

# **Essays on the Structural Models of Executive Compensation**

By

Chen Li

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Industrial Administration)

at the Tepper Business School

at Carnegie Mellon University

2013

Doctoral Committee:

Professor George-Levi Gayle

Professor Jonathan Glover (Co-chair)

Professor Pierre Jinghong Liang

Professor Robert A. Miller (Co-chair)

# Essays on the Structural Models of Executive Compensation

Chen Li

## Abstract

This dissertation is composed of three chapters in which I use both reduced-form approach and structural approach to study executive compensation in S&P1500 firms from 1993 to 2005.

Chapter 1 provides the literature and methodology background of this dissertation. I summarize existing accounting empirical studies on executive compensation under two tasks, that is, (1) testing contract theory and (2) analyzing policies. I compare structural approach with reduced-form approach in terms of their scopes, execution, and comparative advantages. Also, I briefly introduce the steps of implementing structural analysis and close this chapter with a high level plan for the following two chapters.

Chapter 2 focuses on the first task and is based on my job market paper entitled "Mutual Monitoring within Top Management Teams: A Structural Modeling Investigation". I study whether executive compensation reflects that shareholders take advantage of top managers' mutual monitoring. Mutual monitoring as a solution to moral hazard has been extensively studied by theorists, but the empirical results are few and mixed. This chapter semi-parametrically identifies and tests three structural models of principal-two-agent moral hazard. The Mutual Monitoring with Individual Utility Maximization Model is the most plausible one to rationalize the data of executive compensation and stock returns. 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. These results indicate that shareholders seem to recognize and exploit complementary incentive mechanisms, such as mutual monitoring among self-interested top executives, to design compensation.

Chapter 3 focuses on the second task and attempts to answer the question in its title, "Do 2002 Governance Rules affect CEOs' Compensation?" From two non-parametric tests, I found that both the CEOs' compensation contract shape and the distribution of gross abnormal return (performance measure) have significantly changed after 2002. These changes indicate that shareholders may have adjusted CEOs' compensation contract to those governance rules. The results also give confidence to a more sophisticated test using structural approach based on welfare estimation.

## **Acknowledgements**

To be added

## TABLE OF CONTENTS

<b>1.0 LITERATURE AND METHODOLOGY BACKGROUND</b> . . . . .	2
1.1 Introduction . . . . .	2
1.2 Literature Review . . . . .	4
1.2.1 Testing contract theory: three questions . . . . .	4
1.2.1.1 Incentive problems targeted by compensation contracts . . . . .	4
1.2.1.2 Consequences of compensation contracts . . . . .	6
1.2.1.3 Design of compensation contracts . . . . .	7
1.2.2 Analyzing policies . . . . .	8
1.3 A Comparison between Reduced-form Approach and Structural Approach . . . . .	9
1.3.1 Reduced-form approach . . . . .	10
1.3.1.1 Definition . . . . .	10
1.3.1.2 Research challenges . . . . .	10
1.3.2 Structural approach . . . . .	11
1.3.2.1 Definition . . . . .	11
1.3.2.2 How it works . . . . .	12
1.4 When Do We Need Structural Approach? . . . . .	13
1.4.1 Research questions and advantages of reduced-form approach . . . . .	13

1.4.2	Research questions and advantages of structural approach . . . . .	14
1.5	Implementing Structural Approach . . . . .	16
1.6	Plans for Chapter 2 and Chapter 3 . . . . .	22
<b>2.0</b>	<b>MUTUAL MONITORING WITHIN TOP MANAGEMENT TEAMS: A STRUCTURAL MODELING INVESTIGATION . . . . .</b>	<b>23</b>
2.1	Introduction . . . . .	23
2.2	Models . . . . .	31
2.2.1	Timeline . . . . .	32
2.2.2	Technologies . . . . .	34
2.2.3	Managers' Preferences . . . . .	37
2.2.4	Shareholder's Cost Minimization Problem . . . . .	38
2.2.4.1	Objective Function . . . . .	38
2.2.4.2	Participation Constraint . . . . .	39
2.2.4.3	Incentive Compatibility Constraint . . . . .	42
2.2.5	Optimal Contracts . . . . .	46
2.2.6	Shareholder's Profit Maximization . . . . .	48
2.2.7	Summarizing the Three Models . . . . .	50
2.3	Data . . . . .	52
2.3.1	Heterogeneity in the Data . . . . .	52
2.3.2	Key Variables in the Optimal Contracts . . . . .	53
2.3.2.1	Abnormal Stock Returns . . . . .	53
2.3.2.2	Compensation . . . . .	54
2.3.2.3	Measurement Error . . . . .	57
2.3.3	Bond Prices and a Dynamic Consideration . . . . .	58

2.4	Identification . . . . .	59
2.4.1	No Mutual Monitoring Model . . . . .	63
2.4.2	Mutual Monitoring with Total Utility Maximization Model . . . . .	69
2.4.3	Mutual Monitoring with Individual Utility Maximization Model . . . . .	71
2.4.4	Summary of the Identification Results . . . . .	71
2.5	Estimation and Tests . . . . .	73
2.6	Results . . . . .	76
2.6.1	Estimation of the Risk Aversion Parameter and Tests . . . . .	76
2.6.2	Discussion . . . . .	79
2.6.2.1	A Binary Illustration . . . . .	79
2.6.2.2	No Mutual Monitoring versus Mutual Monitoring with Individ- ual Utility Maximization . . . . .	81
2.6.2.3	No Mutual Monitoring Model versus Mutual Monitoring with Total Utility Maximization . . . . .	82
2.7	Extension . . . . .	84
2.7.1	Counterfactual Estimation of Welfare Cost of Moral Hazard . . . . .	84
2.7.2	Testing a Model Observationally Equivalent to Mutual Monitoring with the Individual Utility Maximization Model . . . . .	85
2.8	Conclusion . . . . .	88
<b>3.0</b>	<b>DO 2002 GOVERNANCE RULES AFFECT CEOS' COMPENSATION?</b>	<b>92</b>
3.1	Introduction . . . . .	92
3.2	Background . . . . .	96
3.3	Data . . . . .	97

3.3.1	State variables . . . . .	97
3.3.1.1	Public state . . . . .	98
3.3.1.2	Private state . . . . .	99
3.3.1.3	Distribution of the states . . . . .	100
3.3.2	Abnormal Stock Returns . . . . .	100
3.3.3	Compensation . . . . .	102
3.4	Nonparametric Tests and Results . . . . .	103
3.4.1	Estimating Optimal Compensation and Performance Measure . . . . .	104
3.4.2	Test on the Change in the Distribution of Gross Abnormal Return . . . . .	105
3.4.2.1	Test statistic . . . . .	105
3.4.2.2	Result . . . . .	107
3.4.3	Test on the Change in the Optimal Contract Shape . . . . .	108
3.4.3.1	Test statistic . . . . .	108
3.4.3.2	Result . . . . .	110
3.5	Conclusion . . . . .	110
<b>4.0</b>	<b>APPENDIX TO CHAPTER 2 . . . . .</b>	<b>112</b>
4.1	Proofs . . . . .	112
4.2	Nonparametric Estimation of Compensation and the Probability Density Function of Gross Abnormal Returns in Equilibrium	119
4.3	Tables . . . . .	120
<b>5.0</b>	<b>APPENDIX TO CHAPTER 3 . . . . .</b>	<b>121</b>
5.1	Calculation of wealth change in holding stock and/or options . . . . .	121
5.2	Tables . . . . .	122
	<b>BIBLIOGRAPHY . . . . .</b>	<b>123</b>

## 1.0 LITERATURE AND METHODOLOGY BACKGROUND

### 1.1 INTRODUCTION

Shareholders use compensation contracts to mitigate the agency problems of executives. Those problems stem from the conflicting interests between shareholders and executives when ownership is separated from control, dating back to Berle and Means (1932). Executive compensation has been of interests to academia, practice, and regulators. Researchers study executive compensation in economics, finance, accounting, and management. Their research methods can be theoretical, empirical, experimental, and field survey. To contribute, this dissertation uses nonparametric method and structural model approach, both new to the accounting field, to study executive compensation. This chapter provides the literature and methodology background.

Section 1 mainly focuses on empirical accounting literature in the past decade since a thorough literature review by Bushman and Smith (2001) and concerns expressed by Ittner and Larcker (2002), "there is almost always a very tenuous link between the theoretical notions developed in principal-agent models and the actual research hypotheses and empirical methods used in compensation research". The purpose of this section is not to provide a complete review of previous studies on executive compensation, given that the body of this literature is huge and there have been several existing excellent surveys by Rosen (1992),

Finkelstein and Hambrick (1996), Abowd and Kaplan (1999), Murphy (1999, 2012), Core et al. (2003), Bertrand (2009), and Frydman and Jenter (2010) on empirical findings, Edmans and Gabaix (2009) and Lambert (2001, 2006) on theoretical results, and various textbook treatments of contract theory, for example, Bolton and Dewatripont (2005), among others.

Instead, I restrict the scope to accounting literature to which this dissertation attempts to make contribution and I organize previous findings under two empirical tasks with which this dissertation associates. First, some papers attempt to test contract theory that can rationalize executive compensation. Interactions between theory and reality are at the core of any scientific approach (Salanie, 2003) and executive compensation data by nature can help us examine the empirical relevance of issues studied by contract theory. Second, policies that affect executive compensation have also been investigated. As Murphy (2012) emphasizes, "government intervention has been both a response to and a major driver of time trends in executive compensation over the past century, and that any explanation for pay that ignores political factors is critically incomplete".

Except for these two tasks' own importance in intellectual inquiry, they also provide a good ground on the methodology front for a sharp contrast between the reduced-form approach which is more traditional in accounting and the structural approach which is new.

Section 2 compares those two available and complementary empirical approaches. Section 3 discusses the comparative advantages of the structural approach. Both sections revolve around the two empirical tasks raised in section 1. Section 4 illustrates how to implement the structural approach with critical steps and challenges highlighted. Section 5 sketches the agenda of the following two chapters. One attempts to test multi-agent moral hazard models in the context of top management teams' compensation design. The other investigates the consequences of the 2002 governance rule on CEOs' compensation.

## 1.2 LITERATURE REVIEW

### 1.2.1 Testing contract theory: three questions

Salanie (2003) proposed three important empirical questions in general for researchers who attempt to test contract theory. First, can we find convincing evidence for the presence of a relevant amount of asymmetric information, or is it just a theorist's tale? Second, is there the effect of the various contractual forms on the behavior of the agents who operate under these contracts? Alternatively, do incentives matter? Third, are the observed contracts in real world close enough to the optimum contracts derived from a theoretical analysis of the situation?

These questions invite inquiries in the context of executive compensation. Correspondingly, papers can be classified into three groups based on their answers to the following three questions. First, does executive compensation respond to certain agency problems caused by the friction of information asymmetry? Second, do compensation contracts affect executives' behaviors? Third, are the observed features of executive compensation consistent with theoretical prediction on optimal design?

**1.2.1.1 Incentive problems targeted by compensation contracts** The theoretical agency theory literature and empirical executive compensation literature developed together at the very beginning. Seminal papers in agency theory by Holmstrom (1979, 1982) are tested by Antle and Smith (1986) published in the *Journal of Accounting Research*. In the past decade, empirical accounting researchers attempt to examine the following agency problems using executive compensation data.

First, executive compensation by itself aims at solving certain incentive problems. Reex-

amining the adoption of relative performance evaluation (RPE) in Antle and Smith (1986), Gong et al. (2011) and Albuquerque (2009) provide new evidence supporting the use of RPE. They attribute previous weak support for RPE partially to the lack of detailed information of compensation contract terms and misspecified benchmark group.

Another three papers study the incentive provided in compensation around turnover. Yermack (2006) finds that severance pay deters leaving CEOs from withholding effort and making damage. Ittner et al. (2003) documents that the importance of the retention objective has a significant positive influence on equity grants to newly hired key employees. Balsam and Miharjo (2007) suggest that the negative relationship between voluntary turnover and the intrinsic value of unexercisable in-the-money options, the time value of unexercised options, and the value of restricted shares indicates a retention consideration.

Besides, Banker et al. (2013) study the moral hazard and adverse selection problem reflected in the cash compensation. Hanlon et al. (2003) look for evidence of long-term incentive in the significantly positive relationship between value of stock option and future operating income. Knechel et al. (2013) find implicit incentives provided to Big 4 audit partners.

Second, executive compensation also interacts with other incentive problems and/or monitoring mechanisms. Ortiz-Molina (2007) examines the simultaneity of CEO compensation and capital structure which reflects the interest alignment problem of shareholders and debt-holders. The paper finds that pay-for-performance sensitivity decreases in straight-debt leverage but increases with convertible debt. Stock option policy, among all compensation components, is most sensitive to differences in capital structure. The relationship between CEO compensation and the independence of compensation consultants are studied by Murphy and Sandino (2010) and Cadman et al. (2010). Karuna (2007) examines the influence

of product market competition on executive compensation and Aggarwal et al. (2012) find that pay-performance incentive is negatively related to board size. Ferri and Sandino (2009) find CEO pay decreased in firms in which the proposal was approved relative to a control sample of S&P500 firms, suggesting a role of shareholders' activism. Roulstone (2003) finds that insider trading restrictions explain the cross-sectional difference in the level of total compensation and incentive-based compensation and equity-based incentive.

**1.2.1.2 Consequences of compensation contracts** Even though direct tests on the firm performance improvement attributed to incentive are rare except Aboody et al (2010) who find option repricing increases operating income and cash flows, there exist quite a few papers documenting the effects of compensation contract on executives' managerial activities.

As to financing and investing activities, Young and Yang (2011) reveal a positive association between stock repurchases and earnings per share (EPS)-contingent compensation and suggest net benefits to shareholders from this association. Cheng and Farber (2008) suggest a decrease in option-based compensation reduces CEOs' incentives to take excessively risky investments, resulting in improved profitability. Rajgopal and Shevlin (2002) stock options provide managers with incentives to mitigate risk-related incentive problems.

A series of papers study how compensation contracts influence financial disclosures. McAnally et al (2008) find that some managers may seek to miss earnings targets and benefit from lower strike price on subsequent option grants. Armstrong et al. (2010) find accounting irregularities occur less frequently at firms where CEOs have relatively higher levels of equity incentives. Comrix and Muller (2006) use more income-increasing accounting estimates of pension income when pension income has greater effect on CEO cash compensation. Nagar et al (2003), stock price-based compensation provides incentive to disclose private informa-

tion. Erkens (2011) finds that firms use time-vested stock-based pay to reduce the leakage of R&D-related information to competitors through employee mobility. Mastsunaga and Park (2001) find that CEOs tend to meet analyst forecast in the same quarter of last year.

Other behaviors are studied by Armstrong et al (2012) who find that tax directors are incentivized to reduce tax expenses and Adams and Ferreira (2008) who find director attendance is sensitive to monetary incentives.

**1.2.1.3 Design of compensation contracts** Accounting-based performance measures are extensively examined. Tian et al (2012) look at the earnings component and find that discretionary accrual receives less weight in CEOs' terminal year compensation. Boschen et al (2003) examine the cumulated unexpected good performance and document that CEOs' long-run cumulative financial gain from unexpectedly good accounting performance is not significantly different from zero, but that from unexpectedly good stock price performance is significantly positive. Indiejikian and Nanda (2002) find that CEOs' target bonuses are negatively associated with a proxy for measurement noise in accounting-based performance measures, and positively associated with proxies for firms' growth opportunities and the extent of executives' decision-making authority. Bushman et al. (2006) suggests that the two roles of accounting information, that is valuation and incentive contracting, are related. Cash compensation puts more weight on non-accounting public information captured by stock returns. Banker et al. (2009) confirm the relation of the two roles. Bushman et al. (2004) study the role of earnings timeliness in contract design.

Non-accounting performance measures are also investigated. Stock price-based compensation is studied by Jayaraman and Milbourn (2012) who find positive relationship between pay-for-performance sensitivity and stock liquidity, Hanlon et al. (2003) who find that stock

option grant value is positively related to future operating income, which is discussed by Larcker (2003), Leone et al (2006) who find that asymmetric sensitivity of CEO cash compensation to stock returns reflects that boards intend to reduce ex post settling up in cash compensation. Dechow (2006) discusses this paper and cannot rule out other explanations.

Other performance measures studied include non-profit performance measure in hotel managers' compensation contracts (Banker et al., 2000) and implicit financial incentives in big 4 audit partners' compensation (Knechel et al., 2013).

### **1.2.2 Analyzing policies**

Empirical research is expected to not only evaluate the consequences of previously adopted policies but also predict the outcome of potential not-yet-adopted policies. However, the latter goal requires a good understanding of policy-invariant factors in the decision-making process of both shareholders and executives. Such knowledge can be hardly obtained with traditional empirical method in accounting literature and thus is not provided. By contrast, structural model approach which is relatively new to accounting literature has a comparative advantage in this perspective and will be introduced soon. Before that, I review several papers that evaluate the consequences of various policies.

The Sarbanes-Oxley Act has received much attention. Engel et al (2010) find that audit committee compensation increases due to higher demand for monitoring after SOX. Carter et al (2009) find that the weight of earnings increase in CEOs bonus increased with a decrease in upward earnings management and the cash salary components decreased in the total compensation after SOX. Nekipelov (2007) who estimates a structural model of a linear contract in the apparel retail industry attributes the increase in executive compensation (salary and bonus) across the passage of SOX to the increase of executive managers' risk

aversion. Cohen et al. (2007) document a decline in the pay-for-performance sensitivity after SOX.

Some other policies affecting executive compensation are examined too. Iskandar-Datta and Jia (2013) find the adoption of clawback provisions do not influence either the level or the design of CEOs' compensation contract. Chan et al. (2012) finds that accounting restatements decline after firms initiate such provisions. Ozkan et al. (2012) find that the improved earnings quality and comparability after the adoption of IFRS increases accounting-based pay-for-performance sensitivity (PPS) and RPE. Skantz (2012) suggests that the voluntary option expensing under SFAS 123 may have encouraged inefficiency in CEO pay and the mandatory expensing under SFAS 123(R) may have contributed to the reduction in that inefficiency.

### **1.3 A COMPARISON BETWEEN REDUCED-FORM APPROACH AND STRUCTURAL APPROACH**

Structural approach is usually contrasted with reduced-form approach which is more familiar to accounting researchers and presented in section 1. Except the crucial differences to be discussed soon, it is equally important to realize that the structural approach and the reduced-form approach have two things in common. First, each of the two approaches can accomplish the two tasks in section 1, even though they take different procedures in testing theories and come up with different metrics in policy analyses. Second, both approaches provide quantitative understandings of economic concepts by estimating variables of interest, even though the variables are selected based on research questions that each approach is good at answering.

### 1.3.1 Reduced-form approach

**1.3.1.1 Definition** To clarify, reduced-form approach can have multiple meanings. First, reduced-form refers to the simultaneous equation regression in which all endogenous variables only appear on the left hand side and they are explicitly represented as functions of the exogenous explanatory right hand side variables and unobservables (Reiss and Wolak, 2007).

Second, reduced-form approach may refer to quasi-experimental design that identifies and estimates treatment effect. This treatment effect approach is compared with the structural approach by Heckman and Vytlačil (table V, 2005) and surveyed by Imbens and Wooldridge (2009). This line of research focuses on the effects defined by quasi-experiments, rather than parameters which have explicit economic meanings in theoretical models. Schroeder (2010) introduces treatment effect approach with accounting applications.

Third, reduced-form papers may use explicit economic models to motivate and interpret empirical analyses and they approximate the economic models using simple econometric techniques. Chetty (2009) reviews the sufficient statistic approach in public economic studies in which the welfare analyses are not directly based on deep primitives but instead on sufficient statistics derived from economic models.

**1.3.1.2 Research challenges** To accomplish the two tasks, that is, testing theories and analyzing policies, reduced-form studies encounter at least three challenges. First, to test contract theory, reduced-form approach takes an indirect way by testing implications of models. It appeals to testing comparative statics implied by the equilibria of theoretical models but leaves model structures and assumptions implicit. In order to stay close to the underlying theoretical models, keeping all other things equal is required for this type of tests (Heckman, 2000). This requirement becomes the main challenge, because quite often those

control variables implied by economic models can not be measured or observed.

Second, tests on incentive effects, which try to detect causal effects due to the adoption of incentive devices, often encounter endogeneity problems. One standard solution is to exploit instrumental variables to make the explanatory variables truly exogenous. However, the econometric problems associated with weak instrumental variables render this method unsatisfactory (Larcker and Rusticus, 2010).

Third, to conduct policy analysis, this approach uses a difference-in-difference research design and the policy change is treated as a natural experiment. The key issue here is to find and justify the control group. It turns to be challenging when certain policies are universally adopted by firms whose data researchers have access to. For example, the lack of control groups in most of the studies on SOX gives rise to mixed results, as Leuz (2007) and Dey (2010) point out. Accounting researchers become more serious about the above econometric issues. A group of thought-provoking discussions emerges in Chenhall and Moers (2007), Larcker and Rusticus (2004, 2007), and Van Lent (2007).

### **1.3.2 Structural approach**

**1.3.2.1 Definition** By contrast, structural approach refers to “a branch of economics in which economic theory and statistical method are fused in the analysis of numerical and institutional data” (Hood and Koopmans, 1953, pp. xv). Nowadays, researchers refer to models that combine explicit economic theories with statistical models as structural econometric models.

What separate structural models from nonstructural models is how clearly the connections are made between institutional, economic, and statistical assumptions and the estimated relationships between variables of interest. (Reiss and Wolak, 2007) The structural

approach allows a seamless connection between economic theory and econometric estimation. Under the structural approach, researchers analyze in rigorous theoretical terms how people optimize in face of incentive mechanisms. Structural econometricians use the implications of those mechanisms explicitly as a basis for their empirical investigation.

**1.3.2.2 How it works** To facilitate the comparison, here is a brief introduction of how the structural approach works in the context of executive compensation research. A more detailed illustration is in section 4. The goal of this approach is to make inference about unobservable primitive variables from available data on executive compensation and stock returns. When shareholders design optimal compensation contracts, they act as if they solve an optimization problem based on some primitive variables. We use a theoretical model to characterize the properties of shareholders' optimization problem. Solving the model gives the optimal compensation and a set of equilibrium restrictions. These restrictions are functions of compensation, stock returns, and primitives. They discipline the data and the deeper parameters together, so that we can analyze them consistently within the same framework and mitigate the empirical problem of missing variables.

These restrictions tell us theoretically how the parameters interact with the observables. Along with exclusion restrictions, they help us uniquely recover those parameters from the data. This crucial step is called identification. Then, by examining the consistency between the observed data pattern and the theoretical restrictions derived from the unobservables, we can look for the estimates of parameters that minimize the loss function in this comparison between population and sample properties. Eventually we can test the model by comparing the theoretical restrictions and the sample version of the restrictions. Also, armed with the time/policy-invariant parameters of preferences and technology that are recovered from

historical data and based on the theoretical model, we can predict the potential responses to a policy change which has never happened.

## **1.4 WHEN DO WE NEED STRUCTURAL APPROACH?**

Abowd and Kaplan (1999) propose six questions to answer in studies of executive compensation. They are (1) how much does executive compensation cost the firm? (2) how much is executive compensation worth to the recipient? (3) how well does executive compensation work? (4) what are the effects of executive compensation? (5) how much executive compensation is enough? (6) could executive compensation be improved? Both the reduced-form and structural papers need to answer questions (1) and (2). These measurement issues have been discussed by Antel and Smith (1985, 1986), Core and Guay (2002), and Hall and Leibman (1998). The reduced-form approach and the structural approach complement each other in answering remaining questions with each own comparative advantages.

### **1.4.1 Research questions and advantages of reduced-form approach**

Reduced-form papers can answer question (4) by detecting managerial behaviors driven by certain incentives embedded in compensation contracts, which have been summarized in section 1. Overall, reduced-form approach mainly answers yes-or-no type of questions and focuses on the sign (direction) of association/causality rather than attempts to quantify causal effects.

However, reduced-form approach has its own merits on at least three aspects. First, papers with this approach can use simple econometric techniques to document robust empirical regularities evidenced by statistically significant non-zero coefficients, for example, the

noise-signal trade-off in weighting performance measures in contract design.

Second, this approach can support the existence of certain effect, which may inspire more sophisticated investigation using structural models. Third, reduced-form papers can examine phenomena on which no theory has explained: Masulis et al. (2012) documents that US firms with foreign independent directors (FIDs) are associated with a greater likelihood of intentional financial misreporting and higher CEO compensation.

#### **1.4.2 Research questions and advantages of structural approach**

Compared with reduced-form approach, studies taking the structural approach are able to answer a set of questions that cannot be answered by the reduced-form research.

As to testing theory, first, the structural approach evaluates the predicting performance of an economic model as a whole in order to distinguish between competing theories that may be all able to rationalize the data generating process. The structural approach emphasizes the internal consistency in empirical investigation. The consistency is guaranteed by explicitly building empirical analysis on economic models and compensates for the reduction of inference credibility due to using structures. When researchers pull all equilibrium restrictions, the structural parameters discipline the data within the same framework. For example, the risk aversion parameter affects executives' decisions on both participation and exerting effort rather than shirking, and the technology captured by the distribution parameters of outcome are shared by both shareholders and executives in each party's optimization problem.

Second, this approach makes transparent a track on assumptions which the rejection of models is attributed to or which are required to draw causal economic inferences from the distribution of data (for example, Gayle and Miller (2012)). This explicit tracking enables

empiricists to provide informative feedback to theoretical research, given that theorists care about to what extent their models can help rationalize stylized facts. Only when we bring theoretical structures literally to data, we can realize to what extent the theoretical structures can be recovered from the data we want to understand. This is an important way to advance our knowledge by empirical research. By contrast, reduced-form approach tends to appeal to suboptimality/irrationality to explain the rejection of hypotheses which are derived from economic models and thus seems to be less informative.

As to policy analysis, this approach can estimate primitive parameters which are time-invariant and/or policy-invariant. Such robustness of estimation makes extrapolation reliable and results comparable across studies. Those estimates are used to conduct counterfactual analysis and welfare analysis (in both the evaluation and prediction of policies), When changes in executives' well-beings are unobserved, a direct estimation of deadweight loss is not possible. However, we can draw inferences about executives' preferences over risk and effort and firms' productivity from observed compensation and stock returns through structural parameter estimation. This information can help us predict the welfare changes for a policy that has not yet been implemented. Instead of looking for a control group, the counterfactual analysis uses the primitive parameters in structural models as an anchor and compares the variables of interest before and after a policy based on the same research subjects. It is appealing because social experiments, especially at the executive level, can be almost impossible merely for a trial-and-error purpose.

## 1.5 IMPLEMENTING STRUCTURAL APPROACH

Nevo and Whinston (2010) summarize two significant changes in empirical work since Leamer's (1983) article which criticized the state of applied econometric practice. On one hand, econometric methods have been developed such as nonparametric and semiparametric estimation (Powell, 1994) and identification based on minimal assumptions (Manski, 2003; Tamer, 2010). On the other hand, structural models have been increasingly used. Below I present the procedures of structural approach, based on a static single-agent moral hazard model for the illustration purpose.<sup>1</sup>

- Step 1: Build an economic model

A well-defined economic model serves as the theoretical underpinning of a structural analysis. This economic model is expected to capture the first order effect reflected in the nonexperimental data under consideration. Structural modelers need to select between alternative modeling options while building the economic model, although those options may not give qualitatively different results in theoretical studies. As theorists, we take the following steps to build a principal-agent model.

- (1.a) specify preferences and technologies

The economic model is built on players' utilities which rely on primitive parameters that represent the preferences of both the principal and the agent in the simple moral hazard model. In the context of executive compensation, the principal represents shareholders or board and the agent represents a manager, for example the CEO.

---

<sup>1</sup>Guidelines of implementing the structural approach in other fields can be found at Reiss and Wolak (2007, empirical industrial organization), and Strebulaev and Whited (2012, corporate finance). For nonparametric application, see Matzkin (2007).

We need to consider modeling questions such as whether the magnitude of CEO's risk aversion is affected by his wealth, whether the managerial effort reduces CEO's utility additively or multiplicatively from his pecuniary well-being, and whether the inefficiency due to hidden action should be attributed instead to CEO's limited liability as well, etc.. Answers to these questions ask for some respect on institutional knowledge.

Also, the technology needs to be specified. For example, between a model with continuous effort and one with discrete effort choices, which one would allow us to draw meaningful inference about how much shareholders would lose if they failed to align CEO's interest?

- (1.b) specify information structure and strategic interactions between players

We need to define the common knowledge and the information asymmetry between contracting parties. In a typical moral hazard model, CEO's effort is assumed to be unobservable to shareholders, but preferences and technologies are common knowledge to both parties.

- (1.c) model and solve optimization problems with endogenous and exogenous variables

Researchers need to clearly state the constrained optimization problem for shareholders to solve and managers' possible strategies. The solutions of the optimization problem, either explicit or implicit, and equilibrium restrictions are derived. It is important to distinguish between endogenous variables (determined within model) and exogenous variables (determined out of model), for at least two reasons. Comparative statics that are based on the sensitivity of endogenous variables to exogenous variables can provide testable predications. What's more, in counterfactual analysis, researchers are interested in knowing how welfare that usually depends on endogenous variables will vary with exogenous shocks.

- Step 2: Transit from an economic model to an econometric model

The transition from a theoretical economic model to an empirical econometric model is accomplished by introducing stochastic components into the economic model. This is the watershed where a theoretical model and a structural model depart. Below are the major steps.

- (2.a) define observable and unobservable variables

The goal of empirical studies is to make statistical inferences about unobservables from observables. In addition to the classification of endogenous and exogenous variables, another key classification of variables in a structural model is "observable vs unobservable" from the perspective of researchers instead of players in the theoretical model. This classification depends on what data is available to researchers. For example, risk aversion and personal effort costs are common knowledge in a moral hazard model. In such a sense, they are "observable" to the players. However, they cannot be directly measured by empiricists, so they are unobservable. By contrast, CEO's effort choice is unobservable to both shareholders and researchers, but the realization of performance measure can be observed by both players in the model and researchers.

- (2.b) introduce stochastic components into the theoretical model

According to Reiss and Wolak (2007), there are potential four ways to introduce stochastic components. I discuss them in the context of the simple moral hazard model of executive compensation. The first channel is researchers' uncertainty about contracting environment. It refers to what researchers do not know in the contracting environment and has been answered by step (2.a). The second channel is players' uncertainty about contracting environment. It refers information asymmetry between shareholders and CEOs, which has been discussed in step (1.b).

The third channel is optimization errors on the part of players. It allows players to behave not so rationally as the model predicts, but the deviation from rationality should be independent conditional on other variables of interest. For example, the executive compensation may associate with multiple period stock returns. A static model or repeated short-term model cannot capture it.

The fourth channel is measurement errors in observed variables. For example, we assume that the optimal compensation cannot be directly observed, but instead we can observe the compensation with errors.

For the above stochastic components, we need to make assumptions on both their functional forms and distributions. For example, does the error term enter in an additive way or a multiplicative way into the regression of optimal compensation? Is it necessary to specify a parametric distribution for a random variable? Both Margiotta and Miller (2000) and Gayle and Miller (2012) include an additive error item into the optimal compensation regression. However, the former parameterizes the distribution of performance measure conditional on equilibrium effort as truncated normal, but the latter leaves that distribution to be nonparametrically identified.

- Step 3: Identify the structural model

Identification concerns the empirical investigation with population values of parameters or features of a structural model. Identification is crucial in the structural approach. From one structural model, we can derive a reduced-form model. However, the uniqueness of its reverse process is not always guaranteed. The same observed empirical regularity can be generated by two completely different structural models. In such a case, so called identification failure, the two structural models are observationally equivalent. In other words, the rationale for the data cannot be uniquely determined even if we have infinite data. Identification

failure automatically implies inconsistency in estimation.

Take the compensation gap as an example. As we know from step 1, the optimal compensation is the solution to shareholders' optimization problem and is a function of primitive parameters representing preferences and technologies (or informativeness of performance measure). The gap between two executives' compensation is essentially determined by the differences of their primitive parameters. A model in which the two executives have homogeneous preferences but different technologies and a model in which the two executives have heterogeneous preferences but same technology can give rise to the same observed compensation gap. The primitive parameter values underlying the two models and the implications of the two models are distinct in principle. As a result, it is necessary to investigate whether the available data can distinguish between these two models before estimating any features of either model. This argument motivates chapter 2.

Another example is the outside option in the moral hazard model. Margiotta and Miller (2000) discusses the incomplete identification of this part in their model. Briefly, without further information on the demand and supply of managerial efforts, we cannot distinguish the outside option from the multiplicative effort cost in CEO's utility. We can only identify their ratio.

- (3.a) explore the sources of identification

One source of identification comes from equilibrium conditions derived from the model. They can be equality restrictions or inequality restrictions. The relationships between endogenous variables and exogenous variables and those between observable variables and unobservables together discipline the data and parameters. These relationships can help us set up a mapping from the joint distribution of observable variables to the structures of the model. An N-to-one mapping implies there exist multiple equilibria, but the identification

can still be achieved. However, a one-to-N mapping indicates identification failure. The key is to prove the uniqueness of the inverse process.

Another source of identification is exclusion restrictions. By excluding an exogenous variable from the moment conditions generated by equilibrium restrictions, we obtain more orthogonal moments and identification power.

- (3.b) choose between point identification and set identification

A parametric model with equality restrictions usually can be point identified. However, when a structural model involves strategic interactions, preferences some times are revealed through inequalities in equilibrium. These inequality restrictions, if they are exploited in order to fully represent the model, in nature prevent the model from point identification. Instead, researchers can only achieve set identification with confidence regions of parameters.

- Step 4: Estimate the structural model

Only after we prove that a structural model can be identified from the data, we can move forward to estimation. A traditional GMM estimator can be used if equilibrium restrictions that constitute the moment conditions only incorporate explicit solutions of the theoretical model. Otherwise, simulated moments may be used.

- Step 5: Application in testing theories and analyzing policies

I leave this step to chapter 2 for testing theories and to a continuing project Gayle et al. (2013) for analyzing policies..

## 1.6 PLANS FOR CHAPTER 2 AND CHAPTER 3

Chapter 2, as a response to the first task, uses structural approach and nonparametric method to test three multi-agent moral hazard models of top management teams. This chapter emphasizes the importance and advantages of the structural approach in distinguishing among possible models that can be observationally equivalent in rationalizing the same dataset.

Chapter 3, as a response to the second task, conducts nonparametric analysis on the potential effects of the governance rules enacted around the year 2002 on CEOs' compensation and emphasizes the importance of a careful reduced-form investigation before conducting a fully structural analysis.

## 2.0 MUTUAL MONITORING WITHIN TOP MANAGEMENT TEAMS: A STRUCTURAL MODELING INVESTIGATION

### 2.1 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."<sup>1</sup> 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

---

<sup>1</sup>A 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.

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.<sup>2</sup> However, the optimal compensation is usually derived from primitive parameters<sup>3</sup> 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 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

---

<sup>2</sup>Evidence 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).

<sup>3</sup>For 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.

(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.<sup>4</sup> 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.<sup>5</sup>

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 opti-

---

<sup>4</sup>This 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.

<sup>5</sup>This 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.

mal 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 shareholders 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.<sup>6</sup> 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 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)<sup>7</sup> 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

---

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

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

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 monitoring 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.<sup>8</sup> 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.

---

<sup>8</sup>The 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).

## 2.2 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.

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.

### 2.2.1 Timeline

In a static model, the timeline of the interaction between the risk-neutral shareholders and the two risk-averse managers<sup>9</sup> 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

---

<sup>9</sup>It 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.

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 teams<sup>10</sup>. 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 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.

---

<sup>10</sup>This assumption rules out the revelation mechanism like Ma (1988).

### 2.2.2 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.<sup>11</sup>

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 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

---

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

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.<sup>12</sup> 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.

---

<sup>12</sup>All proofs are in Appendix A.

Formally,

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

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  $h(x) \in \{g_1(x), g_2(x), g(x)\}$ . 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,

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

### 2.2.3 Managers' Preferences

Each manager's preference can be expressed using a negative exponential utility function with multiplicatively separable preference on effort.<sup>13</sup> The two managers have the same coefficient of absolute risk aversion, denoted by  $\rho$ , 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 (2.1) and (2.2), 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  $(j, k)$ . Formally,

$$\text{Manager 1's expected utility} \equiv -\tilde{\alpha}_{1j} E[\exp(-\rho w_1(x)) \mid j, k], \quad (2.1)$$

$$\text{Manager 2's expected utility} \equiv -\tilde{\alpha}_{2k} E[\exp(-\rho w_2(x)) \mid j, k]. \quad (2.2)$$

In particular, on the equilibrium path, manager  $i$  gets his expected utility from compensation

---

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

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  $-\tilde{\alpha}_{i1} \int v_i(x)g(x)f(x)dx$ .

## 2.2.4 Shareholder's Cost Minimization Problem

**2.2.4.1 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:

$$\int [\ln v_1(x) + \ln v_2(x)] f(x) dx. \tag{2.3}$$

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.

**2.2.4.2 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  $-\tilde{\alpha}_0$ .<sup>14</sup> 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

---

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

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:

$$-\alpha_{12} \int v_1(x) f(x) dx \geq -1, \tag{2.4}$$

$$-\alpha_{22} \int v_2(x) f(x) dx \geq -1. \tag{2.5}$$

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 firms<sup>15</sup>, 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*

---

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

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,

$$-\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx \geq -2. \quad (2.6)$$

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.<sup>16</sup> 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 (2.7) is the expected utility of manager 1 (2) given both managers shirking. On the left-hand (right-hand) side of equation (2.8) is the expected utility of manager 1 (2) given both managers working:

$$-\alpha_{11} \int v_1(x)f(x)g(x)dx = -\alpha_{21} \int v_2(x)f(x)g(x)dx, \quad (2.7)$$

$$-\alpha_{12} \int v_1(x)f(x)dx = -\alpha_{22} \int v_2(x)f(x)dx. \quad (2.8)$$

---

<sup>16</sup>More generally, if the equal sharing rule is relaxed, the ratio of  $\alpha_{1j}$  and  $\alpha_{2j}$  will incorporate the relative bargaining power/allocation weight. Under this interpretation, the weight cannot be separately identified, but does not need to be half-half any more.

**2.2.4.3 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 (2.9), 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 ( $\alpha_{11}$ ) by the utility from monetary compensation. And the expectation is taken over the distribution of the gross abnormal return conditional on that manager 1 unilaterally shirks. The inequality (2.10) applies the same constraint, which provides working incentive to manager 2:

$$-\alpha_{12} \int v_1(x) f(x) dx \geq -\alpha_{11} \int v_1(x) f(x) g_1(x) dx, \quad (2.9)$$

$$-\alpha_{22} \int v_2(x) f(x) dx \geq -\alpha_{21} \int v_2(x) f(x) g_2(x) dx. \quad (2.10)$$

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,

$$-\alpha_{12} \int v_1(x) f(x) dx - \alpha_{22} \int v_2(x) f(x) dx, \quad (2.11)$$

$$\geq -\alpha_{11} \int v_1(x) f(x) g(x) dx - \alpha_{21} \int v_2(x) f(x) g(x) dx. \quad (2.12)$$

A caveat is that in this model, I implicitly assume that both working strictly Pareto dominates unilateral shirking<sup>17</sup>. In principle, the optimal compensation schemes also need to satisfy the other two inequality constraints such that both working Pareto dominates

---

<sup>17</sup>If 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 (2.11) 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 the data-generating process.

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.<sup>18</sup>

Note that the empirical optimal contracting approach of this paper assumes that the 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.<sup>19</sup> 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 condi-

---

<sup>18</sup>Formally, to guarantee that both working is Pareto dominant over either manager unilaterally shirking, the current compensation scheme needs to satisfy the following inequalities:

$$\begin{aligned} -\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx &> -\alpha_{11} \int v_1(x)f(x)g_1(x)dx - \alpha_{22} \int v_2(x)f(x)g_1(x)dx \\ -\alpha_{12} \int v_1(x)f(x)dx - \alpha_{22} \int v_2(x)f(x)dx &> -\alpha_{12} \int v_1(x)f(x)g_2(x)dx - \alpha_{21} \int v_2(x)f(x)g_2(x)dx. \end{aligned}$$

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.

<sup>19</sup>If 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.

tional 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 (2.13) is the incentive compatibility constraint for manager 1, and (2.14) is for manager 2:

$$-\alpha_{12} \int v_1(x)f(x)dx \geq -\alpha_{11} \int v_1(x)f(x)g(x)dx, \quad (2.13)$$

$$-\alpha_{22} \int v_2(x)f(x)dx \geq -\alpha_{21} \int v_2(x)f(x)g(x)dx. \quad (2.14)$$

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.

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.

### 2.2.5 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  $\alpha_{ij}$  and  $g_i(x)$  are the same as previously defined,  $\mu_1$  is the shadow price associated with manager 1's incentive compatibility constraint and  $\mu_2$  with manager 2's,  $w_i^*(x)$  is the optimal compensation paid to manager  $i$ .

**Proposition 2.**

$$w_1^*(x) = \frac{1}{\rho} \ln \alpha_{12} + \frac{1}{\rho} \ln \left[ 1 + \mu_1 - \mu_1 \left( \frac{\alpha_{11}}{\alpha_{12}} \right) g_1(x) \right], \quad (2.15)$$

$$w_2^*(x) = \frac{1}{\rho} \ln \alpha_{22} + \frac{1}{\rho} \ln \left[ 1 + \mu_2 - \mu_2 \left( \frac{\alpha_{21}}{\alpha_{22}} \right) g_2(x) \right]. \quad (2.16)$$

*In the no mutual monitoring model,  $w_i^*(x)$  has exactly the same expression in (2.15) and (2.16). 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,  $\mu_1$  and  $\mu_2$  are equal because of the group incentive compatibility constraint. In the mutual monitoring with individual utility maximization model,  $\mu_1$  and  $\mu_2$  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 ( $\rho$ ), their effort tastes ( $\alpha_{ij}$ ), 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.

### 2.2.6 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 ( $\alpha_{i1}$ ). 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

$$\begin{aligned}
E[V * x - w_1^*(x) - w_2^*(x)] \geq \max\{ & E[(V * x - w_1^s - w_2^*(x)) * g_1(x)], \\
& E[(V * x - w_1^*(x) - w_2^s) * g_2(x)], \\
& E[(V * x - w_1^s - w_2^s) * g_1(x) * g_2(x)]\}. \quad (2.17)
\end{aligned}$$

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,

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

### 2.2.7 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.

## 2.3 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.

### 2.3.1 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]

## 2.3.2 Key Variables in the Optimal Contracts

### 2.3.2.1 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]

**2.3.2.2 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),<sup>20</sup> I construct the total compensation by adding 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

---

<sup>20</sup>Among followers are Hall and Liebman (1998), Margiotta and Miller (2000), Gayle and Miller (2009, 2012), and Gayle et al. (2011).

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.

**2.3.2.3 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.

### 2.3.3 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 (2.1) and (2.2) now has a new expression:

$$-\alpha_{ij} \frac{1}{b_{t+1}} E_t \left[ \exp \left( \frac{-\rho w_{it}(x_t)}{b_{t+1}} \right) \mid j, k \right], \quad (2.18)$$

where  $b_t$  is the bond price in year  $t$ , which pays a unit of consumption per period forever.<sup>21</sup> 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.

---

<sup>21</sup>See the detailed construction of the bond prices in Gayle and Miller (2009, page 1748-1749).

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 reinterpretation makes the models fit the framework of Margiotta and Miller (2000).<sup>22</sup> 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.

## 2.4 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

---

<sup>22</sup>This paper is descended from Grossman and Hart (1983) and Fudenberg et al. (1990).

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 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  $\rho^*$ , 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.

### 2.4.1 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  $\alpha_{ij}$ , 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  $\rho$ . 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,  $\alpha_{i2}$  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 (2.4) and (2.5) when the equalities

hold, we can find that only the disutility coefficients  $\alpha_{i2}$  and the risk aversion parameter  $\rho$  are unknown. This indicates that if  $\rho$  can be identified, then  $\alpha_{i2}$  can be expressed as a mapping of  $\rho$  and the observables. In this sense,  $\alpha_{i2}$  are identified respectively for  $i = 1, 2$  up to the risk aversion parameter as follows:

$$\alpha_{12}(\rho) = E_t[v_{1t}(x, \rho)]^{1-b_t}, \quad (2.19)$$

$$\alpha_{22}(\rho) = E_t[v_{2t}(x, \rho)]^{1-b_t}. \quad (2.20)$$

Next, I consider the likelihood ratios  $g_{it}(x)$  for  $i = 1, 2$ . In the formula of optimal compensation (2.15) and (2.16), 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  $\rho$  and some quantities that can be calculated from the data-generating process:

$$g_{1t}(x, \rho) = \frac{1/v_{1t}(x, \rho) - 1/v_{1t}(\bar{x}_1, \rho)}{E[1/v_{1t}(x, \rho)] - 1/v_{1t}(\bar{x}_1, \rho)}, \quad (2.21)$$

$$g_{2t}(x, \rho) = \frac{1/v_{2t}(x, \rho) - 1/v_{2t}(\bar{x}_2, \rho)}{E[1/v_{2t}(x, \rho)] - 1/v_{2t}(\bar{x}_2, \rho)}. \quad (2.22)$$

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,  $\alpha_{i1}$  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 (2.9) and (2.10) as well as the two equalities held in the participation constraints (2.4) and (2.5). 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:

$$\alpha_{11}(\rho) = E_t[v_{1t}(x, \rho)g_{1t}(x, \rho)]^{1-b_t}, \quad (2.23)$$

$$\alpha_{21}(\rho) = E_t[v_{2t}(x, \rho)g_{2t}(x, \rho)]^{1-b_t}. \quad (2.24)$$

Again, these two formulas imply that for any known risk aversion parameter  $\rho$ , the shirking disutility coefficient  $\alpha_{i1}$  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 (2.15) 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  $\mu_1$

and the risk aversion parameter are left unknown. The same procedure applies to identifying the shadow price for manager 2 ( $\mu_2$ ). Consequently, the two shadow prices can be expressed as the mappings of the risk aversion parameter, as follows:

$$\mu_1(\rho) = E_t [v_{1t}(x, \rho)] / v_{1t}(\bar{x}, \rho) - 1, \quad (2.25)$$

$$\mu_2(\rho) = E_t [v_{2t}(x, \rho)] / v_{2t}(\bar{x}, \rho) - 1. \quad (2.26)$$

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 (2.17). 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 (2.27), (2.28), and (2.29).  $\Lambda_{1t}$  ( $\Lambda_{2t}$ ) 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,  $\Lambda_{3t}$  reflects that shareholders' net benefit is also larger than that of having both

managers shirk:

$$\Lambda_{1t}(\rho) = E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[(V * x - w_{1t}^s - w_{2t}^*(x)) * g_{1t}(x, \rho)] \geq 0, (2.27)$$

$$\Lambda_{2t}(\rho) = E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[(V * x - w_{1t}^*(x) - w_{2t}^s) * g_{2t}(x, \rho)] \geq 0, (2.28)$$

$$\begin{aligned} \Lambda_{3t}(\rho) &= E[V * x - w_{1t}^*(x) - w_{2t}^*(x)] - E[(V * x - w_{1t}^s - w_{2t}^s) * g_{1t}(x, \rho) * g_{2t}(x, \rho)] \\ &\geq 0, \end{aligned} (2.29)$$

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,

$$w_{it}^s = \frac{b_{t+1}}{\rho(b_t - 1)} \ln \alpha_{i1}(\rho), \text{ for } i = 1, 2. (2.30)$$

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 (2.31) ((2.32)) 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:

$$\begin{aligned}\Lambda_{4t}(\rho) &= \left\{ -\alpha_{12}(\rho)^{\frac{1}{b_t-1}} E[v_{1t}(x, \rho)g_{2t}(x, \rho)] \right\} \\ &\quad - \left\{ -\alpha_{11}(\rho)^{\frac{1}{b_t-1}} E[v_{1t}(x, \rho)g_{1t}(x, \rho)g_{2t}(x, \rho)] \right\} > 0,\end{aligned}\tag{2.31}$$

$$\begin{aligned}\Lambda_{5t}(\rho) &= \left\{ -\alpha_{22}(\rho)^{\frac{1}{b_t-1}} E[v_{2t}(x, \rho)g_{1t}(x, \rho)] \right\} \\ &\quad - \left\{ -\alpha_{21}(\rho)^{\frac{1}{b_t-1}} E[v_{2t}(x, \rho)g_{1t}(x, \rho)g_{2t}(x, \rho)] \right\} > 0.\end{aligned}\tag{2.32}$$

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 (2.21) and (2.22) 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  $\Lambda_{kt}$  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

$$\Gamma_{\text{N-M}} \equiv \{\rho > 0 : Q_{\text{N-M}}(\rho) = 0\}. \quad (2.33)$$

#### 2.4.2 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 (2.25) and (2.26) 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 (2.25) and (2.26), 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.<sup>23</sup>

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 ( $w_{it}^s$ ) 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

$$\Gamma_{M-T} \equiv \{\rho > 0 : Q_{M-T}(\rho) = 0\}. \quad (2.34)$$

---

<sup>23</sup>This function-wise restriction is constructed in a way similar to the nonnegative restriction on likelihood ratio imposed in Gayle and Miller (2012).

### 2.4.3 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

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

I define  $\Gamma_{\text{M-I}}$ , a set of the risk aversion parameter admissible to this model, as

$$\Gamma_{\text{M-I}} \equiv \{\rho > 0 : Q_{\text{M-I}}(\rho) = 0\}. \quad (2.35)$$

### 2.4.4 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

$$\theta_{\text{N-M}} \equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_{1t}(x), g_{2t}(x), \mu_1, \mu_2),$$

$$\theta_{\text{M-T}} \equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_t(x), \mu),$$

$$\theta_{\text{M-I}} \equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, g_t(x), \mu_1, \mu_2).$$

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  $\rho^*$ , then  $\theta_M^*$  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  $\Gamma_M$  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.<sup>24</sup>

**Proposition 4.** *Consider any data generating process  $(x_n, w_n)$  that satisfies  $w_n = w(x_n)$  for*

---

<sup>24</sup>A 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.

$\forall n$ . Define  $\Gamma_M$  as before for each  $M \in \{N-M, M-T, M-I\}$ . If  $\Gamma_M$  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  $\Gamma_M$  is empty, then  $(x_n, w_n)$  is not generated by the model  $M$ .

## 2.5 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 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  $\rho$  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

$$H_0^M : Q_M(\rho) = 0 \text{ for some } \rho > 0, \text{ i.e., the model } M \text{ cannot be rejected}$$

$$H_A^M : Q_M(\rho) > 0 \text{ for all } \rho, \text{ i.e., the model } M \text{ is rejected.}$$

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  $M \in \{N-M, M-T, M-I\}$  as follows:

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

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  $Q_{N-M, ZS}^{(N)}(\rho)$  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.

## 2.6 RESULTS

### 2.6.1 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."<sup>25</sup>

[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.

---

<sup>25</sup>For a manager with risk aversion parameter  $\rho$ , 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  $CE = \frac{-b}{\rho} \ln EU$ .

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 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.<sup>26</sup> 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

---

<sup>26</sup>These assumptions rule out the possibility of achieving the first best allocation with risk neutral managers.

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 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.

## 2.6.2 Discussion

**2.6.2.1 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 ( $\rho$ ) 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  $prob(x|j)$ ,

that is,  $f(x)$  or  $f(x)h(x)$  in the continuous case. In particular,  $p \equiv \text{prob}(x|\text{work})$  and  $q \equiv \text{prob}(x|\text{shirk})$ ; 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, $-i$ <b>work</b>	$i$ shirk, $-i$ <b>shirk</b>
$x_H$	$p$	$q_N (< p)$	$q_M (< p)$
$x_L$	$1 - p$	$1 - q_N$	$1 - q_M$

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  $x \in \{x_H, x_L\}$ ;  $\rho$  is the risk aversion parameter, and  $\alpha_{ij}$  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 inequality:

$$\begin{aligned}
& p * [-\alpha_{i2}e^{-\rho w_i(x_H)}] + (1 - p) * [-\alpha_{i2}e^{-\rho w_i(x_L)}] \\
& \geq q * [-\alpha_{i1}e^{-\rho w_i(x_H)}] + (1 - q) * [-\alpha_{i1}e^{-\rho w_i(x_L)}] \\
\implies & (\alpha_{i2}p - \alpha_{i1}q)e^{-\rho w_i(x_H)} \leq (\alpha_{i1} - \alpha_{i2} + \alpha_{i2}p - \alpha_{i1}q)e^{-\rho w_i(x_L)} \\
\implies & e^{-\rho[w_i(x_H) - w_i(x_L)]} \leq \frac{\alpha_{i1} - \alpha_{i2}}{\alpha_{i2}p - \alpha_{i1}q} + 1. \tag{2.36}
\end{aligned}$$

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 (2.36)) and fixed wage spread,  $\rho$  is expected to be smaller in these two models, which rationalize the same data as the no mutual monitoring model does.

**2.6.2.2 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.

**2.6.2.3 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  $\rho$  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  $\rho$  and  $q$ , and the mutual monitoring with total utility maximization model works with low values in both  $\rho$  and  $q$ . Even if the group incentive works well, it is not necessary that the contract be flat. A small  $\rho$  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.

## 2.7 EXTENSION

### 2.7.1 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.<sup>27</sup> The first measure of moral hazard cost, denoted by  $\tau_1$ , is the difference between the expected output from both managers working and that from at least one manager shirking, namely, the losses shareholders would incur from managers shirking instead of working, or

$$\tau_1 \equiv E[Vx] - \max\{E[(Vxg_{1t}(x, \rho))], E[(Vxg_{2t}(x, \rho))], E[(Vxg_t(x, \rho))]\}. \quad (2.37)$$

---

<sup>27</sup>Similar analysis can be found in Margiotta and Miller (2000) and Gayle and Miller (2009).

The second measure of moral hazard cost, denoted by  $\tau_{2i}$ , 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  $i = 1, 2$ :

$$\tau_{2i} \equiv w_{i2}^o - w_{i1}^o \quad (2.38)$$

$$= \frac{b_{t+1}}{\rho(b_t - 1)} [\ln \alpha_{i2}(\rho) - \ln \alpha_{i1}(\rho)]. \quad (2.39)$$

The third measure of moral hazard cost, denoted by  $\tau_{3i}$ , 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  $i = 1, 2$ :

$$\tau_{3i} \equiv E[w_i(x)] - w_{i2}^o. \quad (2.40)$$

It could be interesting to investigate the efficiency of the optimal compensation contract by contrasting  $\tau_1$  with  $\tau_{2i}$  and  $\tau_{3i}$ .

### 2.7.2 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.<sup>28</sup> 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 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

---

<sup>28</sup>MacLeod 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.

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.

## 2.8 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.

## 3.0 DO 2002 GOVERNANCE RULES AFFECT CEOS' COMPENSATION?

### 3.1 INTRODUCTION

The Sarbanes-Oxley Act of 2002 (SOX) is a legislative response taken by the U.S. government to a wave of corporate governance failures at many prominent companies, along with several other amendments to the U.S. stock exchanges' regulations.<sup>1</sup> Existing studies have investigated how this set of governance rules enacted in 2002 affects firm behaviors, for example, switching the method of earnings management<sup>2</sup>, reducing investment<sup>3</sup>, and going private/dark<sup>4</sup>. However, the influences of 2002 governance rules on CEOs' compensation are under-explored. This is the focus of this paper.

The importance of examining CEOs' compensation is first determined by the goals that the 2002 governance rules are expected to achieve. One primary goal of these rules is to improve the corporate governance of U.S. firms, for example, mitigating the agency problems in incentive alignment between shareholders and top executives. One important incentive de-

---

<sup>1</sup>A timeline of these rules can be found at Chhaochharia and Grinstein (2007).

<sup>2</sup>Cohen et al (2008) finds that accrual-based earnings management declined after the passage of SOX, but the real earnings management increased at the same time.

<sup>3</sup>Bargeron et al. (2010) finds that, compared with non-US firms, US firms reduced investment in R&D and capital. This finding is consistent with Cohen et al. (2007) and confirms the view that SOX has discouraged corporate risk-taking in investment. Kang et al. (2010) find that firms apply a higher rate to discount the payoff of investment projects and firms with good governance, with a credit rating, and with early compliance of SOX 404 have become more cautious about investment.

<sup>4</sup>Engel et al. (2006) find that small firms chose to go private to avoid the cost of SOX. Leuz et al. (2007) show that the increased deregistration is mainly driven by firms that go dark, rather than private.

vice used by shareholders is executive compensation contract. Naturally, in order to examine the (un)intended effects of 2002 governance rules on the U.S. economy, their consequences for CEOs' compensation seem to be one significant aspect.

Also, as the survey by Murphy (2012) summarizes, "government intervention has been both a response to and a major driver of time trends in executive compensation over the past century, and that any explanation for pay that ignores political factors is critically incomplete". Along this line of research, this paper attempts to answer whether 2002 governance rules have influenced CEOs' compensation. The findings in this paper can enrich the knowledge about how the private compensation contracts in S&P 1500 firms react to the governmental regulations, in the context of 2002 governance rules.

Even though in policy analysis a comprehensive evaluation based on welfare analysis seems more desirable<sup>5</sup>, a careful examination on the changes in the basic properties of CEOs' compensation contract, as a first pass test, is always needed. This paper just does this. Intuitively, if simple tests on the changes of compensation curvature and distribution of performance measure, as this paper does, indicate that there is no significant change in these basic properties of executive compensation, then more sophisticated welfare estimation based on structural models may lose its credibility.

So far researchers have got only limited results on the consequences of 2002 governance rules on CEOs' compensation. Carter et al. (2009) find that in the post-SOX era the weight of earnings increase in CEOs' bonus increased and the cash salary components decreased in the total compensation. Cohen et al. (2007) document a decline in the pay-for-performance sensitivity after 2002.

---

<sup>5</sup>In a continuing project Gayle et al. (2013), we estimate a structural model of both moral hazard and hidden information and attempt to compare the agency costs associated with these two agency problems across the year 2002.

This paper is different from previous studies in several ways. First, CEOs' compensation consists of total wealth. CEOs care about their overall wealth change implied by their compensation packages. Following the concept of current income equivalent first adopted by Antle and Smith (1985, 1986), and later used by Hall and Liebman (1998) and Margiotta and Miller (2000), I construct the total compensation by adding wealth change in options held and wealth change in stocks held into other regular components provided by the ExecComp database including salary, bonus, options, restricted stocks, etc.. The wealth change in holding stocks is equal to the beginning shares of held stocks multiplied by the raw abnormal returns. By holding the options from existing grants rather than disposing this part of wealth into a market portfolio, CEO obtains the net of ending option value and beginning option value multiplied by market portfolio return. This net value is the wealth change in holding options. Including the opportunity costs of holding firm-specific equity enables us to fully capture the incentive that shareholders impose on CEOs.

Second, I apply nonparametric estimation and test in this paper. I assume there are measurement errors in the observed compensation and use a nonparametric regression to estimate the optimal compensation as a function of stock returns. Then, I conduct a nonparametric test on the change of compensation contract shape and a test on the change of the distribution of performance measure that is based on stock returns. The nonparametric estimation of the optimal compensation relies merely on the empirical distribution of the stock returns rather than a particularly specified contractual form. This approach gives us more flexibility in comparing the curvature of compensation and drives the attention to any prompt rational response in the optimal contract to 2002 governance rules.

Third, the design of empirical investigation is derived from a structural model by Gayle and Miller (2012). Their model incorporates two agency problems, namely moral hazard and

hidden information. Accounting information is assumed to convey CEOs' hidden information about firms' prospect. The compensation contract shape can be different between the two private states as they define based on accounting return. Following this intuition, this paper tests the change in contract shape and the change in the distribution of performance measure not only for each public state specified by industry, firm size, and capital structure, but also for private state.

Section 2 discusses the provisions in SOX and how they may affect those two agency problems and correspondingly CEOs' compensation contracts. This discussion helps me justify using the structural model of Gayle and Miller (2012) to evaluate the consequences of 2002 governance rules, from which the research design of this present paper is derived.

Section 3 discusses the data used in this chapter. I compile compensation data from Compustat ExecComp, accounting data from Compustat Fundamentals Annual, and stock market data from Compustat PDE. The sample covers 2,818 firms and 6,450 CEOs over fiscal years from 1993 to 2005, which amounts to 24,535 observations. The size of the sample is mostly limited up to the compensation data available in the ExecComp.

In section 4, using a nonparametric approach, I conduct a probability density equality test on the change in the distribution of gross abnormal returns (performance measure) and a model specification test on the change in the optimal contract shape from the Pre-2002 period to the Post-2002 period. I found that both changes are significant, which is consistent with Holmstrom and Kaplan (2003) who suggest that the overall corporate governance system in U.S. can react quickly to address those problems evidenced by collapse of business monsters.

Section 5 concludes.

## 3.2 BACKGROUND

The 2002 governance rules can influence CEOs' compensation by changing the contracting environment in terms of two types of agency problems, that is, hidden action problem due to the information asymmetry between shareholders and a CEO in a firm on the CEO's productive efforts, namely moral hazard, and hidden information problem due to the information asymmetry between the two contracting parties on the firm's states. CEOs may have superior knowledge about firm prospects. This section uses provisions of SOX to illustrate the potential influences, since SOX is more comprehensive than other contemporaneous rules.

As to the effect of SOX on the moral hazard problem, its provisions serve as a double-edged sword. For example, SOX Section 302 requires that the principal executive officer(s) and the principal financial officer(s) should be responsible for establishing and maintaining internal controls and for disclosing all significant deficiencies. This requirement may divert CEOs' productive efforts and the stock market may not price the improvement on internal controls. As a result, the performance measure (stock returns) may turn to be noisier and the CEOs' cost of working for good performance of the stock returns increases as well. Consequently, the cost of motivating the CEO to work increases.

However, other requirements may help align the interest between shareholders and CEOs. SOX Section 304, which requires that the chief executive officer and chief financial officer of the issuer shall reimburse the issuer for any compensation received during the 12-month period following equity issue filing due to misconduct in financial statement for that equity issue. Such regulation makes the CEOs' compensation less liquid, so it can mitigate their incentives to make myopic investment and to take opportunistic advantage by misreporting financial states. As a result, the interests between shareholders and managers are aligned

more closely, which help shareholders mitigate the moral hazard problem.

As to the effect of SOX on the hidden information problem, the prediction seems less misty. For example, SOX Section 302 requires that the principal executive officer(s) and the principal financial officer(s) certify in each annual or quarterly report filed or submitted that the financial statements and other financial information include fairly present financial conditions and results and do not contain any misleading statement. Enforcing truthful statement of financial conditions makes the potential punishment on misreporting become higher after 2002. As a result, the cost of inducing truth telling would be lower from the perspective of shareholders.

### **3.3 DATA**

#### **3.3.1 State variables**

To facilitate discussions on abnormal returns and compensation which are based on states, I introduce the construction of state variables first. They have clear meanings in the underlying economic model (Gayle and Miller, 2012) that inspires the reduced-form analysis in this paper. One is public state, which is observable to both shareholders and CEOs at the beginning of each contract period. This type of state variables is common knowledge and does not invite any cost to reveal. The other one is private state, which is observable merely to CEOs after they enter into the contract. Shareholders receive the report on the private state variable from CEOs rather than observe that directly by themselves. In optimal contracts, the private state variable is subject to truth telling constraints and induces cost to reveal. I construct these two types of state variables with data available to us as follows.

**3.3.1.1 Public state** I use industry and time varying firm characteristics to generate the public state variables. First, I classify the whole sample into three industrial sectors according to Global Industry Classification Standard (GICS) code, denoting by  $J_{nt}$  for the  $n$ th firm in year  $t$ . The Primary sector ( $J_{nt} = 1$ ) includes firms in energy (GICS: 1010), materials (GICS: 1510), industrials (GICS: 2010, 2020, 2030), or utilities (GICS: 5510). The Consumer good sector ( $J_{nt} = 2$ ) includes firms in consumer discretionary (GICS: 2510, 2520, 2530, 2540, 2550) or consumer staples (GICS: 3010, 3020, 3030). The Service sector ( $J_{nt} = 3$ ) includes firms in health care (GICS: 3510, 3520), financial (GICS: 4010, 4020, 4030, 4040), or information technology and telecommunication services (GICS: 4510, 4520, 5010).

Second, I use categorical variables based on firm size and capital structure (debt-to-equity ratio). The firm size is measured by the total assets on balance sheet at the end of period  $t$  and denoted by  $A_{nt}$ . The capital structure is reflected by the debt-to-equity ratio and denoted by  $C_{nt}$ . Each of the two variables can have two values, i.e. Small ( $S$ ) or Large ( $L$ ). If the total asset of firm  $n$  in year  $t$  is below the median of total asset in its sector,  $A_{nt} = S$ , otherwise  $A_{nt} = L$ . Same rules apply to  $C_{nt}$ . For each firm in a given sector and year, the public state could be one of the four possible combinations with regard to  $(A_{nt}, C_{nt})$ , i.e.  $(A_S, C_S)$ ,  $(A_S, C_L)$ ,  $(A_L, C_S)$ ,  $(A_L, C_L)$ .

Finally, I construct an aggregate indicator variable,  $Z_{nt} = (J_n, A_{nt-1}, C_{nt-1})$ , to denote the observable state. Data used to measure  $Z_{nt}$  comes from Compustat.

The top two rows in Table 1 describe yearly change of firm size (total assets) and that of capital structure (debt-to-equity ratio) respectively, without distinguishing among industrial sectors. The firm size has been increasing and the increasing trend started around late 1990s before the time of corporate governance scandals and subsequent rules. The capital structure presents a smoother pattern.

The top two rows in Table 2 display more aggregate time-series pattern but for each industrial sector. The public state variables are compared between the two periods, before 2002 and after 2002. The total asset increased after 2002 and the debt-to-equity ratio decreased after 2002. The two dimensions of public state does not move systematically and simultaneously, which justifies the necessity of considering both together. Table 2 shows cross-sectional characteristics of total assets, and debt-to-equity. In the Primary sector, both the total asset and debt-to-equity ratio increased after 2002, but in the other two sectors, only the firm size increases.

**3.3.1.2 Private state** After accepting the contractual arrangement, CEOs collect and convey their private information on the firm prospect. The measure of the private state is constructed by equity return evaluated at book value, which is consistent with the concept of comprehensive income in accounting practice. Accounting numbers features the private state in the theoretical framework because a lot of estimations are used to generate accounting numbers. For example, accrual, defined as the difference between realized cash flow and reported earnings, is one of the typical features of accounting as an information system. The smoothing over periods require information about the state of firm which may be excluded from shareholders especially in modern firms where the control right and ownership are separated. Based on estimation, the accounting numbers can convey private information about prospect to shareholders.

Specifically, I define the private state as a binary variable,  $S_{nt}$ .  $S_{nt} = Bad$  if the accounting return  $r_{nt}$  is lower than the average for all firms within the same observable state  $Z_{nt}$  in

year  $t$ , otherwise  $S_{nt} = Good$ .

$$r_{nt} = \frac{Asset_{nt} - Debt_{nt} + Dividend_{nt}}{Asset_{n,t-1} - Debt_{n,t-1}} \quad (3.1)$$

The third row in Table 1 describes the yearly change of accounting returns. It experienced a drop around year 2000, again before the time of the governance rules. Table 2 shows cross-sectional characteristics of the accounting return in Pre-2002 and Post-2002 period respectively. Accounting return is highest in service sector before 2002 and in primary sector after 2002 and the dispersion is highest in service sector whenever. Also, accounting returns increased in primary sector and decreased in the other two after 2002.

**3.3.1.3 Distribution of the states** Table 6 displays the sample distribution across the eight states (4 public \* 2 private) for each sector. The number in the column Total is the number of observations in the corresponding public state. No matter in Pre-2002 or Post-2002 period, the sample clusters in two states, i.e.  $(A_S, C_S)$  and  $(A_L, C_L)$ . The column Bad/Good reports the ratio of sample size in the Bad state to that in the Good state given certain public state. Overall the ratios are close to one, though it is more often that the Bad state has slightly more observations than the Good state. This implies that the two private states have balanced size and the accounting return is right-skewed.

### 3.3.2 Abnormal Stock Returns

I get raw prices and adjustment factors from the Compustat PDE dataset. For each firm in the sample, I calculate monthly compounded returns adjusted for splitting and repurchasing for each fiscal year, and subtract return to a value-weighted market portfolio (NYSE/NASDAQ/AMEX) from this raw return to get the raw abnormal return for its cor-

responding 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 and consequently comparable with each other.

Table 1 displays the time-series pattern of abnormal stock returns. Though firms outperformed market in those booming years and the abnormal returns drop after year 2002, the standard deviations have been very high.

Table 2 compares cross-sectional characteristics of raw abnormal returns between Pre-2002 and Post-2002 periods. After 2002, the abnormal returns increased in all of the three sectors. The most profitable sector was service sector before 2002 and switched to primary sector after 2002, but the largest dispersion in abnormal returns has been in service sector whenever. The cross sectional variation and time series fluctuation in abnormal returns partially induces the variation and fluctuation in compensation discussed later.

To be more relevant to the interest in 2002 governance rules, in Table 5, I further contrast abnormal returns from Pre-2002 period with those from Post-2002 period by both public and private states. Consistent with what has been found in Table 2, Primary sector has increased raw abnormal returns in all states after 2002. In Consumer Goods sector, abnormal returns of small firms increase after 2002, but large firms show decreasing abnormal returns except firms with low debt-to-equity ratio and in the bad state. In Service sector, abnormal returns of small firms increase after 2002, but large firms again show decreasing abnormal returns except firms with high debt-to-equity ratio and in the bad state. Also, no matter in which sector and public state, the abnormal returns in good state is always first order stochastic dominate those in bad state and the divergence between private states tends to be larger than that among public states for a given sector, indicating that the private state variable I am using can predict outcome well, which is required by the principal-agent model with

hidden information in state. Overall, firms present different abnormal return distribution in different states.

### **3.3.3 Compensation**

CEOs care about their overall wealth change implied by their compensation packages. In the ExecComp database, available to us 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. CEOs' wealth changes with their holdings in firm-specific equity as well. They can always offset the aggregate risks imposed in their compensation package by adjusting their market portfolio but cannot avoid being exposed to non-diversifiable risks of holding firm stocks and options. As a result, CEOs' wealth changes in holding firm-specific equity are reflected in the value change given that they cannot diversify those idiosyncratic risks.

Following the concept of wealth change adopted by Antle and Smith (1985, 1986), Hall and Liebman (1998), and Margiotta and Miller (2000), I construct the total compensation by adding wealth change in options held and wealth change in stocks held into other regular components like salary, bonus, options, restricted stocks, and so on. The wealth change in holding stocks is equal to the beginning shares of held stocks multiplied by the raw abnormal returns. By holding the options from existing grants rather than disposing this part of wealth into a market portfolio, CEO obtains the net of ending option value and beginning option value multiplied by market portfolio return. More detailed procedure goes to Appendix.

Table 3 shows the time-series pattern of each component as well as the total compensation. The documented soaring CEOs' compensation seems to be inverse after 2002. Also, I

find that the level and fluctuation in equity-based compensation components have more influence on those of the total compensation, consistent with the previous analysis and justifying the importance of adopting a more comprehensive measure of CEOs' wealth.

Table 4 summarizes total compensation by public and private states, again contrasting Pre-2002 against Post-2002 period. The trend between the two periods is consistent with that observed in abnormal returns in Table 5. The post-2002 compensation is always higher or insignificantly lower than the pre-2002 compensation in all states in the three sectors. Compensation in bad states is lower than that in good states all over the sample. Also, large firms seem to pay higher compensation, which is consistent with previous findings from time-series change of compensation that the moral hazard cost can explain executive compensation and large firms have more severe problems to be compensated (Gayle and Miller, 2009).

### 3.4 NONPARAMETRIC TESTS AND RESULTS

I conduct a probability density equality test on the change in the distribution of gross abnormal returns (performance measure) and a model specification test on the change in the optimal contract shape from the Pre-2002 period to the Post-2002 period. I found that both changes are significant, which is consistent with Holmstrom and Kaplan (2003) who suggest that the overall corporate governance system in U.S. can react quickly to address those problems evidenced by collapse of business monsters.

In the structural estimation discussed in next subsections, the effort costs corresponding to working and shirking both change across the passage of 2002. These results imply that the productivity changes, no matter it is captured by the managerial input (captured by

effort costs) or by the output (captured by the distribution of the gross abnormal return). As a response to the changes in these fundamentals, the optimal contract changes as well.

Before moving to the structural model identification and estimation, I first explore the empirical pattern of the gross abnormal return and optimal compensation for two reasons. First, they are key elements in the model such that their changes from Pre-2002 to Post-2002 period reflect the essential changes in other structural parameters, especially those measures of agency costs. Second, the distribution of gross abnormal returns and the curvature of the optimal compensation, which I focus on in this section, can be both nonparametrically estimated before I introduce more complicated structures as I need for other primitives, for instance the risk-aversion parameter. In this section, I briefly describe the method used to derive consistent estimators of these two variables, nonparametrically test on their changes over the two periods and report the results of testing statistics.

### **3.4.1 Estimating Optimal Compensation and Performance Measure**

Equity-based compensation is designed to align the interests of CEOs to shareholders and consequently eliminate the moral hazard problem. While including stock returns into the performance measure metric, I am aware of two issues. First, the stock return which is used as a performance measure in the optimal contract should be closely tied to CEOs' efforts but eliminate stochastic variations that are out of CEOs' control. Second, the performance measure should reflect the outcome sharing between shareholders and CEOs, that is, reflect returns before compensation payment.

Taking into account these two points, I construct the performance measure, gross abnormal return as I call, in the following steps. First I subtract the market portfolio return from the annual return to a firm stock in the same corresponding fiscal year. The residual

captures the idiosyncratic components in firm stock returns. This non-diversifiable portion generates the incentive to work rather than shirk. Given that neither the gross abnormal return nor the optimal compensation can be directly observed from the data, I construct consistent estimators of them as discussed below.

$\tilde{x}_{nt}$  is the raw abnormal returns and  $\tilde{w}_{nt}$  is the total compensation of firm n in year t.  $(Z_{nt}, S_{nt})$  are state variables defined previously. First I nonparametrically estimate the optimal compensation by running the following regression<sup>6</sup>

$$\hat{w}_{nt} = \frac{\sum_{m=1, m \neq n}^N \tilde{w}_{mt} * I\{Z_{mt} = Z_{nt}, S_{mt} = S_{nt}\} K\left(\frac{\tilde{x}_{mt} - \tilde{x}_{nt}}{h_x}, \frac{v_{m,t-1} - v_{n,t-1}}{h_v}\right)}{\sum_{m=1, m \neq n}^N I\{Z_{mt} = Z_{nt}, S_{mt} = S_{nt}\} K\left(\frac{\tilde{x}_{mt} - \tilde{x}_{nt}}{h_x}, \frac{v_{m,t-1} - v_{n,t-1}}{h_v}\right)} \quad (3.2)$$

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 by

$$x_{nt} \equiv \tilde{x}_{nt} + \frac{\hat{w}_{nt}}{v_{n,t-1}} \quad (3.3)$$

Now the consistent estimate of optimal compensation conditional on state  $(Z, S)$  is given by

$$w_{nt}(x_{nt}|Z, S) = \frac{\sum_{m=1, m \neq n}^N \hat{w}_{mt} * I\{Z_{mt} = Z, S_{mt} = S\} K\left(\frac{x_{mt} - x_{nt}}{h_x}\right)}{\sum_{m=1, m \neq n}^N I\{Z_{mt} = Z, S_{mt} = S\} K\left(\frac{x_{mt} - x_{nt}}{h_x}\right)} \quad (3.4)$$

### 3.4.2 Test on the Change in the Distribution of Gross Abnormal Return

**3.4.2.1 Test statistic** Given that the optimal compensation depends on the underlying distribution of the performance measure (gross abnormal return in the paper), I first test the

---

<sup>6</sup> $K(\cdot)$  is a multivariate standard normal kernel density function as below.

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

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

change in the distribution of gross abnormal return from Pre-2002 to Post-2002. The riskier the distribution of gross abnormal returns is in the mean-spread sense, the more costly is for shareholders to motivate risk-averse CEOs.

The probability density function (PDF) of gross abnormal return  $x_{nt}$  is nonparametrically estimated by

$$f(x_{nt}|Z, S) = \frac{\sum_{m=1}^N I\{Z_{mt} = Z, S_{mt} = S\} K\left(\frac{x_{mt} - x_{nt}}{h_x}\right)}{\sum_{m=1}^N I\{Z_{mt} = Z, S_{mt} = S\}} \quad (3.5)$$

I have two series of gross abnormal returns  $\left\{x_{nt}^{(N)}\right\}_{pre}^{n=\overline{1, N_1}}$  from Pre-2002 period and  $\left\{x_{nt}^{(N)}\right\}_{post}^{n=\overline{1, N_2}}$  from Post-2002 period. Assuming that  $\left\{x_{nt}^{(N)}\right\}_{pre}^{n=\overline{1, N_1}}$  has PDF  $f_{pre}(\cdot)$  and  $\left\{x_{nt}^{(N)}\right\}_{post}^{\overline{1, N_2}}$  has PDF  $f_{post}(\cdot)$  respectively. To test whether the underlying gross abnormal returns distribute differently between Pre-2002 versus Post-2002 period is equivalent to test on the equality between the PDF in the two periods, i.e. test  $H_0^{PDF} : f_{pre}(x) = f_{post}(x)$  for almost all  $x$ . Based on the test for equality of PDFs proposed by Li and Racine (2007, pp. 363), I calculate the statistics  $T^{PDF}$  by state and sector.  $T^{PDF}$  follows standard normal distribution  $N(0, 1)$ . Please refer to the appendix for details of  $T^{PDF}$  construction.

I have two series of gross abnormal returns  $\left\{x_{nt}^{(N)}\right\}_{pre}^{n=\overline{1, N_1}}$  and  $\left\{x_{nt}^{(N)}\right\}_{post}^{n=\overline{1, N_2}}$  from Pre-2002 and Post-2002 periods respectively. Assuming that  $\left\{x_{nt}^{(N)}\right\}_{pre}^{n=\overline{1, N_1}}$  has PDF  $f_{pre}(\cdot)$  and  $\left\{x_{nt}^{(N)}\right\}_{post}^{\overline{1, N_2}}$  has PDF  $f_{post}(\cdot)$  respectively. To test whether the underlying gross abnormal returns distribute differently before versus after 2002 is equivalent to test on the equality between the PDF of each of the two periods, or test  $H_0^{PDF} : f_{pre}(x) = f_{post}(x)$  for almost all  $x$ . If  $H_0^{PDF}$  is false, then the statistics  $T^{PDF}$  constructed below has  $P(T^{PDF} > C) \rightarrow 1$ , for any positive constant C. Based on the test for equality of PDFs proposed by Li and Racine

(2007, pp. 363), I calculate the statistics  $T^{PDF}$  by state and sector at

$$T^{PDF} = (N_1 N_2 h^2)^{1/2} \frac{(I_n^b - c_{n,b})}{\hat{\sigma}_b} \xrightarrow{d} N(0, 1)$$

where

$$I_n^b = \left(\frac{1}{N_1}\right)^2 \sum_{m=1}^{N_1} \sum_{n=1}^{N_1} K_{h,mn}^{x-pre} + \left(\frac{1}{N_2}\right)^2 \sum_{m=1}^{N_2} \sum_{n=1}^{N_2} K_{h,mn}^{x-post} - \frac{2}{N_1 N_2} \sum_{m=1}^{N_1} \sum_{n=1}^{N_2} K_{h,mn}^{x-pre,x-post}$$

$$c_{n,b} = \frac{k(0)}{h} \left[ \frac{1}{N_1} + \frac{1}{N_2} \right]$$

$$\hat{\sigma}_b^2 = \frac{h}{N_1 N_2} \left\{ \begin{aligned} & \sum_{m=1}^{N_1} \sum_{n=1}^{N_1} (N_2/N_1) (K_{h,mn}^{x-pre})^2 + \sum_{m=1}^{N_2} \sum_{n=1}^{N_2} (N_1/N_2) (K_{h,mn}^{x-post})^2 \\ & + 2 \sum_{m=1}^{N_1} \sum_{n=1}^{N_2} (K_{h,mn}^{x-pre,x-post})^2 \end{aligned} \right\}$$

and the kernel density functions are defined as

$$K_{h,mn}^{x-pre} \equiv (1/h) k_x \left( \frac{x_{mt}^{pre} - x_{nt}^{pre}}{h} \right)$$

$$K_{h,mn}^{x-post} \equiv (1/h) k_x \left( \frac{x_{mt}^{post} - x_{nt}^{post}}{h} \right)$$

$$K_{h,mn}^{x-pre,x-post} \equiv (1/h) k_x \left( \frac{x_{mt}^{pre} - x_{nt}^{post}}{h} \right)$$

$$k_x(\cdot) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_x} \exp \left\{ -\frac{1}{2} \left( \frac{x_{mt} - x_{nt}}{h} \right)^2 \right\}$$

$T^{PDF}$  is a one-sided test, therefore I reject  $H_0^{PDF}$  at level 1% if  $T^{PDF} > 2.33$ , reject  $H_0^{PDF}$  at level 5% if  $T^{PDF} > 1.64$ , and reject  $H_0^{PDF}$  at level 10% if  $T^{PDF} > 1.28$ .

**3.4.2.2 Result** Table 7.1 reports the statistics  $T^{PDF}$  for the 24 states (3 sectors \* 2 private states \* 4 public states). Except in the good state of  $(A_S, C_L)$  in the Primary sector (the value of the static is 0.93) and in the bad state of  $(A_S, C_L)$  in the Consumer Goods sector (the value of the static is 0.75), in all other states the values of  $T^{PDF}$  are above the critical value of 5% confidence level (1.64). Consequently, we can reject  $H_0^{PDF}$  in these

remaining 22 states.

This favorable result indicates that there may exist significant changes in the distribution of gross abnormal returns from Pre-2002 period to Post-2002 period. The changes in executive compensation, if any, cannot be only due to behavior changes of CEOs.

### 3.4.3 Test on the Change in the Optimal Contract Shape

**3.4.3.1 Test statistic** The test on the change in contract shape is equivalent to a model specification test on the significance of a dummy variable in the standard Nadaraya-Watson kernel regression of optimal compensation on the gross abnormal return. The dummy variable  $I^{2002}$  equals to 1 if the observation is from Post-2002 period and equals to 0 otherwise.

The null hypothesis states as follows.  $H_0^W : \Pr [w(x, I^{2002}) = W(x)] = 1$ , meaning that

the contract shape is not significantly different between Pre-2002 and Post-2002 period. The testing statistic for  $H_0^W$  is  $T^W$ , which follows the standard normal distribution  $N(0, 1)$ . Please refer to the appendix for the detail of constructing  $T^W$ .

The test on the change in contract shape is equivalent to a model specification test on the significance of a dummy variable in the standard Nadaraya-Watson kernel regression of optimal compensation on gross abnormal returns. The dummy variable  $I^{2002}$  equals to 1 if the observation is from Post-2002 period and equals to 0 otherwise.  $H_0^W : \Pr [w(x, I^{2002}) = W(x)] =$  ■

1 or the contract shape is not significantly different between Pre-2002 and Post-2002 period. If  $H_0^W$  is false, then the squared difference in nonparametric estimates of the functions  $w(x, I^{2002})$  and  $W(x)$  should be beyond certain critical values in the distribution of the

statistics  $T^W$ . I tend to reject  $H_0^W$  when  $T^W$  is large. Specifically,

$$T^W = \sigma_{11}^{-1} \left[ Nh^{(p+q)/2} \cdot \Gamma - h^{-(p+q)/2} \gamma_{12} - h^{(q-p)/2} \gamma_{22} - h^{(p+q)/2} H^{-p} \gamma_{32} \right] \\ \xrightarrow{d} N(0, 1)$$

where

$$\Gamma = \frac{1}{N} \sum_{n=1}^N \{w_n^h - W_n^H\}^2 A_n$$

I get two nonparametric estimators of optimal compensation,  $w_n^h$  and  $W_n^H$  and their densities  $f_n^h$  and  $f_n^H$  as follows.

$$w_n^h \equiv \frac{\sum_{m=1}^N w_m * I\{Z_m = Z_n, S_m = S_n\} K_h \left( \frac{x_m - x_n}{h_x}, I_m^{2002} \right)}{\sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_h \left( \frac{x_m - x_n}{h_x}, I_m^{2002} \right)}$$

$$W_n^H \equiv \frac{\sum_{m=1}^N w_m * I\{Z_m = Z_n, S_m = S_n\} K_H \left( \frac{x_m - x_n}{h_x} \right)}{\sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_H \left( \frac{x_m - x_n}{h_x} \right)}$$

$$f_n^h = \frac{1}{N} \sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_h \left( \frac{x_m - x_n}{h_x}, I_m^{2002} \right)$$

$$f_n^H = \frac{1}{N} \sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_h \left( \frac{x_m - x_n}{h_x} \right)$$

$A$  is a non-negative  $N$ -dimension weighting vector with element corresponding to each observation.  $A_n = 1$  if  $x_n$  falls into the 2.5%-97.5% range of gross abnormal returns within the same state identified by  $Z_n$  and  $S_n$ , otherwise  $A_n = 0$ .  $p$  is the dimension of non-tested regressors and  $q$  is the dimension of the tested regressors.  $p = q = 1$  in the context.

Also,

$$\sigma_{nh}^2 = \frac{\sum_{m=1}^N w_m^2 * I\{Z_m = Z_n, S_m = S_n\} K_h \left( \frac{x_m - x_n}{h_x}, I_m^{2002} \right)}{\sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_h \left( \frac{x_m - x_n}{h_x}, I_m^{2002} \right)} - (w_n^h)^2$$

$$\sigma_{nH}^2 = \frac{\sum_{m=1}^N w_m^2 * I\{Z_m = Z_n, S_m = S_n\} K_h\left(\frac{x_m - x_n}{h_x}\right)}{\sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_h\left(\frac{x_m - x_n}{h_x}\right)} - (W_n^H)^2$$

$$\sigma_{11}^2 = \frac{2C_{11}}{N} \sum_{n=1}^N \frac{(\sigma_{nh}^2)^2 A_n^2}{f_n^h}$$

$$C_{12} = \left(\frac{1}{2\sqrt{\pi}}\right)^{p+q}, C_{22} = \left(\frac{1}{\sqrt{2\pi}}\right)^p, C_{32} = \left(\frac{1}{2\sqrt{\pi}}\right)^p, C_{11} = \left(\frac{1}{2\sqrt{2\pi}}\right)^{p+q}$$

$$\gamma_{12} = \frac{C_{12}}{N} \sum_{n=1}^N \frac{\sigma_{nh}^2 A_n}{f_n^h}, \gamma_{22} = -\frac{2C_{22}}{N} \sum_{n=1}^N \frac{\sigma_{nh}^2 A_n}{f_n^H}, \gamma_{32} = \frac{C_{32}}{N} \sum_{n=1}^N \frac{\sigma_{nH}^2 \widetilde{A}_n}{f_n^H}$$

$$\widetilde{A}_n = \frac{\sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_H\left(\frac{x_m - x_n}{h_x}\right) A_m}{\sum_{m=1}^N I\{Z_m = Z_n, S_m = S_n\} K_H\left(\frac{x_m - x_n}{h_x}\right)}$$

$T^W$  is a one-sided test, therefore I reject  $H_0^W$  at level 1% if  $T^W > 2.33$ , reject  $H_0^W$  at level 5% if  $T^W > 1.64$ , and reject  $H_0^W$  at level 10% if  $T^W > 1.28$ .

**3.4.3.2 Result** In Table 7.2 I report the test static for the change of compensation contract shape ( $T^W$ ) by state and sector, in the same manor as it is done for the test on the change of the distribution of gross abnormal returns. In all states, the static is above 2.33. It implies that the null hypothesis  $H_0^W$  that there is no change in compensation contract shape is rejected.

This result suggests that the contract shape has significantly changed since 2002, and thus it is worth taking further studies to estimate the quantitative welfare effect and look for the driving forces behind such changes.

### 3.5 CONCLUSION

As a legislative response to those corporate governance failures seen by the beginning of this century, a set of rules including SOX and several other amendments on stock exchanges'

regulations were enacted. Regulatory intervention has been controversial because of the "one-size-fit-all" criticism. It is important to examine whether those rules have influenced the corporate governance of firms. One aspect is executive compensation which reflects both executives' attributes captured by CEOs' preference parameters in structural models and production technology captured by the distribution of performance measure.

As a first pass investigation, this paper nonparametrically tests the change of the distribution of the performance measure (gross abnormal returns) and the change of CEOs' compensation contract shape. I find that both changes are significant from Pre-2002 to Post-2002 period. These findings about the basic properties of CEOs' compensation suggest that the optimal compensation contracts between shareholders and CEOs may be adaptive to the exogenous shock of 2002. The change in CEOs' compensation may not be purely driven by behavioral change of CEOs but can also be the result of production change. Also, the significant change in contract shape indicates that agency costs embedded in the compensation contracts may change after 2002 as well. This invites a more sophisticated study using structural approach in Gayle et al. (2013).

## 4.0 APPENDIX TO CHAPTER 2

### 4.1 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:

$$g_1(x) \equiv \frac{f(x \mid \text{manager 1 shirks, manager 2 works})}{f(x \mid \text{manager 1 works, manager 2 works})} \quad (4.1)$$

$$= \frac{g_1(x)f(x)}{f(x)} \quad (4.2)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x \mid \text{manager 1 works, manager 2 shirks})} \quad (4.3)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x)} \quad (4.4)$$

$$g_2(x) \equiv \frac{f(x \mid \text{manager 1 works, manager 2 shirks})}{f(x \mid \text{manager 1 works, manager 2 works})} \quad (4.5)$$

$$= \frac{g_2(x)f(x)}{f(x)} \quad (4.6)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{f(x \mid \text{manager 1 shirks, manager 2 works})} \quad (4.7)$$

$$= \frac{f(x \mid \text{manager 1 shirks, manager 2 shirks})}{g_1(x)f(x)}. \quad (4.8)$$

Using (4.1) and (4.7) 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.  $\square$

*Proof of Proposition 2.* See the first-order conditions (FOCs) in the proof of Proposition 3 later.  $\square$

*Proof of Proposition 3. No Mutual Monitoring Model*

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

$$\begin{aligned}
L = & E [\ln v_{1t}(x) + \ln v_{2t}(x)] \\
& - \mu_1 \left[ \alpha_{12}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)] - \alpha_{11}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)g_1(x)] \right] \\
& - \mu_2 \left[ \alpha_{22}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)] - \alpha_{21}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)g_2(x)] \right] \\
& - \mu_3 \left[ \alpha_{12}^{\frac{1}{b_t-1}} E_t [v_{1t}(x)] - 1 \right] \\
& - \mu_4 \left[ \alpha_{22}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)] - 1 \right].
\end{aligned} \tag{4.9}$$

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

$$1/v_{1t}(x) = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} - \mu_1\alpha_{11}^{\frac{1}{b_t-1}}g_1(x). \tag{4.10}$$

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

$$1/v_{2t}(x) = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}} - \mu_2\alpha_{21}^{\frac{1}{b_t-1}}g_2(x). \quad (4.11)$$

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

$$1/v_{1t}(\bar{x}_1) = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} \quad (4.12)$$

$$1/v_{2t}(\bar{x}_2) = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}}. \quad (4.13)$$

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

$$E[1/v_{1t}(x)] = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} - \mu_1\alpha_{11}^{\frac{1}{b_t-1}} \quad (4.14)$$

$$E[1/v_{2t}(x)] = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}} - \mu_2\alpha_{21}^{\frac{1}{b_t-1}}. \quad (4.15)$$

The binding participation constraint for each manager gives

$$\alpha_{12}^* = E_t[v_{1t}(x)]^{1-b_t} \quad (4.16)$$

$$\alpha_{22}^* = E_t[v_{2t}(x)]^{1-b_t}. \quad (4.17)$$

The binding incentive compatibility constraint gives

$$\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g_1(x)] = \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g_2(x)] = 1. \quad (4.18)$$

Multiply both sides of (4.10) and integrate over  $f(x)$ ; it follows that

$$1 = (\mu_1 + \mu_3)\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] - \mu_1\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g_1(x)], \quad (4.19)$$

and plugging (4.16) and (4.17) into the preceding, it follows that

$$\mu_3 = 1.$$

Multiply both sides of (4.11) and integrate over  $f(x)$ ; it follows that

$$1 = (\mu_2 + \mu_4)\alpha_{22}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)] - \mu_2\alpha_{21}^{\frac{1}{b_t-1}} E_t [v_{2t}(x)g_2(x)], \quad (4.20)$$

and plugging (4.17) and (4.18) into the preceding, it follows that

$$\mu_4 = 1.$$

Multiplying (4.12) 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 (4.13) 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 (4.10), (4.12), and (4.14) 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

$$g_1(x) = \frac{1/v_{1t}(x) - 1/v_{1t}(\bar{x}_1)}{E[1/v_{1t}(x)] - 1/v_{1t}(\bar{x}_1)} \quad (4.21)$$

$$g_2(x) = \frac{1/v_{2t}(x) - 1/v_{2t}(\bar{x}_2)}{E[1/v_{2t}(x)] - 1/v_{2t}(\bar{x}_2)}. \quad (4.22)$$

Plug into (4.18); 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

$$\begin{aligned} L &= E_t [\ln v_{1t}(x) + \ln v_{2t}(x)] \\ &- \eta_0 \left[ \alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] - 2 \right] \\ &- \eta_1 \left\{ \begin{aligned} &\left[ \alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] \right] \\ &- \left[ \alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] + \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)] \right] \end{aligned} \right\}. \end{aligned} \quad (4.23)$$

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

$$1/v_{1t}(x) = \eta_0 \alpha_{12}^{\frac{1}{b_t-1}} + \eta_1 \alpha_{12}^{\frac{1}{b_t-1}} - \eta_1 \alpha_{11}^{\frac{1}{b_t-1}} g(x). \quad (4.24)$$

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

$$1/v_{2t}(x) = \eta_0 \alpha_{22}^{\frac{1}{b_t-1}} + \eta_1 \alpha_{22}^{\frac{1}{b_t-1}} - \eta_1 \alpha_{21}^{\frac{1}{b_t-1}} g(x). \quad (4.25)$$

Multiply both sides of (4.24) with  $v_{1t}(x)$  and then integrating over  $f(x)$ , we get

$$1 = (\eta_0 + \eta_1)\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] - \eta_1\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)]. \quad (4.26)$$

Similarly, from (4.25), we get

$$1 = (\eta_0 + \eta_1)\alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] - \eta_1\alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)]. \quad (4.27)$$

Recall that

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

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

$$1/v_{1t}(\bar{x}) = (\eta_0 + \eta_1)\alpha_{12}^{\frac{1}{b_t-1}} \quad (4.28)$$

$$1/v_{2t}(\bar{x}) = (\eta_0 + \eta_1)\alpha_{22}^{\frac{1}{b_t-1}}. \quad (4.29)$$

Binding participation constraint gives

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] = 2. \quad (4.30)$$

Binding incentive compatibility constraint gives

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] + \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] = \alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] + \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)]. \quad (4.31)$$

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

$$\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] = \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)]. \quad (4.32)$$

Similarly, if both work,

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] = \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)]. \quad (4.33)$$

Combining (4.30) and (4.31), we can immediately get

$$\alpha_{11}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)g(x)] = \alpha_{21}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)g(x)] = 1 \quad (4.34)$$

$$\alpha_{12}^{\frac{1}{b_t-1}} E_t[v_{1t}(x)] = \alpha_{22}^{\frac{1}{b_t-1}} E_t[v_{2t}(x)] = 1 \quad (4.35)$$

$$\alpha_{12}^* = E_t[v_{1t}(x)]^{1-b_t} \quad (4.36)$$

$$\alpha_{22}^* = E_t[v_{2t}(x)]^{1-b_t}. \quad (4.37)$$

Add (4.28) and (4.29). Then use binding IC and plug in (4.36) and (4.37):

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

Plug  $\eta_0^*$  into (4.28) and (4.29); 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

$$E[1/v_{1t}(x)] = (\eta_0 + \eta_1)\alpha_{12}^{\frac{1}{b_t-1}} - \eta_1\alpha_{11}^{\frac{1}{b_t-1}} \quad (4.38)$$

$$E[1/v_{2t}(x)] = (\eta_0 + \eta_1)\alpha_{22}^{\frac{1}{b_t-1}} - \eta_1\alpha_{21}^{\frac{1}{b_t-1}}. \quad (4.39)$$

Plug  $\eta_0^*$  and  $\eta_1^*$  into (4.28) and (4.38); we get

$$\alpha_{11}^* = \left[ \frac{E_t[v_{1t}(x)] - v_{1t}(\bar{x})}{1 - v_{1t}(\bar{x})E[1/v_{1t}(x)]} \right]^{1-b_t}. \quad (4.40)$$

Similarly, combining (4.29) and (4.39), we get

$$\alpha_{21}^* = \left[ \frac{E_t[v_{2t}(x)] - v_{2t}(\bar{x})}{1 - v_{2t}(\bar{x})E[1/v_{2t}(x)]} \right]^{1-b_t}. \quad (4.41)$$

Plug  $\eta_0^*$  and  $\eta_1^*$  into (4.24) and (4.25), respectively; using (4.38), (4.39), (4.32), and (4.33), 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.  $\square$

*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 confirmed by multiplying the Lagrangian multiplier with the associated constraint and finding that the product equals zero. Then the proposition is proved.  $\square$

## 4.2 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):<sup>1</sup>

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

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:

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

### 4.3 TABLES

---

<sup>1</sup> $K(\cdot)$  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.

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.

## 5.0 APPENDIX TO CHAPTER 3

### 5.1 CALCULATION OF WEALTH CHANGE IN HOLDING STOCK AND/OR OPTIONS

Due to the data availability, for each sample year, we can not exactly observe all the inputs of Black-Sholes formula for grants carried from years before 1993, the beginning year of our sample. Compustat ExecComp dataset only provides the valuation information for those options newly granted after year 1993, including number of underlying stock shares, exercise prices, expiration dates and issue dates. However, we need to know these Black-Sholes inputs for options granted before year 1993 to completely value the wealth change of CEOs by estimating the value of unexercised options and updating it each year. Instead, we assume that all options are not exercised until expiration dates. For the same reason, we apply FIFO rule to derive Black-Sholes inputs for options granted before year 1993, i.e. earlier issued options will be exercised earlier too. Together, we use the average length of holding period for each CEO to infer the issue dates and exercised prices for options granted before 1993. The same routines apply to those non-zero options granted before the year when the CEO entered our sample. We apply the dividend-adjusted Black-Sholes formula to re-evaluate the CEOs' call options for each CEO in each year. See footnote for the details.<sup>1</sup>

---

<sup>1</sup>Below  $c$  is the call option value,  $K$  is the exercise price,  $T_m$  is the time to maturity (in years),  $S$  is the

## 5.2 TABLES

---

underlying security price,  $q$  is the dividend yield,  $r$  is the risk free rate,  $\sigma$  is implied volatility.  $N(\cdot)$  defines a standard normal cumulative distribution function.

$$\begin{aligned}c &= S e^{-q T_m} N(d_1) - K e^{-r T_m} N(d_2) \\d_1 &= \frac{\ln(S/K) + (r - q + \sigma^2/2) T_m}{\sigma \sqrt{T_m}} \\d_2 &= d_1 - \sigma \sqrt{T_m}\end{aligned}$$

Table 1: Time-series Summary  
 (Asset in millions of 2006 US\$, Standard deviations in parentheses)

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Asset	6638 (18876)	6425 (18577)	6525 (19065)	6937 (21505)	6966 (22482)	7163 (24622)	8358 (29319)	9696 (35187)	10420 (40695)	10108 (38649)	10705 (42530)	11052 (47376)	12258 (56338)
Debt/Equity	2.640 (3.978)	2.466 (3.495)	2.568 (3.727)	2.466 (3.577)	2.343 (3.621)	2.408 (3.749)	2.546 (3.648)	2.638 (3.847)	2.474 (3.541)	2.416 (3.524)	2.467 (3.533)	2.385 (3.556)	2.272 (3.330)
Accounting Return	1.161 (0.275)	1.150 (0.247)	1.175 (0.277)	1.156 (0.277)	1.182 (0.303)	1.150 (0.292)	1.134 (0.285)	1.130 (0.334)	1.099 (0.296)	1.055 (0.283)	1.126 (0.228)	1.149 (0.268)	1.085 (0.238)
Abnormal Return	0.058 (0.321)	-0.013 (0.278)	-0.064 (0.364)	-0.038 (0.331)	-0.016 (0.385)	-0.156 (0.399)	-0.190 (0.425)	0.123 (0.470)	0.163 (0.369)	0.065 (0.331)	0.071 (0.354)	0.048 (0.317)	0.007 (0.298)
Observations	1434	1575	1550	1620	1690	1689	1501	1452	1506	1615	1580	1609	1531

Table 2: Cross-section Summary of Firm Characteristics  
(Asset in millions of 2006 US\$, Standard deviations in parentheses)

Sector	Primary		Consumer		Service	
	(1) Pre	(2) Post	(3) Pre	(4) Post	(5) Pre	(6) Post
Total Asset	4597 (7677)	6477 (11117)	3516 (8103)	5168 (12439)	13499 (42502)	17568 (70163)
Debt/Equity	1.819 (1.450)	2.102 (2.575)	1.571 (1.646)	1.537 (2.119)	3.650 (5.263)	2.982 (4.323)
Accounting Return	1.108 (0.228)	1.138 (0.217)	1.126 (0.282)	1.108 (0.253)	1.175 (0.335)	1.115 (0.261)
Abnormal Return	-0.031 (0.319)	0.063 (0.277)	-0.029 (0.383)	0.014 (0.320)	0.022 (0.433)	0.044 (0.353)
Observations	5331	1386	4013	1136	6288	2198

Table 3: Time Series of Compensation Components

Year	A. Mean (in thousands of US\$ (2006)) and Standard Deviation (in parentheses)													
	Salary		Bonus		Value of options Granted		Values of Restricted Stocks		Change in Wealth from Options Held		Change in Wealth from Stock Held		Total Compensation	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1993	656	(371)	506	(1116)	796	(2335)	144	(772)	453	(5144)	886	(12024)	3742	(14112)
1994	649	(367)	548	(1256)	1068	(2852)	159	(762)	786	(4843)	-39	(10134)	3409	(13158)
1995	652	(372)	552	(938)	1004	(2976)	173	(697)	308	(6481)	-282	(14361)	2730	(17425)
1996	649	(372)	618	(1107)	1412	(3767)	221	(981)	713	(8075)	234	(15413)	4199	(18957)
1997	646	(377)	715	(1476)	1888	(5528)	311	(1495)	1233	(11636)	510	(16379)	5709	(22388)
1998	650	(360)	655	(1193)	2072	(4982)	293	(1387)	-705	(12720)	-1388	(15781)	1978	(23375)
1999	672	(364)	754	(1475)	2488	(6309)	388	(2496)	419	(13558)	-3316	(16540)	1886	(24639)
2000	666	(373)	835	(1882)	3340	(8602)	459	(2119)	1618	(17070)	3425	(18441)	10930	(28789)
2001	676	(430)	759	(3615)	3313	(8717)	435	(1967)	1372	(16010)	3422	(17171)	10444	(24873)
2002	681	(398)	753	(1503)	2556	(5580)	459	(1884)	313	(11458)	2599	(15849)	7785	(22419)
2003	697	(385)	864	(1495)	1715	(3023)	690	(2930)	2794	(11920)	1653	(17450)	8958	(23473)
2004	697	(371)	992	(1704)	1773	(3292)	812	(2372)	415	(11236)	1652	(15454)	6806	(21841)
2005	703	(444)	1056	(1849)	1689	(4014)	923	(2344)	-501	(11830)	1082	(14753)	5613	(21800)

B. Percentage of Total Compensation (in %)

Year	Salary	Bonus	Value of options Granted	Values of Restricted Stocks		Change in Wealth from Options Held	Change in Wealth from Stock Held
				Restricted Stocks	Restricted Stocks		
1993	17.5	13.5	21.3	3.9	12.1	23.7	
1994	19.0	16.1	31.3	4.7	23.1	-1.1	
1995	23.9	20.2	36.8	6.3	11.3	-10.3	
1996	15.5	14.7	33.6	5.3	17.0	5.6	
1997	11.3	12.5	33.1	5.4	21.6	8.9	
1998	32.9	33.1	104.8	14.8	-35.7	-70.2	
1999	35.6	40.0	131.9	20.6	22.2	-175.8	
2000	6.1	7.6	30.6	4.2	14.8	31.3	
2001	6.5	7.3	31.7	4.2	13.1	32.8	
2002	8.7	9.7	32.8	5.9	4.0	33.4	
2003	7.8	9.6	19.1	7.7	31.2	18.5	
2004	10.2	14.6	26.1	11.9	6.1	24.3	
2005	12.5	18.8	30.1	16.4	-8.9	19.3	

Table 4: Cross-section Summary of Compensation  
(In thousands of US\$ (2006), standard deviations in parentheses)

State	Primary			Consumer			Service					
	Bad	Good	Bad	Good	Bad	Good	Bad	Good				
Z \ Period	Pre (1)	Post (2)	Pre (3)	Post (4)	Pre (5)	Post (6)	Pre (7)	Post (8)	Pre (9)	Post (10)	Pre (11)	Post (12)
1 ( $A_S C_S$ )	657 (7750)	1389 (11800)	4474 (16904)	6904 (13700)	-1272 (15504)	-1133 (10585)	6965 (28180)	9646 (24230)	1095 (15041)	1262 (13361)	7860 (28052)	7560 (22274)
2 ( $A_S C_L$ )	248 (7062)	2426 (5474)	3632 (7740)	6940 (10115)	580 (9021)	844 (6365)	7164 (18797)	7277 (17749)	1445 (12922)	1407 (11899)	7574 (21780)	8238 (21079)
3 ( $A_L C_S$ )	2855 (9936)	8039 (18117)	6458 (15787)	12888 (24020)	894 (23933)	873 (30287)	12083 (39369)	12227 (42431)	6370 (30391)	7646 (31371)	15174 (41707)	17083 (38319)
4 ( $A_L C_L$ )	3216 (10563)	6465 (12648)	6359 (13021)	10687 (18947)	3493 (21668)	6954 (24369)	12680 (29889)	14694 (29214)	6727 (23607)	6816 (20956)	14465 (29805)	16689 (33114)

Note: A: Asset. C: Capital structure (debt-to-equity ratio). S: Small. L: Large.

Table 5: Cross-section Summary of Abnormal Returns  
(Standard deviations in parentheses.)

Z	Primary		Consumer				Service					
	Bad	Good	Bad	Good	Bad	Good	Bad	Good	Bad	Good		
	Pre (1)	Post (2)	Pre (3)	Post (4)	Pre (5)	Post (6)	Pre (7)	Post (8)	Pre (9)	Post (10)	Pre (11)	Post (12)
1 ( $A_S C_S$ )	-0.094 (0.337)	-0.015 (0.281)	0.037 (0.346)	0.140 (0.282)	-0.141 (0.377)	-0.115 (0.298)	0.123 (0.417)	0.161 (0.328)	-0.115 (0.469)	-0.051 (0.419)	0.143 (0.499)	0.168 (0.399)
2 ( $A_S C_L$ )	-0.097 (0.329)	0.006 (0.314)	0.103 (0.352)	0.173 (0.342)	-0.139 (0.405)	-0.078 (0.434)	0.146 (0.430)	0.163 (0.347)	-0.071 (0.450)	-0.028 (0.405)	0.174 (0.456)	0.180 (0.383)
3 ( $A_L C_S$ )	-0.108 (0.286)	0.053 (0.277)	0.002 (0.275)	0.131 (0.238)	-0.077 (0.328)	-0.137 (0.208)	0.047 (0.275)	0.021 (0.221)	-0.164 (0.369)	-0.012 (0.334)	0.111 (0.419)	0.099 (0.319)
4 ( $A_L C_L$ )	-0.083 (0.278)	0.013 (0.218)	0.033 (0.285)	0.087 (0.257)	-0.132 (0.332)	-0.029 (0.285)	0.068 (0.343)	0.063 (0.278)	0.008 (0.345)	-0.011 (0.234)	0.111 (0.341)	0.077 (0.265)

Note: A: Asset. C: Capital structure (debt-to-equity ratio). S: Small. L: Large.

Table 6: Distribution of states in Pre-SOX and Post-SOX period

Pre-SOX							
Z	Primary		Consumer		Service		
	Bad/Good	Total (N)	Bad/Good	Total (N)	Bad/Good	Total (N)	
1 ( $A_S C_S$ )	1.2	2162	1.3	1704	1.3	2933	
2 ( $A_S C_L$ )	1.4	941	1.5	796	1.4	875	
3 ( $A_L C_S$ )	1.0	943	1.1	798	1.0	877	
4 ( $A_L C_L$ )	1.2	2167	1.3	1714	1.4	2943	
Total	1.2	6213	1.3	5012	1.3	7628	

Post-SOX							
Z	Primary		Consumer		Service		
	Bad/Good	Total (N)	Bad/Good	Total (N)	Bad/Good	Total (N)	
1 ( $A_S C_S$ )	1.1	2162	1.1	1704	1.0	2933	
2 ( $A_S C_L$ )	1.5	941	0.9	796	1.4	875	
3 ( $A_L C_S$ )	1.1	943	0.8	798	1.2	877	
4 ( $A_L C_L$ )	1.3	2167	1.0	1714	1.6	2943	
Total	1.2	6213	1.0	5012	1.3	7628	

Note: A: Asset. C: Capital structure (debt-to-equity ratio). S: Small. L: Large.

Table 7.1 Test on PDF of Gross Abnormal Returns ( $T^{PDF}$ )

Z	A. Primary		B. consumer		C. Service	
	Bad	Good	Bad	Good	Bad	Good
1 ( $A_S C_S$ )	14.07	14.91	10.69	12.07	13.75	18.40
2 ( $A_S C_L$ )	4.72	0.93	0.75	5.09	8.10	2.03
3 ( $A_L C_S$ )	10.85	11.65	10.61	3.23	12.96	10.79
4 ( $A_L C_L$ )	26.53	8.46	9.58	10.93	64.76	20.13

Table 7.2 Test on Contract Shape ( $T^W$ )

Z	A. Primary		B. consumer		C. Service	
	Bad	Good	Bad	Good	Bad	Good
1 ( $A_S C_S$ )	33.40	8.40	12.52	15.08	9.79	14.55
2 ( $A_S C_L$ )	43.47	34.46	14.67	9.56	5.79	4.11
3 ( $A_L C_S$ )	101.32	57.69	7.77	2.30	16.28	11.21
4 ( $A_L C_L$ )	84.11	147.08	25.74	54.59	35.57	56.26

Note: A: Asset. C: Capital structure (debt-to-equity ratio). S: Small. L: Large. Both tests are one-sided test and both statistics follow  $N(0,1)$ .

## BIBLIOGRAPHY

- Abowd, JM, and DS Kaplan, 1999, Executive compensation: Six questions that need answering., *Journal of Economic Perspectives* 13, 145–167.
- ACHARYA, VIRAL V., STEWART C. MYERS, and RAGHURAM G. RAJAN, 2011, The internal governance of firms, *The Journal of Finance* 66, 689–720.
- Adams, René B., and Daniel Ferreira, 2008, Do directors perform for pay?, *Journal of Accounting and Economics* 46, 154 – 171.
- Aggarwal, Rajesh K., Mark E. Evans, and Dhananjay Nanda, 2012, Nonprofit boards: Size, performance and managerial incentives, *Journal of Accounting and Economics* 53, 466 – 487.
- Albuquerque, Ana, 2009, Peer firms in relative performance evaluation, *Journal of Accounting and Economics* 48, 69 – 89.
- Antle, Rick, and Abbie Smith, 1985, Measuring executive compensation: Methods and an application, *Journal of Accounting Research* 23, 296–325.
- , 1986, An empirical investigation of the relative performance evaluation of corporate executives, *Journal of Accounting Research* 24, 1–39.
- Armstrong, Chris, Alan Jagolinzer, and David Larcker, 2010, Performance-based incentives for internal monitors, *Rock Center for Corporate Governance at Stanford University Working Paper Series*.
- Armstrong, Christopher S., Jennifer L. Blouin, and David F. Larcker, 2012, The incentives for tax planning, *Journal of Accounting and Economics* 53, 391 – 411.

- Armstrong, Christopher S., Wayne R. Guay, and Joseph P. Weber, 2010, The role of information and financial reporting in corporate governance and debt contracting, *Journal of Accounting and Economics* 50, 179 – 234.
- Arya, Anil, John Fellingham, and Jonathan Glover, 1997, Teams, repeated tasks, and implicit incentives, *Journal of Accounting and Economics* 23, 7–30.
- Balsam, Steven, and Setiyono Miharjo, 2007, The effect of equity compensation on voluntary executive turnover, *Journal of Accounting and Economics* 43, 95 – 119.
- Banker, Rajiv D, Masako N Darrrough, Rong Huang, and Jose M Plehn-Dujowich, 2013, The relation between ceo compensation and past performance, *The Accounting Review* 88, 1–30.
- BANKER, RAJIV D., RONG HUANG, and RAMACHANDRAN NATARAJAN, 2009, Incentive contracting and value relevance of earnings and cash flows, *Journal of Accounting Research* 47, 647–678.
- Berle, Adolf Augustus, and Gardiner Coit Means, 1932, *The modern corporation and private property* new york, macmillan reprinted in 1944 edn.
- Bertrand, Marianne, 2009, Ceos, *Annual Review of Economics* 1, 121–150.
- Bolton, Patrick, and Mathias Dewatripont, 2005, *Contract Theory* (The MIT Press).
- Boschen, John F., Augustine Duru, Lawrence A. Gordon, and Kimberly J. Smith, 2003, Accounting and stock price performance in dynamic ceo compensation arrangements, *The Accounting Review* 78, 143–168.
- Bushman, Robert, Qi Chen, Ellen Engel, and Abbie Smith, 2004, Financial accounting information, organizational complexity and corporate governance systems, *Journal of Accounting and Economics* 37, 167–201.
- Bushman, Robert, Zhonglan Dai, and Weining Zhang, 2012, Management team incentive alignment and firm value, *working paper*.
- Bushman, Robert, Ellen Engel, and Abbie Smith, 2006, An analysis of the relation between the stewardship and valuation roles of earnings, *Journal of Accounting Research* 44, 53–83.
- Bushman, Robert M., and Abbie J. Smith, 2001, Financial accounting information and corporate governance, *Journal of Accounting and Economics* 32, 237–333.

- Cadman, Brian, Mary Ellen Carter, and Stephen Hillegeist, 2010, The incentives of compensation consultants and ceo pay, *Journal of Accounting and Economics* 49, 263 – 280.
- Carter, Mary Ellen, Luann J. Lynch, and Sarah L. C. Zechman, 2009, Changes in bonus contracts in the post-sarbanes-oxley era, *Review of Accounting Studies* 14, 480–506.
- Chan, Lilian H., Kevin C.W. Chen, Tai-Yuan Chen, and Yangxin Yu, 2012, The effects of firm-initiated clawback provisions on earnings quality and auditor behavior, *Journal of Accounting and Economics* 54, 180 – 196.
- Che, Yeon-Koo, and Seung-Weon Yoo, 2001, Optimal incentives for teams, *American Economic Review* 91, 525–541.
- Cheng, Qiang, and David B. Farber, 2008, Earnings restatements, changes in ceo compensation, and firm performance, *The Accounting Review* 83, 1217–1250.
- Chenhall, Robert H, and Frank Moers, 2007, The issue of endogeneity within theory-based, quantitative management accounting research, *European Accounting Review* 16, 173–196.
- Chernozhukov, Victor, Han Hong, and Elie Tamer, 2007, Estimation and confidence regions for parameter sets in econometric models<sup>1</sup>, *Econometrica* 75, 1243–1284.
- Chetty, Raj, 2009, Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods, *Annual Review of Economics* 1, 451–488.
- Coates, John C., 2007, The goals and promise of the sarbanes-oxley act, *Journal of Economic Perspectives* 21, 91–116.
- Cohen, Daniel A., Aiysha Dey, and Thomas Z. Lys, 2007, The sarbanes oxley act of 2002: Implications for compensation contracts and managerial risk-taking, *working paper*.
- Comprix, Joseph, and Karl A. Muller, 2006, Asymmetric treatment of reported pension expense and income amounts in ceo cash compensation calculations, *Journal of Accounting and Economics* 42, 385 – 416.
- , 2011, Pension plan accounting estimates and the freezing of defined benefit pension plans, *Journal of Accounting and Economics* 51, 115 – 133.
- Core, John, and Wayne Guay, 2002, Estimating the value of employee stock option portfolios and their sensitivities to price and volatility, *Journal of Accounting Research* 40, 613–630.

- Core, John E., Wayne Guay, and David F. Larcker, 2003, Executive equity compensation and incentives: A survey, *Federal Reserve Bank of New York Economic Policy Review*.
- Core, John E., and Wayne R. Guay, 2001, Stock option plans for non-executive employees, *Journal of Financial Economics* 61, 253 – 287.
- Dechow, Patricia M., 2006, Asymmetric sensitivity of ceo cash compensation to stock returns: A discussion, *Journal of Accounting and Economics* 42, 193 – 202 <ce:title>Conference Issue on Implications of Changing Financial Reporting Standards</ce:title>.
- Dey, Aiysha, 2010, The chilling effect of sarbanes–oxley: A discussion of sarbanes–oxley and corporate risk-taking, *Journal of Accounting and Economics* 49, 53–57.
- Edmans, Alex, and Xavier Gabaix, 2009, Is ceo pay really inefficient? a survey of new optimal contracting theories, *European Financial Management* 15, 486–496.
- EncinosaIII, William E., Martin Gaynor, and James B. Rebitzer, 2007, The sociology of groups and the economics of incentives: Theory and evidence on compensation systems, *Journal of Economic Behavior and Organization* 62, 187 – 214.
- Engel, Ellen, Rachel M. Hayes, and Xue Wang, 2010, Audit committee compensation and the demand for monitoring of the financial reporting process, *Journal of Accounting and Economics* 49, 136 – 154 <ce:title>Conference Issue on Current Issues in Accounting & Reassessing the Regulation of Capital Markets</ce:title> <xocs:full-name>University of Rochester, William E. Simon Graduate School of Business Administration</xocs:full-name>.
- ERKENS, DAVID H., 2011, Do firms use time-vested stock-based pay to keep research and development investments secret?, *Journal of Accounting Research* 49, 861–894.
- Fama, Eugene F., 1980, Agency problems and the theory of the firm, *Journal of political economy* 88, 288–307.
- Ferri, Fabrizio, and Tatiana Sandino, 2009, The impact of shareholder activism on financial reporting and compensation: The case of employee stock options expensing, *The Accounting Review* 84, 433–466.
- Finkelstein, Sydney, Donald C Hambrick, and Albert A Cannella, 1996, *Strategic leadership* (West St. Paul, Minn.).

- Frydman, Carola, and Dirk Jenter, 2010, Ceo compensation, *Annual Review of Financial Economics* 2, 75–102.
- Fudenberg, Drew, Bengt Holmstrom, and Paul Milgrom, 1990, Short-term contracts and long-term agency relationships, *Journal of Economic Theory* 51, 1 – 31.
- Gayle, George-Levi, Chen Li, and Robert A. Miller, 2013, The consequences of 2002 governance rules on ceos’ compensation, *Tepper School of Business, Carnegie Mellon University, Working Paper*.
- Gayle, George-Levi, and Robert A. Miller, 2009, Has moral hazard become a more important factor in managerial compensation?, *American Economic Review* 99, 1740–1769.
- , 2012, Identifying and testing models of managerial compensation, *Tepper School of Business, Carnegie Mellon University, Working Paper*.
- Glover, Jonathan, 2012, Explicit and implicit incentives for multiple agents, *working paper*.
- Gong, Guojin, Laura Yue Li, and Jae Yong Shin, 2011, Relative performance evaluation and related peer groups in executive compensation contracts, *The Accounting Review* 86, 1007–1043.
- Grossman, Sanford J., and Oliver D. Hart, 1983, An analysis of the principal-agent problem, *Econometrica* 51, pp. 7–45.
- Hall, B., and J. Leibman, 1998, Are CEOs really paid like bureaucrats?, *The Quarterly Journal of Economics* 103, 653–691.
- Hambrick, Donald C, 2007, Upper echelons theory: An update., *Academy of management review* 32, 334–343.
- Hanlon, Michelle, Shivaram Rajgopal, and Terry Shevlin, 2003, Are executive stock options associated with future earnings?, *Journal of Accounting and Economics* 36, 3 – 43  
<ce:title>Conference Issue on</ce:title>.
- Heckman, James J, 2000, Causal parameters and policy analysis in economics: A twentieth century retrospective, *The Quarterly Journal of Economics* 115, 45–97.
- , and Edward Vytlacil, 2005, Structural equations, treatment effects, and econometric policy evaluation1, *Econometrica* 73, 669–738.

- Henderson, Andrew D., and James W. Fredrickson, 2001, Top management team coordination needs and the ceo pay gap: A competitive test of economic and behavioral views, *The Academy of Management Journal* 44, 96–117.
- Hermalin, Benjamin E., 1998, Toward an economic theory of leadership: Leading by example, *The American Economic Review* 88, pp. 1188–1206.
- Holmstrom, B., 1979, Moral hazard and observability, *The Bell Journal of Economics* 10, 74–91.
- Holmstrom, Bengt, 1982, Moral hazard in teams, *The Bell Journal of Economics* 13, 324–340.
- , and Steven N. Kaplan, 2003, The state of u.s. corporate governance, *European Corporate Governance Institute Finance Working Paper* September.
- Holmstrom, Bengt, and Paul Milgrom, 1990, Regulating trade among agents, *Journal of Institutional and Theoretical Economics* 146, 85–105.
- Hood, William Calvin, Tjalling C Koopmans, and Yale University, 1953, *Studies in econometric method* . vol. 14 (Wiley New York).
- Imbens, Guido W, and Jeffrey M Wooldridge, 2009, Recent developments in the econometrics of program evaluation, *Journal of Economic Literature* 47, 5–86.
- Indjejikian, Raffi J., and Dhananjay (DJ) Nanda, 2002, Executive target bonuses and what they imply about performance standards, *The Accounting Review* 77, 793–819.
- Iskandar-Datta, Mai, and Yonghong Jia, 2013, Valuation consequences of clawback provisions, *The Accounting Review* 88, 171–198.
- Itoh, Hideshi, 1990, Coalitions, incentives, and risk sharing, *Unpublished manuscript*.
- , 1992, Cooperation in hierarchical organizations: An incentive perspective, *Journal of Law, Economics and Organization* 8, 321–345.
- , 1993, Coalitions, incentives, and risk sharing, *Journal of Economic Theory* 60, 410–427.

- Ittner, Christopher, and David Larcker, 2002, Empirical managerial accounting research: are we just describing management consulting practice?, *European Accounting Review* 11, 787–794.
- Ittner, Christopher D., Richard A. Lambert, and David F. Larcker, 2003, The structure and performance consequences of equity grants to employees of new economy firms, *Journal of Accounting and Economics* 34, 89 – 127.
- Jayaraman, Sudarshan, and Todd T. Milbourn, 2012, The role of stock liquidity in executive compensation, *The Accounting Review* 87, 537–563.
- Kandel, Eugene, and Edward P. Lazear, 1992, Peer pressure and partnerships, *Journal of Political Economy* 100, pp. 801–817.
- Karuna, Christo, 2007, Industry product market competition and managerial incentives, *Journal of Accounting and Economics* 43, 275 – 297.
- KNECHEL, W. ROBERT, LASSE NIEMI, and MIKKO ZERNI, 2013, Empirical evidence on the implicit determinants of compensation in big 4 audit partnerships, *Journal of Accounting Research* 51, 349–387.
- Knez, Marc, and Duncan Simester, 2001, Firm-wide incentives and mutual monitoring at continental airlines, *Journal of Labor Economics* 19, 743–772.
- Kreps, D.M., 1990, Corporate culture and economic theory, *Perspectives on positive political economy* 90, 109–10.
- Lambert, Richard A., 2001, Contracting theory and accounting, *Journal of Accounting and Economics* 32, 3 – 87.
- Lambert, Richard A, 2006, Agency theory and management accounting, vol. 1 . pp. 247–268 (Elsevier).
- Landier, Augustin, Julien Sauvagnat, David Sraer, and David Thesmar, 2012, Bottom-up corporate governance, *Review of Finance*.
- Larcker, David F, 2003, Discussion of “Are executive stock options associated with future earnings?”, *Journal of Accounting and Economics* 36, 91–103.
- , and Tjomme O Rusticus, 2007, Endogeneity and empirical accounting research, *European Accounting Review* 16, 207–215.

- Larcker, David F., and Tjomme O. Rusticus, 2010, On the use of instrumental variables in accounting research, *Journal of Accounting and Economics* 49, 186 – 205.
- Leamer, Edward E, 1983, Let’s take the con out of econometrics, *The American Economic Review* 73, 31–43.
- Leone, Andrew J., Joanna Shuang Wu, and Jerold L. Zimmerman, 2006, Asymmetric sensitivity of ceo cash compensation to stock returns, *Journal of Accounting and Economics* 42, 167 – 192 <ce:title>Conference Issue on Implications of Changing Financial Reporting Standards</ce:title>.
- Leuz, Christian, 2007, Was the Sarbanes-oxley act of 2002 really this costly? a discussion of evidence from event returns and going-private decisions, *Journal of Accounting and Economics* 44, 146–165.
- Li, Qi, and Jeffrey Scott Rachine, 2006, *Nonparametric Econometrics: Theory and Practice* (Princeton University Press).
- Li, Zhichuan, 2011, Mutual monitoring and corporate governance, *Arizona State University, working paper*.
- Ma, Ching-To, 1988, Unique implementation of incentive contracts with many agents, *The Review of Economic Studies* 55, 555–572.
- Macho-Stadler, Ines, and J. David Perez-Castrillo, 1993, Moral hazard with several agents, *International Journal of Industrial Organization* 11, 73–100.
- MacLeod, W.B., 1995, Incentives in organizations: An overview of some of the evidence and theory, *Trends in Business Organization: Do Participation and Cooperation Increase Competitiveness?* pp. 832–854.
- MacLeod, W. Bentley, and James M. Malcomson, 1989, Implicit contracts, incentive compatibility, and involuntary unemployment, *Econometrica* 57, pp. 447–480.
- Main, Brian G. M., Charles A. O’Reilly III, and James Wade, 1993, Top executive pay: Tournament or teamwork?, *Journal of Labor Economics* 11, 606–628.
- Manski, Charles F, 2003, *Partial identification of probability distributions* (Springer).
- Margiotta, Mary M., and Robert A. Miller, 2000, Managerial compensation and the cost of moral hazard, *International Economic Review* 41, 669–719.

- Masulis, Ronald W., Cong Wang, and Fei Xie, 2012, Globalizing the boardroom—the effects of foreign directors on corporate governance and firm performance, *Journal of Accounting and Economics* 53, 527 – 554.
- Matsunaga, Steven R., and Chul W. Park, 2001, The effect of missing a quarterly earnings benchmark on the ceo’s annual bonus, *The Accounting Review* 76, 313–332.
- Matzkin, Rosa L., 2007, Chapter 73 nonparametric identification, vol. 6, Part B of *Handbook of Econometrics* . pp. 5307 – 5368 (Elsevier).
- McAnally, Mary Lea, Anup Srivastava, and Connie D. Weaver, 2008, Executive stock options, missed earnings targets, and earnings management, *The Accounting Review* 83, 185–216.
- Mirrlees, J. A., 1975, The theory of moral hazard and unobservable behaviour, *mimeo, Oxford*.
- Murphy, K., 2012, Executive compensation: Where we are, and how we got there, *Handbook of the Economics of Finance. Elsevier Science North Holland (Forthcoming)*.
- Murphy, Kevin J., 1999a, Chapter 38 executive compensation, vol. 3, Part B of *Handbook of Labor Economics* . pp. 2485 – 2563 (Elsevier).
- , 1999b, Executive compensation, *Working paper*.
- , and Tatiana Sandino, 2010, Executive pay and independent compensation consultants, *Journal of Accounting and Economics* 49, 247 – 262.
- Nagar, Venky, Dhananjay Nanda, and Peter Wysocki, 2003, Discretionary disclosure and stock-based incentives, *Journal of Accounting and Economics* 34, 283 – 309.
- Nekipelov, Denis, 2007, Empirical content of a continuous-time principal-agent model: The case of the retail apparel industry, *Working paper*.
- Nevo, Aviv, and Michael D Whinston, 2010, Taking the dogma out of econometrics: Structural modeling and credible inference, *The Journal of Economic Perspectives* 24, 69–81.
- Ortiz-Molina, Hernan, 2007, Executive compensation and capital structure: The effects of convertible debt and straight debt on ceo pay, *Journal of Accounting and Economics* 43, 69 – 93.

- OZKAN, NESLIHAN, ZVI SINGER, and HAIFENG YOU, 2012, Mandatory ifrs adoption and the contractual usefulness of accounting information in executive compensation, *Journal of Accounting Research* 50, 1077–1107.
- Pagan, Adrian, and Aman Ullah, 1999, *Nonparametric Econometrics* (Cambridge University Press).
- Powell, James L, 1994, Estimation of semiparametric models, vol. 4 . pp. 2443–2521 (Elsevier).
- Prendergast, Canice, 1999, The provision of incentives in firms, *Journal of Economic Literature* 37, 7–63.
- Rajgopal, Shivaram, and Terry Shevlin, 2002, Empirical evidence on the relation between stock option compensation and risk taking, *Journal of Accounting and Economics* 33, 145 – 171.
- Ramakrishnan, Ram T. S., and Anjan V. Thakor, 1991, Cooperation versus competition in agency, *Journal of Law, Economics and Organization* 7, 248–286.
- Reiss, Peter C, and Frank A Wolak, 2007, Structural econometric modeling: Rationales and examples from industrial organization, vol. 6 . pp. 4277–4415 (Elsevier).
- Rosen, S., 1992, Contracts and the market for executives, in L. Werin, and H. Wijkander, ed.: *Contract Economics* pp. 181–211.
- Roulstone, Darren T., 2003, The relation between insider-trading restrictions and executive compensation, *Journal of Accounting Research* 41, 525–551.
- Salanie, Bernard, 2003, Testing contact theory, *CESifo Economic Studies* 49, 461–477.
- Schroeder, Douglas A., 2010, *Accounting and causal effects: econometric challenges* (Springer: New York).
- SEC, 2002, Sarbanes-oxley act of 2002, .
- Skantz, Terrance R., 2012, Ceo pay, managerial power, and sfas 123(r), *The Accounting Review* 87, 2151–2179.
- Tamer, Elie, 2010, Partial identification in econometrics, *Annual Review of Economics* 2, 167–195.

- Tirole, Jean, 1992, *Collusion and the Theory of Organizations* vol. II . chap. 3, pp. 151–206 (Advances in Economic Theory Sixth World Congress).
- Van Lent, Laurence, 2007, Endogeneity in management accounting research: A comment, *European Accounting Review* 16, 197–205.
- Varian, Hal. R., 1990, Monitoring agents with other agents, *Journal of Institutional and Theoretical Economics* 146, 153–174.
- White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica: Journal of the Econometric Society* pp. 817–838.
- Yermack, David, 2006, Golden handshakes: Separation pay for retired and dismissed ceos, *Journal of Accounting and Economics* 41, 237 – 256.
- Young, Steven, and Jing Yang, 2011, Stock repurchases and executive compensation contract design: The role of earnings per share performance conditions, *The Accounting Review* 86, 703–733.
- Zhang, Ivy Xiyang, 2007, Economic consequences of the sarbanes-oxley act of 2002, *Journal of Accounting and Economics* 44, 74 – 115 <ce:title>Conference Issue on Corporate Governance: Financial Reporting, Internal Control, and Auditing</ce:title>.

?