrium of this model. As a result, $f(x)$ can be identified directly from the empirical distribution of the gross abnormal returns using a nonparametric density estimator. Also, following the same logic, the optimal compensation can be nonparametrically identified from the data as well. Then I show that those unobservable structural parameters can be sequentially derived as

mappings of the risk aversion parameter and the observables.

First, I consider the disutility coefficients of working, that is, α_{i2} for $i = 1, 2$. 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 is indifference between rejecting the job to pursue an outside option and accepting the offer and working diligently during the following period. In the economic model, this means that the participation constraint in the shareholders' optimization problem is binding, that is, each manager's expected utility conditional on his subsequent effort choice (working) is equal to the value of his outside option, which is normalized to be -1 .

Rearranging the terms in the dynamic version of the (4) and (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:

$$\alpha_{12}(\rho) = E_t[v_{1t}(x, \rho)]^{1-b_t}, \quad (19)$$

$$\alpha_{22}(\rho) = E_t[v_{2t}(x, \rho)]^{1-b_t}. \quad (20)$$

Next, I consider the likelihood ratios $g_{it}(x)$ for $i = 1, 2$. In the formula of optimal compensation (15) and (16), 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, $\lim_{x \rightarrow \infty} g_{it}(x) = 0$, then $\bar{w}_{it} \equiv w_{it}(\bar{x}_{it})$ satisfying $g_{it}(\bar{x}_{it}) = 0$ 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:

$$g_{1t}(x, \rho) = \frac{1/v_{1t}(x, \rho) - 1/v_{1t}(\bar{x}_1, \rho)}{E[1/v_{1t}(x, \rho)] - 1/v_{1t}(\bar{x}_1, \rho)}, \quad (21)$$

$$g_{2t}(x, \rho) = \frac{1/v_{2t}(x, \rho) - 1/v_{2t}(\bar{x}_2, \rho)}{E[1/v_{2t}(x, \rho)] - 1/v_{2t}(\bar{x}_2, \rho)}. \quad (22)$$

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 $g_{it}(\bar{x}_i, \rho) = 0$, 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 I consider the disutility coefficients of shirking, that is, α_{i1} for $i = 1, 2$. When

shareholders design the optimal contracts to induce both managers to work, they need to provide sufficient incentive through the compensation not only to induce the managers to stay in the firm but also to motivate them to exert effort in the shareholders' interests. As a result, the optimal compensation schemes should make each manager's expected utility from working and receiving the monetary compensation at the end of the period the same as his expected utility from shirking during the following period. In the economic model, this means that the incentive compatibility constraint in the shareholders' optimization problem is binding for each manager. In the econometric model, the data generated from this model satisfy the two equalities held in the incentive compatibility constraints (9) and (10) as well as the two equalities held in the participation constraints (4) and (5). These together help us derive the disutility coefficients $\alpha_{11}(\rho)$ and $\alpha_{21}(\rho)$ as the mappings of the risk aversion parameter, as follows:

$$\alpha_{11}(\rho) = E_t[v_{1t}(x, \rho)g_{1t}(x, \rho)]^{1-b_t}, \quad (23)$$

$$\alpha_{21}(\rho) = E_t[v_{2t}(x, \rho)g_{2t}(x, \rho)]^{1-b_t}. \quad (24)$$

Again, these two formulas imply that 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 $g_{1t}(\bar{x}_1) = 0$ in the formula of the optimal compensation $w_{1t}^*(x)$ in (15) and evaluate both sides at \bar{x}_1 . Note that on the left-hand side of that formula, $w_{1t}(\bar{x}_1)$ can be identified and estimated by the highest compensation that manager 1 receives. On the right-hand side, given that the disutility coefficients have been identified as previously and $g_{1t}(\bar{x}_1)$ drops off, only the shadow price μ_1 and the risk aversion parameter are left unknown. The same procedure applies to identifying the shadow price for manager 2 (μ_2). Consequently, the two shadow prices can be expressed as the mappings of the risk aversion parameter, as follows:

$$\mu_1(\rho) = E_t[v_{1t}(x, \rho)]/v_{1t}(\bar{x}, \rho) - 1, \quad (25)$$

$$\mu_2(\rho) = E_t[v_{2t}(x, \rho)]/v_{2t}(\bar{x}, \rho) - 1. \quad (26)$$

Collectively, all primitives 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 no mutual monitoring

model to delimit the identified set 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 benefit of motivating both managers to work is the expected increase in the equity value of the firm in the contract period. Recall the mathematical expression of this profit maximization preference in (17). 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 (27), (28), and (29). 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:

$$\Lambda_{1t}(\rho) = E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[(V * x - w_{1t}^s - w_{2t}^*(x)) * g_{1t}(x, \rho)] \geq 0, \quad (27)$$

$$\Lambda_{2t}(\rho) = E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[(V * x - w_{1t}^*(x) - w_{2t}^s) * g_{2t}(x, \rho)] \geq 0, \quad (28)$$

$$\begin{aligned} \Lambda_{3t}(\rho) &= E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[(V * x - w_{1t}^s - w_{2t}^s) * g_{1t}(x, \rho) * g_{2t}(x, \rho)] \\ &\geq 0, \end{aligned} \quad (29)$$

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,

$$w_{it}^s = \frac{b_{t+1}}{\rho(b_t - 1)} \ln \alpha_{i1}(\rho), \text{ for } i = 1, 2. \quad (30)$$

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 inequality in (31) ((32)) following reflects this restriction for manager 1 (2). The first term of the top (bottom) line 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:

$$\begin{aligned} \Lambda_{4t}(\rho) &= \left\{ -\alpha_{12}(\rho)^{\frac{1}{b_t-1}} E[v_{1t}(x, \rho)g_{2t}(x, \rho)] \right\} \\ &\quad - \left\{ -\alpha_{11}(\rho)^{\frac{1}{b_t-1}} E[v_{1t}(x, \rho)g_{1t}(x, \rho)g_{2t}(x, \rho)] \right\} > 0, \end{aligned} \quad (31)$$

$$\begin{aligned} \Lambda_{5t}(\rho) &= \left\{ -\alpha_{22}(\rho)^{\frac{1}{b_t-1}} E[v_{2t}(x, \rho)g_{1t}(x, \rho)] \right\} \\ &\quad - \left\{ -\alpha_{21}(\rho)^{\frac{1}{b_t-1}} E[v_{2t}(x, \rho)g_{1t}(x, \rho)g_{2t}(x, \rho)] \right\} > 0. \end{aligned} \quad (32)$$

The third source of equilibrium restrictions comes from the requirement that the likelihood ratio $g_i(x)$ be nonnegative. Recall the identification of \bar{x}_i , which is obtained by satisfying $g_i(\bar{x}_i) = 0$; $\forall x > \bar{x}_i$ ($i = 1, 2$), the formula of $g_i(x)$ in (21) and (22) is guaranteed to be nonnegative. However, the product $g_1(x)g_2(x)$ is another likelihood ratio such that the following restriction must be satisfied:

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

Collectively, the preceding restrictions implied by the no mutual monitoring model can be summarized by a function $Q_{N-M}(\rho)$ as

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

Note that the $Q_{N-M}(\rho)$ 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, $Q_{N-M}(\rho)$ 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 defined as

$$\Gamma_{N-M} \equiv \{\rho > 0 : Q_{N-M}(\rho) = 0\}. \quad (33)$$

4.2 Mutual Monitoring with Total Utility Maximization Model

The intuition of the identification here is similar to that for the no mutual monitoring model. The only departure is that the two differences between the no mutual monitoring model and the mutual monitoring with total utility maximization model lead to two extra restrictions

that the risk aversion parameter needs to satisfy.

The first difference is that the two managers are now motivated as a single agent. This implies that the shadow prices of the two managers' incentive compatibility constraints are no longer differentiable and thus (25) and (26) are equal such that

$$E_t [v_{1t}(x)] / v_{1t}(\bar{x}) - 1 = E_t [v_{2t}(x)] / v_{2t}(\bar{x}) - 1.$$

Consequently, the two managers' incentive compatibility constraints have the same shadow price in the shareholders' optimization problem. Using (25) and (26), the shadow price associated with manager 1 (2) enters into the following equality restriction as the first (second) term:

$$\Psi_{2t}(\rho) = \frac{E_t[v_{1t}(x, \rho)]}{v_{1t}(\bar{x}, \rho)} - \frac{E_t[v_{2t}(x, \rho)]}{v_{2t}(\bar{x}, \rho)} = 0.$$

The second difference is that shareholders contrast the optimal effort pair (both working) with both shirking. This implies that the two managers' compensation schemes have the same informative inference to back out the likelihood ratio $g_t(x)$. I ensure that the two likelihood ratios are equal in unit mass by imposing the following restriction:

$$\Psi_{3t}(\rho) = E_t[\mathbf{1}\{g_{1t}(x, \rho) = g_{2t}(x, \rho)\} - 1] \geq 0,$$

where $\mathbf{1}\{g_{1t}(x, \rho) = g_{2t}(x, \rho)\}$ is an index function equal to 1 if the condition is satisfied and zero otherwise.²²

I further explore other restrictions implied by the mutual monitoring with total utility maximization model to bound the identified set of the risk aversion parameters, which appeal to shareholders' preferences on the optimal effort level. Compared with the same set of restrictions in the no mutual monitoring model, the difference here is that the suboptimal effort level as a benchmark becomes both managers shirking, meaning that the shareholders prefer incentivizing both managers working to both shirking. The top (bottom) inequality expresses this restriction using the likelihood ratio with respect to manager 1 (2)'s compensation:

$$\begin{aligned} \Lambda_{6t}(\rho) &= E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[V * x * g_1(x, w_1, w_2) - w_{1t}^s - w_{2t}^s] \geq 0, \\ \Lambda_{7t}(\rho) &= E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[V * x * g_2(x, w_1, w_2) - w_{1t}^s - w_{2t}^s] \geq 0, \end{aligned}$$

where the fixed compensation paid to both managers if the shareholders prefer them shirking

²²This function-wise restriction is constructed in a way similar to the nonnegative restriction on likelihood ratio imposed in Gayle and Miller (2012).

(w_{it}^s) is the same as previously defined.

Define $Q_{M-T}(\rho)$ as subsequently to collect all the preceding restrictions implied by the mutual monitoring with total utility maximization model. It has the same distance-minimizing feature and has the following expression:

$$Q_{M-T}(\rho) \equiv \sum_{t=1}^T \left\{ \sum_{l=6}^7 [\min(0, \Lambda_{lt}(\rho))]^2 + [\Psi_{2t}(\rho)]^2 + [\min(0, \Psi_{3t}(\rho))]^2 \right\}.$$

Then the risk aversion parameter admissible to this model belongs to the set defined as

$$\Gamma_{M-T} \equiv \{\rho > 0 : Q_{M-T}(\rho) = 0\}. \quad (34)$$

4.3 Mutual Monitoring with Individual Utility Maximization Model

Compared with the mutual monitoring with total utility maximization model, shareholders still compare between symmetric effort pairs but individualize the incentive for each manager. To highlight the difference between this model and the one with total utility maximization, here the shadow price of the incentive compatibility constraint for each manager is distinct. As a result, the associated restriction once used in the mutual monitoring with total utility maximization model is dropped, that is, $\Psi_{2t}(\rho)$. However, similar to the mutual monitoring with total utility maximization model, the two compensation schemes of the two managers have the same inference about the likelihood ratio because the contract is based on symmetric effort only. Thus the associated restriction maintains, that is, $\Psi_{3t}(\rho)$.

Collecting the restrictions implied by the mutual monitoring with individual utility maximization model as

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

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

$$\Gamma_{M-I} \equiv \{\rho > 0 : Q_{M-I}(\rho) = 0\}. \quad (35)$$

4.4 Summary of the Identification Results

For each model, all primitives introduced into the econometric model can be recovered from the data generating process along with the risk aversion parameter. Denote $M \in \{\text{N-M}$

(no mutual monitoring model), M-T (mutual monitoring with total utility maximization model), M-I (mutual monitoring with individual utility maximization model)}. Denote the set of structural parameters by

$$\begin{aligned}\theta_{N-M} &\equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_{1t}(x), g_{2t}(x), \mu_1, \mu_2), \\ \theta_{M-T} &\equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_t(x), \mu), \\ \theta_{M-I} &\equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_t(x), \mu_1, \mu_2).\end{aligned}$$

The following proposition formally states this result.

Proposition 3 *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 $(x_t, w_{it}, \bar{w}_{it})$ for $i = 1, 2$, that is, $\theta_M^* = \theta_M(\rho^*)$.*

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 $Q_M(\rho)$ 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. Intuitively, if the model can rationalize the data, there must exist some nonnegative values of the risk aversion parameter such that the data restrictions embedded in the function $Q_M(\rho)$ are satisfied. In other words, the corresponding set Γ_M is nonempty. Formally, the following proposition establishes that the restrictions implied by model M set a sharp and tight bound for the identified set of the risk aversion parameter.²³

Proposition 4 *Consider any data generating process (x_n, w_n) that satisfies $w_n = w(x_n)$ for $\forall n$. Define Γ_M as before for each $M \in \{N-M, M-T, M-I\}$. If Γ_M is not empty, then (x_n, w_n) is observationally equivalent to every data process generated by the model M parameterized by each $\rho \in \Gamma_M$. If Γ_M is empty, then (x_n, w_n) is not generated by the model M .*

5 Estimation and Tests

Recall that the $Q_M(\rho)$ function has a distance-minimizing feature. If the data are generated by the model M , the observables in the data should satisfy the equilibrium restrictions

²³A caveat is that the tight bound under the mutual monitoring with total utility maximization model asks for the assumption that both working strictly Pareto dominates unilateral shirking in the managers' subgame.

parameterized by the equalities and inequalities implied by the model. Mathematically, this means that there must exist some nonnegative values of the risk aversion parameter ρ such that the population value $Q_M(\rho)$ is zero. As a result, I can define for each model M the null hypothesis and alternative hypothesis as

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

I calculate a sample analogue of $Q_M(\rho)$, denoted by $Q_{M,ZS}^{(N)}(\rho)$, for each firm type Z in each sector S by replacing each element in $Q_M(\rho)$ with its sample analogue. In particular, the expectation valued by an integral is consistently estimated by an average weighted by the corresponding kernel densities. Here $v_{it}(\bar{x}_{it})$ is replaced with $\exp\left(\frac{-\rho\bar{w}_{it}(\bar{x}_{it})}{b_{t+1}}\right)$, where $\bar{w}_{it} = \max\{w_{it}^1, \dots, w_{it}^{N_{ZS}}\}$ in the no mutual monitoring model, and is replaced with $\exp\left(\frac{-\rho\bar{w}_{it}(\bar{x}_t)}{b_{t+1}}\right)$, where $\bar{x}_t = \max\{\arg \max_x(w_{1t}(x)), \arg \max_x(w_{2t}(x))\}$, in the other two models with mutual monitoring. The value of $Q_{M,ZS}^{(N)}(\rho)$ is the sum of yearly equality and inequality restrictions within firm type Z and industrial sector S . Formally, I obtain the sample analogue of $Q_M(\rho)$ for each model $M \in \{\text{N-M}, \text{M-T}, \text{M-I}\}$ as follows:

$$\begin{aligned} Q_{\text{N-M},ZS}^{(N)}(\rho) & \equiv \sum_{t=1}^T \left\{ \sum_{l=1}^5 \left[\min(0, \Lambda_{it,ZS}^{(N)}) \right]^2 + \left[\Psi_{1t,ZS}^{(N)} \right]^2 \right\}, \\ Q_{\text{M-T},ZS}^{(N)}(\rho) & \equiv \sum_{t=1}^T \left\{ \sum_{l=6}^7 \left[\min(0, \Lambda_{it,ZS}^{(N)}) \right]^2 + \left[\Psi_{2t,ZS}^{(N)} \right]^2 + \left[\min(0, \Psi_{3t,ZS}^{(N)}) \right]^2 \right\}, \\ Q_{\text{M-I},ZS}^{(N)}(\rho) & \equiv \sum_{t=1}^T \left\{ \sum_{l=6}^7 \left[\min(0, \Lambda_{it,ZS}^{(N)}) \right]^2 + \left[\min(0, \Psi_{3t,ZS}^{(N)}) \right]^2 \right\}. \end{aligned}$$

Let us summarize the differences among the preceding three criterion functions. The suboptimal effort pair unfavorable to the shareholders is different between the no mutual monitoring model and the other two models incorporating mutual monitoring such that the restrictions corresponding to the shareholders' profit maximization are $\Lambda_{it,ZS}^{(N)}$ ($l = 1, 2, 3$) in the criterion function of the no mutual monitoring model but $\Lambda_{it,ZS}^{(N)}$ ($l = 6, 7$) in the other two models of mutual monitoring. The restriction on the uniqueness of Nash equilibrium is only required by the no mutual monitoring model, so its criterion function $Q_{\text{N-M},ZS}^{(N)}(\rho)$ includes two unique terms $\Lambda_{it,ZS}^{(N)}$ ($l = 4, 5$). The restrictions on the likelihood ratios generate the term $\Psi_{1t,ZS}^{(N)}$ in the no mutual monitoring model to guarantee that the likelihood ratio associated with both managers shirking satisfies the integral-to-one property. The mutual monitoring with total utility maximization model also has a unique restriction on the equalized shadow

prices of the two managers' incentive compatibility constraints, that is, $\Psi_{2t,ZS}^{(N)}$, because the incentive compatibility constraint is based on total utility. In the two models of mutual monitoring, 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 $\Psi_{3t,ZS}^{(N)}$.

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 in a frequency sense, the model should be rejected. Define the 95 percent 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 percent critical value of the test statistic. This value can be consistently estimated by the subsampling algorithm used in Gayle and Miller (2012), which is modified from Chernozhukov et al. (2007). Consequently, I reject the model M for firm type Z in sector S if the set $\Gamma_{M,ZS}^{(N)}$ is empty. If it is not empty, I obtain the 95 percent confidence region of the risk aversion parameter set.

6 Results

6.1 Estimation of the Risk Aversion Parameter and Tests

Table 5 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. The three panels in the table correspond to the three models. The column "Risk Aversion" reports the 95 percent confidence region of the identified set of the risk aversion parameter,

where a blank parenthesis means 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 value in the column "Risk Aversion."²⁴

[INSERT TABLE 5 HERE]

A comparison of confidence regions between the three models shows that the level of the estimated risk aversion parameter is highest under the no mutual monitoring model, is second highest under the mutual monitoring with individual utility maximization model, and is close to zero under the mutual monitoring with total utility maximization model when the sets are not empty. Note that for the same industrial sector and firm type, whenever, between the no mutual monitoring model and the mutual monitoring with individual utility maximization model, the confidence regions are not perfectly overlapped, the mutual monitoring with individual utility maximization 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 no mutual monitoring model in panel A of Table 5 as an example. Firstly, 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.

Second, 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

²⁴For 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. Thus the certainty equivalent to this gamble is $CE = \frac{-b}{\rho} \ln EU$.

no common interval of the confidence regions of the risk aversion parameter, which means that the no mutual monitoring 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 mutual monitoring with total utility maximization model, and in panel C, for the mutual monitoring with individual utility maximization model, I do the same analysis and report the common confidence regions subject to different levels of homogeneity of managers' risk preferences.

The main results from the estimation of the risk aversion parameter are summarized as follows. The no mutual monitoring 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 no mutual monitoring model cannot be rejected only in the service sector, which accommodates firms with a larger size and higher financial leverage. 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 this model is rejected if the managers are assumed to have homogeneous risk preferences.

The mutual monitoring with total utility maximization model is rejected in three types of firms because of the empty identified set of the risk aversion parameter, that is, large firms in the primary sector and small firms with high financial leverage in the service sector. However, when the identified set is not empty, the estimated confidence regions of the risk aversion parameter all cover values close to zero. This indicates that the mutual monitoring with total utility maximization model can rationalize the data in some types of firms but has to go with managers who are risk-neutral in an economic meaning. Such near risk neutrality contradicts the model itself, which assumes up front that managers are risk-averse and the moral hazard problem exists.²⁵ This contradiction rejects the mutual monitoring with total utility maximization model.

The mutual monitoring with individual utility maximization 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. This common interval covers a range lower than what single-agent models predict, but it is still at a reasonable level. A comparable result is found in Gayle and Miller (2012). In their paper, the estimated risk aversion parameter under a pure moral hazard model is lower than that under a hybrid moral hazard model in which the CEO has private

²⁵These assumptions rule out the possibility of achieving the first best allocation with risk neutral managers.

information about the firm’s states and shareholders pay a premium to induce truthful report. In their pure moral hazard model, the states of the firm are public information, and managers’ expected utilities are equalized across states such that the variation in CEOs’ compensation curvature is mitigated. Given that in the mutual monitoring with individual utility maximization model, the two managers have the same risk aversion parameter and same off-equilibrium distribution of the output, the results here can be compared with the two-states setting in Gayle and Miller (2012). Overall, the mutual monitoring with individual utility maximization model is more robust than the no mutual monitoring model in explaining the observed executive compensation which attempts to mitigate the moral hazard in top management teams.

6.2 Discussion

6.2.1 A Binary Illustration

Before comparing the results in pair of the models, 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 $j \in \{1 = \textit{shirk}, 2 = \textit{work}\}$ 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 $\textit{prob}(x|j)$, that is, $f(x)$ or $f(x)h(x)$ in the continuous case. In particular, $p \equiv \textit{prob}(x|\textit{work})$ and $q \equiv \textit{prob}(x|\textit{shirk})$; subscripts correspond to no mutual monitoring (N) or mutual monitoring (M).

Model	With/Without Mutual Monitoring	Without	With
$x \setminus j$	i work, $-i$ work	i shirk, $-i$ work	i shirk, $-i$ shirk
x_H	p	$q_N (< p)$	$q_M (< p)$
x_L	$1 - p$	$1 - q_N$	$1 - q_M$

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 $q \in \{q_N, q_M\}$ and $\{\alpha_{ij}\}_{i=1,2,j=1,2}$, the optimal compensation scheme for manager i satisfies the following in-

equality:

$$\begin{aligned}
& p * [-\alpha_{i2}e^{-\rho w_i(x_H)}] + (1 - p) * [-\alpha_{i2}e^{-\rho w_i(x_L)}] \\
\geq & q * [-\alpha_{i1}e^{-\rho w_i(x_H)}] + (1 - q) * [-\alpha_{i1}e^{-\rho w_i(x_L)}] \\
\implies & (\alpha_{i2}p - \alpha_{i1}q)e^{-\rho w_i(x_H)} \leq (\alpha_{i1} - \alpha_{i2} + \alpha_{i2}p - \alpha_{i1}q)e^{-\rho w_i(x_L)} \\
\implies & e^{-\rho[w_i(x_H) - w_i(x_L)]} \leq \frac{\alpha_{i1} - \alpha_{i2}}{\alpha_{i2}p - \alpha_{i1}q} + 1. \tag{36}
\end{aligned}$$

Note that the right-hand side of the last line is an amount negatively related to 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 two models of mutual monitoring have $q_M < q_N$ because the incentive compatibility constraint is relaxed owing to mutual monitoring and thus the suboptimal effort pair is both shirking. Given the binding incentive compatibility constraint (equality held in (36)) and fixed wage spread, ρ is expected to be smaller in these two models, which rationalize the same data as the no mutual monitoring model does.

6.2.2 No Mutual Monitoring versus Mutual Monitoring with Individual Utility Maximization

From the preceding binary example, if the risk aversion is fixed, the incentive is muted owing to using a contract with lower q . This implies that the compensation schemes that can be rationalized by a model using mutual monitoring tend to be flatter (i.e., smaller wage spread). If the mutual monitoring with individual utility maximization model is observationally equivalent to the no mutual monitoring model, shareholders seem to have adopted a wage spread larger than they are supposed to use. This tends to suggest a rejection of the model.

However, this is not true if the managers' risk aversion is actually lower than the level indicated by the no mutual monitoring model. In such a case, there is more demand of incentive in the contracts using mutual monitoring and thus a steeper compensation scheme is needed for less risk-averse managers. The estimated risk aversion parameter under the mutual monitoring with individual utility maximization model is indeed smaller than that under the no mutual monitoring model. The confidence regions of the mutual monitoring with individual utility maximization model cover the lower range of the nonoverlapped intervals between the two models in all firms in the primary sector and consumer goods

sector as well as in the large firms with low debt-to-equity ratio in the service sector. This is consistent with the preceding theoretical prediction. Thus the mutual monitoring with individual utility maximization model can rationalize the data. Because the variation in the inference of two managers' compensation about the same joint signal is attributed to the effort disutilities, the managers appear to be less risk averse.

Recall that the shareholders' profit maximization restriction plays a key role in delimiting the identified set of risk aversion parameters. The difference in the suboptimal effort benchmark explains the two models' different predictions for the risk aversion parameter. As the assumption on homogeneous risk preference is strengthened, shareholders' net profit in implementing the optimal effort pair (both managers work) shrinks correspondingly. This is a welfare explanation for the rejection of the no mutual monitoring model when more restrictive assumptions on homogeneous risk preference are taken.

Also, within the single output framework, the specification of each manager's individual contribution to the joint performance in the no mutual monitoring model demonstrates that individual incentives can be provided even with one single output. This rules out the possibility that the mutual monitoring with individual utility maximization model outperforms simply because individual incentives are not plausible. Instead, the outperforming tends to suggest that the shareholders indeed recognize the comparative advantage of using mutual monitoring to mitigate the moral hazard in top management teams.

6.2.3 No Mutual Monitoring Model versus Mutual Monitoring with Total Utility Maximization

One key feature of the mutual monitoring with total utility maximization model is that it equalizes the expected utility between the two managers both on and off the equilibrium path. From the researcher's viewpoint, when a model of less volatility in managers' utility payoffs across effort levels can reconcile the observed curvature of the compensation schemes, the managers will unsurprisingly appear to be less risk averse. In such a sense, the risk aversion parameter is expected to be lower under the mutual monitoring with total utility maximization model. To see this from the binary example, the mutual monitoring with total utility maximization model makes ρ and q decrease on both sides such that the wage spread may maintain. To rationalize the data, the no mutual monitoring model works with high values in both ρ and q , and the mutual monitoring with total utility maximization model works with low values in both ρ and q . Even if the group incentive works well, it is not necessary that the contract be flat. A small ρ is not inconsistent with the observed nonflat compensation scheme.

However, the mutual monitoring with total utility maximization model is rejected because

the value of the risk aversion parameter, which makes the model able to rationalize the data, is unrealistically small, which contradicts the assumption of risk-averse managers. This rejection result also eases an earlier concern about missing restrictions to guarantee that asymmetric effort pairs are not in Pareto-dominant equilibrium. Testing additional restrictions does not change the result of rejection. The question essentially is whether there are other models that can better rationalize the observed compensation and stock returns. The results in this paper answer this question affirmatively. To be cautious, the test on this model is a test joint with an assumption that both managers working is a Pareto-dominant strategy.

There are several potential reasons for the rejection. First, the equal sharing rule can be misspecified. If the shareholders anticipate an incorrect sharing rule, they may fail to induce proper incentives. In turn, if the real bargaining power within the team is away from symmetry, the model can be rejected. Testing how sensitive the rejection of the mutual monitoring with total utility maximization model is to the sharing weight can be a task of future studies. However, the model with individual utility maximization also implicitly assumes a symmetric bargaining power between the two managers because in equilibrium, the Pareto-dominant allocation provides both managers the same expected utility at the value of the same outside option. In such a sense, the equal sharing rule is less likely to be the reason for rejection. Second, the side contracts of utility transfer are not enforceable in reality, though side contracts without utility transfer may work as the mutual monitoring with individual utility maximization model indicates. Managers failing to honor this type of side contracts and converging to the bad equilibrium too often can cause the model to be unable to rationalize the data. Third, the mutual monitoring with total utility maximization model may better fit the moral hazard problem of lower level employees if there are easier ways for them to transfer payment (utility). As a result, even though this model might have an empirical ground, it cannot survive the test given the sample used in this paper.

7 Extension

7.1 Counterfactual Estimation of Welfare Cost of Moral Hazard

Armed with the estimates of primitive parameters, a natural follow-up task is to estimate the welfare cost of moral hazard. I consider three metrics, as follows.²⁶ The first measure of moral hazard cost, denoted by τ_1 , is the difference between the expected output from both managers working and that from at least one manager shirking, namely, the losses

²⁶Similar analysis can be found in Margiotta and Miller (2000) and Gayle and Miller (2009).

shareholders would incur from managers shirking instead of working, or

$$\tau_1 \equiv E[Vx] - \max\{E[(Vxg_{1t}(x, \rho))], E[(Vxg_{2t}(x, \rho))], E[(Vxg_t(x, \rho))]\}. \quad (37)$$

The second measure of moral hazard cost, denoted by τ_{2i} , is the pecuniary benefit the manager i would gain from shirking instead of working. It is equal to the difference between the certainty equivalent to working under perfect monitoring (w_{i2}^o) and that to shirking (w_{i1}^o), which are derived from participation constraint for $i = 1, 2$:

$$\tau_{2i} \equiv w_{i2}^o - w_{i1}^o \quad (38)$$

$$= \frac{b_{t+1}}{\rho(b_t - 1)} [\ln \alpha_{i2}(\rho) - \ln \alpha_{i1}(\rho)]. \quad (39)$$

The third measure of moral hazard cost, denoted by τ_{3i} , is the cost shareholders would be willing to pay for perfect monitoring. It can be reflected in the difference between manager i 's expected compensation under the current optimal contract ($E[w_i(x)]$) and the certainty equivalent to working under perfect monitoring, respectively, for manager $i = 1, 2$:

$$\tau_{3i} \equiv E[w_i(x)] - w_{i2}^o. \quad (40)$$

It could be interesting to investigate the efficiency of the optimal compensation contract by contrasting τ_1 with τ_{2i} and τ_{3i} .

7.2 Testing a Model Observationally Equivalent to Mutual Monitoring with the Individual Utility Maximization Model

In this section, I discuss another potentially testable model that is observationally equivalent to the mutual monitoring with individual utility maximization model. It relies on self-enforcing punishment in a repeated game in the spirit of folk theorem. The comparison is summarized in the table below, followed by a detailed discussion.

Model	Self-Enforcing Punishment	Mutual Monitoring with Individual Utility Maximization
· Equilibrium	Subgame perfect	Pareto dominant
· Managers' interaction	Noncooperative	Cooperative
· Research challenge	Unknown discount factor and profitable deviation	
· Most related theory papers	Arya et al. (1997); Che and Yoo (2001)	Itoh (1993)

By modifying the game structure to create credible threats, (work, work) can be sustained as a subgame-perfect equilibrium. This new structure is observationally equivalent to the Pareto-dominant equilibrium in the mutual monitoring with individual utility maximization model in the sense that the modifications do not affect identification because the threats are off the equilibrium path, that is, they are self-enforcing but are never played.

I assume that the two managers can observe each other's effort choice and that the trigger strategy is based on this observation. Note that in the current mutual monitoring with individual utility maximization model, both the participation constraint and the incentive compatibility constraint are binding at the outside option which is normalized to -1 . To make the punishment strictly individually rational, the outside option in the current model is renamed as "accept the offer but resign." Then, I introduce the fourth option for the managers to choose as "reject the offer," which brings even lower utility for each manager, but at the same level for each, regardless of what choice the other makes, say, a number $m < -1$. That is, shareholders design the optimal contract such that there is some rent for the managers to stay.²⁷ So never forming the team, that is, "both managers reject," is a stage game Nash equilibrium with the payoffs strictly lower than "accept and work." It is thus a self-enforcing punishment the managers can put on the shirker in the team. Because shareholders want to keep both managers, the participation constraint will meet at the option of "resign" rather than "reject." The (work, work) equilibrium can be sustained if the managers are patient and the profitable deviation in the stage game is not very large.

Because (work, work) is supported by the trigger strategy as a subgame-perfect equilibrium, the data are still generated from the equilibrium in which both managers work in each period. In the infinitely repeated game, (shirk, shirk) is not an equilibrium, and the

²⁷MacLeod and Malcomson (1989) discuss the role of exit cost in a subgame-perfect equilibrium under a single-agent setting, but here the idea of creating the rent of stay is similar.

trigger strategy never happens. In such a sense, this structure is observationally equivalent to the one laid out in the paper where the two managers play a Pareto-dominant strategy (work, work). If a finitely repeated model applies as Arya et al. (1997) suggests, presumably, shareholders implement the group-incentive contract (the first period contract in their paper) for a time duration longer than the duration of the two managers in the sample because their second-period individual incentive contract is a credible threat but the contract type is assumed the same in the panel data.

The mutual monitoring incentive arising from repeated interactions sounds appealing, and there is a large body of theoretical research on this topic, though rare empirical study. My model does not rule out this type of structure, which uses self-enforcing punishment in a repeated game to support a subgame-perfect equilibrium, but there is no sufficient data to distinguish it from the mutual monitoring with individual utility maximization model. In particular, the first issue is that the discount factor needs to be estimated. It may be borrowed from previous studies, so it may not be a severe concern. The second issue is that the profitable deviation in the stage game needs to be identified and estimated too, regardless of the normalization in the rejection payoff m . Accomplishing this task requires other sources to identify and estimate the value of managers' options off the equilibrium path, but this is not infeasible.

8 Conclusion

Hidden action and free riding are two fundamental frictions in the moral hazard problem in top management teams. To mitigate the problem, shareholders can base top managers' individual compensation on stock performance and exploit mutual monitoring among managers, as theory suggests. Previous structural estimation papers find that the welfare costs of moral hazard can, to a large extent, help explain the increases in executive compensation over past decades (Gayle and Miller 2009). To examine the importance of moral hazard more closely, this paper investigates whether shareholders exploit uncodified incentives, such as mutual monitoring, in the optimal compensation design. This is an empirical question. If shareholders only provide individual incentives in the optimal compensation, then it seems meaningless to examine the consequences of group-based incentives, for example, studying the association between the relative characteristics of top executive compensation and firm performance.

The theory-based empirical investigation in this paper attempts to answer the preceding question more directly. This paper identifies and tests three competing structural models that are explicitly based on theoretical models of principal-multiagent moral hazard. The

three models are intended to capture the crucial considerations in shareholders' optimal compensation design, that is, whether and how the managers can monitor each other. If shareholders do not exploit the mutual monitoring, the no mutual monitoring model applies. If shareholders exploit the mutual monitoring, the other two models fit into this class. Furthermore, if shareholders consider the two managers as a unitary decision maker, the mutual monitoring with total utility maximization model characterizes this case. Otherwise, if shareholders consider the two managers as self-interested decision makers, the mutual monitoring with individual utility maximization model applies.

For each model, this paper exploits the equilibrium restrictions to delimit the identified set of the risk aversion parameter to which all other primitive parameters in the same model can be indexed. The hypothesis tests are based on the confidence region of the identified set. The nonparametric technique used in this paper can, to certain extent, alleviate concern about overusing auxiliary assumptions. This concern applies to many structural modeling papers. The set identification method allows me to examine a richer set of equilibrium restrictions by incorporating both equality and inequality moment conditions into the criterion functions of the tests.

To analyze the results of the hypothesis tests and draw conclusions, we need to delve into a discussion about the assumption of homogeneity of managers' risk preferences. Under the mutual monitoring with total utility maximization model, the identified sets of the risk aversion parameter are either empty or close to zero (meaning risk neutrality). If we assume that the managers are risk averse to some degree, this model is rejected. Under the no mutual monitoring 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, we have 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 ordinary people, it is unclear to what extent they among themselves are distinguishable in terms of risk aversion based on the characteristics of their employers. By contrast, the mutual monitoring with individual utility maximization 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 the most robust explanatory power for the correlation between the observed executive compensation and stock returns.

Although the management literature has found that "attention to executive groups, rather than to individuals, often yields better explanations of organizational outcomes" (Hambrick, 2007, page 334), its emphasis is on behavioral integration and collective cognition based on demographic characteristics. This paper may advance our understanding of

how economic incentives work in public firms; that is, shareholders respond to moral hazard by taking advantage of mutual monitoring in designing optimal compensation, and top managers engage in mutual monitoring in self-interest.

Internal governance is gaining attention from both theorists (Acharya et al. 2011) and empiricists (Armstrong et al. 2010; Landier et al. 2012). It is unlikely that outsiders know more about the top executives than compensation designers. The unconditional explanation provided by the mutual monitoring with individual utility maximization model tends to suggest that, from the compensation designers' perspective, mutual monitoring as one type of internal governance mechanism is exploited to mitigate the moral hazard in top management teams, even though each manager engages in mutual monitoring only to maximize his own utility. Armed with empirical evidence, this paper calls for attention to the positive effects of managerial coordination such as mutual monitoring in the same way that external governance mechanisms, such as takeovers and labor market competition, have been well explored.

Also, the results in this paper invite two issues for future investigation. First, in this paper I assume that the mutual monitoring is free for managers to enforce. Relaxing this assumption can generate cross-sectional variation in the effectiveness of mutual monitoring. Traditionally, in studying the determinants and consequences of executive compensation, researchers mainly focus on corporate governance factors relying on explicit provisions. This paper suggests that researchers may also need to consider factors that affect the enforcement of mutual monitoring when managers are engaged as self-interested decision makers. For example, theoretical studies have suggested factors 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), among other factors.

Second, it could be interesting to figure out how the mutual monitoring is enforced, which is under-explored in this paper. When coordination between managers turns out to be useful to shareholders, investment in human resources to facilitate cooperation is in demand. For example, maintaining a stable and close network within top management teams may be beneficial to a firm, but could be otherwise detrimental if the managers tend to collude against shareholders' interests. In this sense, investigating the nature of managerial coordination in firms, as this paper does, has real implications.

A Proofs

Proof of Lemma 1. The assumption that manager 1's marginal influence on the distribution of the single output x is unconditional on manager 2's effort choice implies that the deviation of x 's probability density from manager 1 working to manager 1 shirking is the same no matter whether manager 2 works or shirks, and vice versa. Denote $f(\bullet|j, k)$ as the PDF of x conditional on the two managers' effort choices, mathematically:

$$g_1(x) \equiv \frac{f(x \mid \text{manager 1 shirks, manager 2 works})}{f(x \mid \text{manager 1 works, manager 2 works})} \quad (41)$$

$$= \frac{g_1(x)f(x)}{f(x)} \quad (42)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x \mid \text{manager 1 works, manager 2 shirks})} \quad (43)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x)} \quad (44)$$

$$g_2(x) \equiv \frac{f(x \mid \text{manager 1 works, manager 2 shirks})}{f(x \mid \text{manager 1 works, manager 2 works})} \quad (45)$$

$$= \frac{g_2(x)f(x)}{f(x)} \quad (46)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x \mid \text{manager 1 shirks, manager 2 works})} \quad (47)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{g_1(x)f(x)}. \quad (48)$$

Using (41) and (47) gives

$$\begin{aligned} & \int g_1(x)g_2(x)f(x)dx \\ &= \int \frac{f(x \mid \text{manager 1 shirks, manager 2 works})}{f(x \mid \text{manager 1 works, manager 2 works})} \\ & * \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x \mid \text{manager 1 shirks, manager 2 works})} f(x)dx \\ &= \int \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x)} f(x)dx \\ &= 1. \end{aligned}$$

The last equality is by the definition of a PDF. ■

Proof of Proposition 2. See the first-order conditions (FOCs) in the proof of Proposition 3 later. ■

Proof of Proposition 3. No Mutual Monitoring Model

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

$$\begin{aligned}
L = & E [\ln v_{1t}(x) + \ln v_{2t}(x)] \\
& - \mu_1 \left[\alpha_{12}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)] - \alpha_{11}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)g_1(x)] \right] \\
& - \mu_2 \left[\alpha_{22}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)] - \alpha_{21}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)g_2(x)] \right] \\
& - \mu_3 \left[\alpha_{12}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)] - 1 \right] \\
& - \mu_4 \left[\alpha_{22}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)] - 1 \right].
\end{aligned} \tag{49}$$

The First Order Condition (FOC hereafter) w.r.t. $v_{1t}(x)$ is

$$1/v_{1t}(x) = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} - \mu_1\alpha_{11}^{\frac{1}{b_t-1}} g_1(x). \tag{50}$$

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

$$1/v_{2t}(x) = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}} - \mu_2\alpha_{21}^{\frac{1}{b_t-1}} g_2(x). \tag{51}$$

Evaluate the FOCs at the threshold values of shirking distribution, respectively, to get

$$1/v_{1t}(\bar{x}_1) = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} \tag{52}$$

$$1/v_{2t}(\bar{x}_2) = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}}. \tag{53}$$

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

$$E [1/v_{1t}(x)] = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} - \mu_1\alpha_{11}^{\frac{1}{b_t-1}} \tag{54}$$

$$E [1/v_{2t}(x)] = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}} - \mu_2\alpha_{21}^{\frac{1}{b_t-1}}. \tag{55}$$

The binding participation constraint for each manager gives

$$\alpha_{12}^* = E_t[v_{1t}(x)]^{1-b_t} \tag{56}$$

$$\alpha_{22}^* = E_t[v_{2t}(x)]^{1-b_t}. \tag{57}$$

The binding incentive compatibility constraint gives

$$\alpha_{11}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)g_1(x)] = \alpha_{21}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)g_2(x)] = 1. \tag{58}$$

Multiply both sides of (50) and integrate over $f(x)$; it follows that

$$1 = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)] - \mu_1\alpha_{11}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)g_1(x)], \quad (59)$$

and plugging (56) and (57) into the preceding, it follows that

$$\mu_3 = 1.$$

Multiply both sides of (51) and integrate over $f(x)$; it follows that

$$1 = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)] - \mu_2\alpha_{21}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)g_2(x)], \quad (60)$$

and plugging (57) and (58) into the preceding, it follows that

$$\mu_4 = 1.$$

Multiplying (52) by $E_t [v_{1t}(x)]$ and using $\mu_3 = 1$, it follows that

$$\begin{aligned} E_t [v_{1t}(x)]/v_{1t}(\bar{x}_1) &= \mu_1 + \mu_3 \\ \mu_1 &= E_t [v_{1t}(x)]/v_{1t}(\bar{x}_1) - 1. \end{aligned}$$

Similarly, multiplying (53) by $E_t [v_{2t}(x)]$ and using $\mu_4 = 1$, it follows that

$$\begin{aligned} E_t [v_{2t}(x)]/v_{2t}(\bar{x}_2) &= \mu_2 + \mu_4 \\ \mu_2 &= E_t [v_{2t}(x)]/v_{2t}(\bar{x}_2) - 1. \end{aligned}$$

Equations (50), (52), and (54) together give

$$\begin{aligned} 1/v_{1t}(\bar{x}_1) - E [1/v_{1t}(x)] &= \mu_1\alpha_{11}^{\frac{1}{b_t-1}} \\ 1/v_{1t}(\bar{x}_1) - 1/v_{1t}(x) &= \mu_1\alpha_{11}^{\frac{1}{b_t-1}} g_1(x) \end{aligned}$$

and

$$g_1(x) = \frac{1/v_{1t}(x) - 1/v_{1t}(\bar{x}_1)}{E [1/v_{1t}(x)] - 1/v_{1t}(\bar{x}_1)} \quad (61)$$

$$g_2(x) = \frac{1/v_{2t}(x) - 1/v_{2t}(\bar{x}_2)}{E [1/v_{2t}(x)] - 1/v_{2t}(\bar{x}_2)}. \quad (62)$$

Plug into (58); it follows that

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

Mutual Monitoring with Total Utility Maximization Model

We want to show that $\theta^* = \theta(\rho^*)$.

The Lagrangian for the shareholders' cost minimization problem is

$$\begin{aligned}L_{EC} &= E_t [\ln v_{1t}(x) + \ln v_{2t}(x)] \\ &\quad - \eta_0 \left[\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] - 2 \right] \\ &\quad - \eta_1 \left\{ \begin{aligned} &\left[\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] \right] \\ &- \left[\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] + \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)] \right] \end{aligned} \right\}.\end{aligned}\tag{63}$$

The First Order Condition (FOC hereafter) w.r.t. $v_{1t}(x)$ is

$$1/v_{1t}(x) = \eta_0 \alpha_{12}^{\frac{1}{b_t-1}} + \eta_1 \alpha_{12}^{\frac{1}{b_t-1}} - \eta_1 \alpha_{11}^{\frac{1}{b_t-1}} g(x).\tag{64}$$

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

$$1/v_{2t}(x) = \eta_0 \alpha_{22}^{\frac{1}{b_t-1}} + \eta_1 \alpha_{22}^{\frac{1}{b_t-1}} - \eta_1 \alpha_{21}^{\frac{1}{b_t-1}} g(x).\tag{65}$$

Multiply both sides of (64) with $v_{1t}(x)$ and then integrating over $f(x)$, we get

$$1 = (\eta_0 + \eta_1) \alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] - \eta_1 \alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)].\tag{66}$$

Similarly, from (65), we get

$$1 = (\eta_0 + \eta_1) \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] - \eta_1 \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)].\tag{67}$$

Recall that

$$g(\bar{x}) = 0, \forall x > \bar{x}.$$

Evaluate the FOCs at the threshold of the both-manager shirking distribution,

$$1/v_{1t}(\bar{x}) = (\eta_0 + \eta_1)\alpha_{12}^{\frac{1}{b_t-1}} \quad (68)$$

$$1/v_{2t}(\bar{x}) = (\eta_0 + \eta_1)\alpha_{22}^{\frac{1}{b_t-1}}. \quad (69)$$

Binding participation constraint gives

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] = 2. \quad (70)$$

Binding incentive compatibility constraint gives

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] = \alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] + \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)]. \quad (71)$$

The utility transfer constraint implies that the following equation is held if both managers shirk:

$$\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] = \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)]. \quad (72)$$

Similarly, if both work,

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] = \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)]. \quad (73)$$

Combining (70) and (71), we can immediately get

$$\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] = \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)] = 1 \quad (74)$$

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] = \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] = 1 \quad (75)$$

$$\alpha_{12}^* = E_t[v_{1t}(x)]^{1-b_t} \quad (76)$$

$$\alpha_{22}^* = E_t[v_{2t}(x)]^{1-b_t}. \quad (77)$$

Add (68) and (69). Then use binding IC and plug in (76) and (77):

$$2 = \eta_0 \left[\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] \right] + 0$$

$$\eta_0^* = 1.$$

Plug η_0^* into (68) and (69); we get

$$\begin{aligned} \eta_1^* &= \left[\frac{E_t[v_{1t}(x)]}{v_{1t}(\bar{x})} - 1 \right] \\ &= \left[\frac{E_t[v_{2t}(x)]}{v_{2t}(\bar{x})} - 1 \right]. \end{aligned}$$

Take the expectation over FOCs; we get

$$E[1/v_{1t}(x)] = (\eta_0 + \eta_1)\alpha_{12}^{\frac{1}{b_t-1}} - \eta_1\alpha_{11}^{\frac{1}{b_t-1}} \quad (78)$$

$$E[1/v_{2t}(x)] = (\eta_0 + \eta_1)\alpha_{22}^{\frac{1}{b_t-1}} - \eta_1\alpha_{21}^{\frac{1}{b_t-1}}. \quad (79)$$

Plug η_0^* and η_1^* into (68) and (78); we get

$$\alpha_{11}^* = \left[\frac{E_t[v_{1t}(x)] - v_{1t}(\bar{x})}{1 - v_{1t}(\bar{x})E[1/v_{1t}(x)]} \right]^{1-b_t}. \quad (80)$$

Similarly, combining (69) and (79), we get

$$\alpha_{21}^* = \left[\frac{E_t[v_{2t}(x)] - v_{2t}(\bar{x})}{1 - v_{2t}(\bar{x})E[1/v_{2t}(x)]} \right]^{1-b_t}. \quad (81)$$

Plug η_0^* and η_1^* into (64) and (65), respectively; using (78), (79), (72), and (73), we get

$$\frac{1 - v_{1t}(\bar{x})/v_{1t}(x)}{1 - v_{1t}(\bar{x})E[1/v_{1t}(x)]} = g^*(x) = \frac{1 - v_{2t}(\bar{x})/v_{2t}(x)}{1 - v_{2t}(\bar{x})E[1/v_{2t}(x)]}.$$

Mutual Monitoring with Individual Utility Maximization Model

See the proof for the no mutual monitoring model. The only difference is that the likelihood ratio is the same. ■

Proof of Proposition 4. In the cost minimization problem, the objective function is quasi-concave and the constraints are linear in $v_i(x)$. Consequently, the FOCs that are used to derive the parameters can uniquely determine the solution to the optimal contracting problem if the complementary slackness conditions are satisfied. This can be easily confirmed by multiplying the Lagrangian multiplier with the associated constraint and finding that the product equals zero. Then the proposition is proved. ■

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

Either the gross abnormal return or the optimal compensation cannot be directly observed from real data. I construct their consistent estimators as discussed subsequently. Here \tilde{x}_{nt} represents the abnormal returns, and \tilde{w}_{imt} is manager i 's total compensation from firm n in

year t . (Z_{nt}, S_{nt}) are firm type variables, defined before. I nonparametrically estimate the optimal compensation using the following kernel regression (Pagan and Ullah 1999):²⁸

$$\begin{aligned} w_{int} &\equiv E_t[\tilde{w}_{int}|\tilde{x}_{nt}, V_{n,t-1}] \\ &= \frac{\sum_{m=1, m \neq n}^N \tilde{w}_{imt} * I\{Z_{mt} = Z_{nt}, S_{mt} = S_{nt}\} K\left(\frac{\tilde{x}_{mt} - \tilde{x}_{nt}}{h_x}, \frac{v_{m,t-1} - v_{n,t-1}}{h_v}\right)}{\sum_{m=1, m \neq n}^N I\{Z_{mt} = Z_{nt}, S_{mt} = S_{nt}\} K\left(\frac{\tilde{x}_{mt} - \tilde{x}_{nt}}{h_x}, \frac{v_{m,t-1} - v_{n,t-1}}{h_v}\right)}, \end{aligned}$$

where $v_{n,t-1}$ is the market value of firm n at the end of year $t-1$. Then I calculate the gross abnormal returns by

$$x_{nt} \equiv \tilde{x}_{nt} + \frac{w_{1nt}}{V_{n,t-1}} + \frac{w_{2nt}}{V_{n,t-1}}.$$

The PDF of gross abnormal return x_{nt} is nonparametrically estimated by a kernel estimator:

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

²⁸ $K(\cdot)$ is a multivariate standard normal kernel density function:

$$K(\cdot) = \exp\left\{-0.5 * \left(\frac{\bar{x}_{mt} - \bar{x}_{nt}}{h_x}\right)^2\right\} * \exp\left\{-0.5 * \left(\frac{\bar{v}_{mt} - \bar{v}_{nt}}{h_v}\right)^2\right\} \frac{|S|^{-1/2}}{(2\pi)h_x h_v},$$

where $S = \text{cov}(\tilde{x}, v)$, $h_{x/v} = 1.06 * sd_{x/v} * N_{I\{\cdot\}}^{-1/5}$, is a cross-validation bandwidth and $(\bar{x}, \bar{v}) = (\tilde{x}, v)S^{-1/2}$, (\tilde{x}_t, v_{t-1}) are the raw abnormal returns and raw one-year lagged market value.

References

- Acharya, V., S. C. Myers, and R. G. Rajan. 2011. The Internal governance of firms. *Journal of Finance* 66 (3): 689–720.
- Antle, R., and A. Smith. 1985. Measuring executive compensation: Methods and an application. *Journal of Accounting Research* 23: 296–325.
- Antle, R., and A. Smith. 1986. An empirical investigation of the relative performance evaluation of corporate executives. *Journal of Accounting Research* 24: 1–39.
- Armstrong, C., A. D. Jagolinzer, and D. F. Larcker 2010. *Performance-Based Incentives for Internal Monitors*. Working paper, Rock Center for Corporate Governance at Stanford University Working Paper Series No. 76, Stanford University Graduate School of Business Research Paper No. 2052.
- Arya, A., J. Fellingham, and J. Glover. 1997. Teams, repeated tasks, and implicit incentives. *Journal of Accounting and Economics* 23: 7–30.
- Bolton, P., and M. Dewatripont. 2005. *Contract Theory*. Cambridge, MA: MIT Press.
- Bushman, R., Z. Dai, and W. Zhang. 2012. *Management Team Incentive Alignment and Firm Value*. Working paper, University of North Carolina at Chapel Hill.
- Che, Y.-K., and S.-W. Yoo. 2001. Optimal incentives for teams. *American Economic Review* 91 (3): 525–541.
- Chernozhukov, V., H. Hong, and E. Tamer. 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75 (5): 1243–1284.
- Core, J. E., W. Guay, and D. F. Larcker. 2003. *Executive Equity Compensation and Incentives: A Survey*. New York: Federal Reserve Bank of New York Economic Policy Review.
- Fama, E. F. 1980. Agency problems and the theory of the firm. *Journal of Political Economy* 88 (2): 288–307.
- Fudenberg, D., B. Holmstrom, and P. Milgrom. 1990. Short-term contracts and long-term agency relationships. *Journal of Economic Theory* 51 (1): 1–31.
- Gayle, G.-L., and R. A. Miller. 2009. Has moral hazard become a more important factor in managerial compensation? *American Economic Review* 99 (5): 1740–1769.

- Gayle, G.-L., and R. A. Miller. 2012. *Identifying and Testing Models of Managerial Compensation*. Working paper, Tepper School of Business, Carnegie Mellon University.
- Gayle, G.-L., C. Li, and R. A. Miller. 2012. *The Consequences of the Sarbanes-Oxley Act on CEOs' Compensation*. Working paper, Tepper School of Business, Carnegie Mellon University.
- Glover, J. 2012. Explicit and implicit incentives for multiple agents. *Foundations and Trends in Accounting* (forthcoming).
- Grossman, S. J., and O. D. Hart. 1983. An analysis of the principal-agent problem. *Econometrica* 51 (1): 7–45.
- Hambrick, D. C. 2007. Upper echelons theory: An update. *Academy of Management Review* 32 (2): 334–343.
- Hall, B., and J. Liebman. 1998. Are CEOs really paid like bureaucrats? *Quarterly Journal of Economics* 103: 653–691.
- Henderson, A. D., and J. W. Fredrickson. 2001. Top management team coordination needs and the CEO pay gap: A competitive test of economic and behavioral views. *Academy of Management Journal* 44 (1): 96–117.
- Hermalin, B. E. 1998. Toward an economic theory of leadership: Leading by example. *American Economic Review* 88 (5): 1188–1206.
- Holmstrom, B., and P. Milgrom. 1990. Regulating trade among agents. *Journal of Institutional and Theoretical Economics* 146: 85–105.
- Itoh, H. 1990. Coalitions, incentives, and risk sharing. *Unpublished Manuscript*.
- Itoh, H. 1993. Coalitions, incentives, and risk sharing. *Journal of Economic Theory* 60: 410–427.
- Kandel, E., and E. P. Lazear. 1992. Peer pressure and partnerships. *Journal of Political Economy* 100 (4): 801–817.
- Kreps, D. M. 1990. Corporate culture and economic theory. *Perspectives on Positive Political Economy* 90: 109–110.
- Landier, A., J. Sauvagnat, and D. Sraer. 2012. Bottom-up corporate governance. *Review of Finance* (forthcoming).

- Li, Z. 2011. *Mutual Monitoring and Corporate Governance*. Working paper, Arizona State University.
- Ma, C. 1988. Unique Implementation of incentive contracts with many agents. *Review of Economic Studies* 55: 555-572.
- MacLeod, W. B. 1995. Incentives in organizations: An overview of some of the evidence and theory. In *Trends in Business Organization: Do Participation and Cooperation Increase Competitiveness?*, edited by H. Siebert, 832-854. Kiel, Germany: Kiel Institute of World Economics.
- MacLeod, W. B., and J. M. Malcomson. 1989. Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica* 57 (2): 447-480.
- Main, B. G. M., C. A. O'Reilly III, and J. Wade. 1993. Top executive pay: Tournament or teamwork? *Journal of Labor Economics* 11 (4): 606-628.
- Margiotta, M. M., and R. A. Miller. 2000. Managerial compensation and the cost of moral hazard. *International Economic Review* 41 (3): 669-719.
- Matzkin, R. L. 2007. Nonparametric identification. In *Handbook of Econometrics*, vol. 6, part B, edited by J. J. Heckman and E. E. Leamer, 5307-5368. New York: Elsevier.
- Mirrlees, J. A. 1975. *The Theory of Moral Hazard and Unobservable Behaviour*. Mimeo, Oxford University.
- Murphy, K. J. 1999. Executive compensation. In *Handbook of Labor Economics*, vol. 3, part B, edited by O. C. Ashenfelter and D. Card, 2485-2563. New York: Elsevier.
- Murphy, K. 2012. Executive compensation: Where we are, and how we got there. In *Handbook of the Economics of Finance*, edited by G. Constantinides, M. Harris, and R. Stulz. New York: Elsevier Science North Holland (forthcoming).
- Pagan, A., and A. Ullah. 1999. *Nonparametric Econometrics*. Cambridge: Cambridge University Press.
- Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37: 7-63.
- Ramakrishnan, R. T. S., and A. V. Thakor. 1991. Cooperation versus competition in agency. *Journal of Law, Economics, and Organization* 7 (2): 248-286.

Tirole, J. 1992. Collusion and the Theory of Organizations. In *Advances in Economic Theory Sixth World Congress (II)*, edited by J.-J. Laffont, 151-206.

Varian, H. R. 1990. Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics* 146: 153–174.

Table 1: Cross-Sectional Summary on Firm Characteristics

Sector	Primary	Consumer Goods	Service
Assets	4704 (7423)	3059 (5035)	4688 (8307)
Debt/Equity	1.84 (1.40)	1.52 (1.41)	2.56 (3.36)
Abnormal Returns	-0.014 (0.29)	-0.026 (0.31)	-0.007 (0.33)
Market Value	3285 (4808)	3440 (5181)	3417 (5059)
Observations	6583	5004	8023

Note: Both Assets (the Total Assets on Balance Sheet) and Market Value are measured in millions of 2006 \$US. To calculate the abnormal return, 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 the compounded returns for the corresponding fiscal year. I drop firm-year observations if the firm changed its fiscal year end, such that all compensations and stock returns are twelve-month based.

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

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

Note: Compensation is measured in thousands of 2006 \$US. Mean is reported and standard deviation is in the parenthesis below. In the first three columns, the third row for each type of firms reports the number of observations.

Table 3: Time-Series Summary of Compensation Components for Each Manager

Year	Salary		Bonus		Values of Restricted Stocks		Values of Granted Options		Changes in Wealth from Stocks Held		Changes in Wealth from Options Held		Total Compensation		No. of Observations
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	
1993	545 (265)	413 (209)	347 (408)	240 (315)	92 (290)	61 (230)	599 (1000)	384 (701)	834 (2000)	493 (1465)	1186 (3653)	468 (2090)	4759 (6920)	2457 (4369)	1423
1994	548 (264)	411 (208)	398 (437)	269 (332)	97 (317)	73 (253)	852 (1320)	492 (873)	1171 (2433)	631 (1557)	742 (2748)	300 (1765)	5045 (7251)	2422 (3634)	1561
1995	518 (262)	407 (207)	402 (444)	272 (328)	107 (332)	79 (273)	738 (1236)	454 (817)	1050 (2375)	561 (1569)	813 (3024)	300 (1677)	4941 (7505)	2519 (4414)	1543
1996	536 (263)	407 (203)	418 (463)	276 (324)	125 (363)	76 (258)	1007 (1504)	630 (1020)	1474 (2818)	879 (1938)	799 (2934)	451 (2147)	6059 (8478)	3156 (4743)	1610
1997	528 (261)	401 (203)	443 (472)	305 (362)	113 (362)	95 (310)	1098 (1574)	739 (1230)	1754 (3113)	1230 (2582)	1330 (3949)	751 (2686)	7472 (9566)	4272 (6634)	1682
1998	481 (255)	390 (196)	383 (461)	271 (340)	121 (369)	94 (317)	1112 (1601)	798 (1308)	1111 (2637)	701 (2046)	537 (2921)	208 (1985)	5487 (8420)	3099 (5602)	1679
1999	522 (272)	407 (200)	440 (512)	299 (350)	117 (363)	82 (269)	1256 (1705)	852 (1341)	1514 (3061)	969 (2338)	313 (2585)	77 (1835)	5788 (8544)	3484 (6339)	1482
2000	563 (283)	431 (217)	417 (486)	314 (399)	124 (394)	94 (306)	1306 (1598)	941 (1455)	1915 (3357)	1347 (2667)	1979 (4527)	1071 (2974)	9295 (10990)	5510 (8244)	1400
2001	590 (276)	444 (223)	411 (523)	274 (373)	148 (414)	114 (348)	1594 (1910)	1132 (1567)	2298 (3561)	1569 (2802)	2089 (4277)	923 (2408)	10471 (11265)	5642 (7296)	1418
2002	584 (279)	427 (205)	467 (539)	320 (410)	162 (435)	130 (359)	1429 (1759)	947 (1393)	1510 (2969)	936 (2200)	1825 (4177)	781 (2312)	8386 (10157)	4308 (6403)	1500
2003	600 (282)	441 (207)	528 (589)	352 (429)	219 (505)	174 (414)	1231 (1636)	753 (1114)	2379 (3679)	1399 (2577)	1075 (3599)	616 (2712)	8326 (9959)	4666 (7130)	1456
2004	584 (278)	432 (203)	599 (617)	398 (437)	288 (551)	218 (438)	1246 (1632)	779 (1234)	1753 (3134)	1011 (2226)	1458 (3821)	772 (2715)	8189 (9681)	4381 (6124)	1464
2005	563 (270)	430 (208)	597 (610)	388 (413)	345 (592)	267 (478)	1076 (1555)	622 (1009)	1286 (2883)	807 (2138)	1132 (3528)	546 (2234)	7473 (9825)	3911 (6080)	1392

Note: "1st" is the highest paid manager and "2nd" is the second highest paid. Each component is measured in thousands of 2006 \$US. The mean of each component is reported with standard deviation in the parenthesis below. The Changes in Wealth from Stocks Held is equal to the beginning shares of held stocks multiplied by the abnormal returns. The Changes in Wealth from Options Held is the difference between the ending option value and the beginning option value multiplied by market portfolio return.

Table 4: The Distribution of Positions Held by the Two Highest Paid Managers

Sector	Primary		Consumer Goods		Service	
Compensation Rank	1st	2nd	1st	2nd	1st	2nd
Functional	0.01	0.01	0.01	0.01	0.01	0.02
General 1	0.21	0.08	0.17	0.08	0.21	0.09
General 2	0.18	0.51	0.23	0.50	0.23	0.51
Functional & General 1	0.04	0.10	0.06	0.11	0.04	0.08
Functional & General 2	0.05	0.18	0.07	0.15	0.06	0.15
General 1 & General 2	0.50	0.12	0.45	0.14	0.44	0.13
Functional & General 1 & 2	0.01	0.01	0.01	0.01	0.01	0.02
Number of observations	6583	6583	5004	5004	8023	8023

Note: "1st" is the highest paid manager and "2nd" is the second highest paid. For each type of manager, I count the frequency of holding certain types of positions as follows. "Functional" = 1 if the manager holds one of the following positions: CTO, CIO, COO, CFO, CMO but not any others. "General 1" = 1 if the manager holds one of the following positions: Chairman, President, CEO, or Founder but not any others. "General 2" = 1 if the manager holds one of the following positions: Executive Vice-President, Senior Vice-President, Vice-President, Vice-Chair, or Other (defined in the database) but not any others. "Functional & General 1" = 1 if the manager holds at least one position from each of the Functional category and the General 1 category but none from the General 2 category. Same rule applies to "Functional & General 2" and "General 1 & General 2". "Functional & General 1 & General 2" = 1 if the manager holds at least one position from each of the three categories.

Table 5: The Risk Aversion Parameter's 95 % Confidence Regions for Different Specifications

A: No Mutual Monitoring: different likelihood ratio/different shadow price of IC							
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: Mutual Monitoring with Total Utility Maximization: same likelihood ratio/same shadow price of IC							
Sector	[A, D/E]	Risk Aversion		Certainty Equivalent	Homogeneous within Size	Homogeneous within Sector	Homogeneous across Sectors
Primary	[S, S]	(0.10, 0.13)	(0.003, 0.004)				
	[S, L]	(0.16, 0.21)	(0.005, 0.006)	(,)			
	[L, S]	(,)	(,)				
	[L, L]	(,)	(,)	(,)	(,)	(,)	
Consumer Goods	[S, S]	(0.05, 0.06)	(0.001, 0.002)				
	[S, L]	(0.16, 0.21)	(0.005, 0.006)	(,)			
	[L, S]	(0.02, 0.03)	(0.001, 0.001)				
	[L, L]	(0.03, 0.04)	(0.001, 0.001)	(0.03, 0.03)	(,)		
Service	[S, S]	(2E-9, 0.03)	(2E-9, 0.001)				
	[S, L]	(,)	(,)	(,)			
	[L, S]	(0.02, 0.02)	(0.001, 0.001)				
	[L, L]	(0.05, 0.06)	(0.001, 0.002)	(,)	(,)	(,)	(,)

C: Mutual Monitoring with Individual Utility Maximization: same likelihood ratio/different shadow price of IC							
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)	

Note: IC is short for the incentive compatibility constraint. 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 is the amount paid to avoid a gamble with equal probability to win and lose \$1 million and is measured in \$ million with the median of the bond price in the sample period.