

How Well Does Agency Theory Explain Executive Compensation?

George-Levi Gayle, Chen Li, and Robert A. Miller

As the share of all income going to the top 1 percent has risen over the past four decades, so has the share of top incomes coming from labor income relative to capital income. The rise in labor income is mainly due to the explosion in executive compensation over the same period—mostly because of the increase in executives being paid with stocks, options, and bonuses. The principal-agent model explains the reason for such compensation instead of a flat salary. Yet hundreds of papers in economics, finance, accounting, and management have reached no consensus on whether executive compensation is efficient or whether empirically it conforms to the prediction of the principal-agent theory. In this article, we argue that this lack of consensus is due to two issues: The first is a measurement issue, and the second is that the exact prediction of the principal-agent model depends on many objects unobservable to the econometrician. We illustrate how using theory-based estimation together with a model-motivated measure of total compensation can help overcome these issues. Finally, using a model-consistent measure of compensation and theory-based estimation, we conclude that executive compensation broadly conforms to the principal-agent theory; however, each situation and the variables used have to be carefully modeled, identified, and estimated. (JEL D82, L25, M12, M52)

Federal Reserve Bank of St. Louis *Review*, Third Quarter 2018, 100(3), pp. 201-36.
<https://doi.org/10.20955/r.2018.201-36>

1 INTRODUCTION

Over the past three to four decades, income inequality in the United States has substantially increased. One measure of inequality used in academic research and the popular press is the share of all income that goes to the top income earners (especially the top 1 percent). In the United States, top income shares dropped dramatically from 1929 to 1950 but have increased dramatically since 1980.¹ In the early part of the twentieth century, top incomes were made up of primarily capital income; however, today's top incomes are divided 50/50 between labor and capital income. The increase from labor income is primarily from the

George-Levi Gayle is an associate professor at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis. Chen Li is an assistant professor at the Zicklin School of Business, Baruch College, CUNY. Robert A. Miller is the Richard M. Cyert and Morris DeGroot Professor of Economics and Statistics at the Tepper School of Business, Carnegie Mellon University.

© 2018, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

explosion of executive compensation since 1980, paid mainly with firm-denominated securities, that is, stocks, options, and bonuses.²

Why are executives paid in firm-denominated securities? The principal-agent model is the main theoretical underpinning for why managers are compensated with stocks, options, and bonuses instead of a flat salary. The model captures the economic interactions of an uninformed party (the principal) who delegates tasks to the informed party (the agent) whose private action can affect both parties' benefits and whose interest is not perfectly aligned with the uninformed party. Modern firms are characterized by a dispersed ownership structure; the shareholders of a firm delegate the business operation to professional managers. Unlike the input of physical capital that can be easily measured, the input of managerial effort is hardly measurable and cannot be directly traded. A principal-agent problem, called moral hazard, arises when self-interested managers intend to secretly choose an effort level different from what would maximize the benefits of shareholders. To align interests, shareholders have to base executive compensation on the output of managerial effort, for example, the stock price. The unobservability of managerial effort is the main reason why executive compensation is largely based on firm equity rather than a flat salary.

Principal-agent models use techniques that characterize the optimal incentive mechanism for aligning the principal's and the agent's interests by simplifying assumptions about their preferences, technologies, and information structures. However, some uncertainties in the economy may also affect output, which risk-averse managers want to be insured against. Shareholders have to pay an extra amount as a risk premium to managers while balancing between incentives and insurance. An efficient compensation contract provides insurance at a sufficient amount that can guarantee the manager makes the effort that shareholders desire. In addition to the information asymmetry on effort, a manager may take advantage of his private information regarding the firm's state, which shareholders do not have access to. The optimal contract also has to provide the incentive for the manager to truthfully reveal the private information on the firm's state at an extra cost to the firm.

In contrast to the sophisticated, complex compensation schemes in the real world, there is a question of whether abstract principal-agent models can provide good explanations and predictions on executive compensation. Empirical research on managerial compensation seeks to examine whether the observed compensation schemes conform to an optimal contract supported by the principal-agent model. Furthermore, this line of research identifies and quantifies the effects of asymmetric information and assesses its impact on welfare, competition, and policy. Ultimately, if these models are a good description of the complex real-world compensation practices, they can be used to understand one aspect of the increase in inequality over the past four decades.

Empirically, there are three main ways of evaluating the output of a model. The first is to test a major prediction of the model while leaving unspecified the main structure of the model, for example, the positive-correlation test of agency theory. For examples of this approach for insurance markets, see Chiappori and Salanie (2000), Cardon and Hendel, (2001), and Cohen (2005), among others.³ For examples for performance-pay settings, see Jensen and Murphy (1990), Hall and Liebman (1998), and Aggarwal and Samwick (1999), among others.⁴ The

second way of evaluating the output of a model is to specify the complete structure of the model and then derive over-identifying restrictions. A model that has empirical content imposes restrictions on the relationship between variables observed in the data. These restrictions can be used to recover the parameters of the model. If the number of independent restrictions that the model imposes on the observables is more than the number of parameters that need to be estimated, then the additional restrictions are called over-identifying restrictions. These over-identifying restrictions can be used to test the validity of the model. Without such restrictions, the model could be rationalized by any data. The final way is to specify the complete structure and perform out-of-sample validation. For example, say a researcher uses data from before 1980 to estimate a principal-agent model of executive compensation and then uses the estimated model to predict compensation after 1980. A test of the model would be to see if the model can predict the post-1980 rapid increase in executive compensation. The last two ways are the focus of theoretical-based estimation. Theoretical-based estimation is normally called for by the need to move beyond testing a model and to quantify welfare, efficiency, and the potential impact of policy reforms.

There are hundreds of papers in economics, finance, accounting, and management on whether executive compensation is efficient⁵ and whether empirically it conforms to the prediction of the principal-agent model. Most of this research is based on testing one of the major predictions of the principal-agent model while leaving unspecified the main structure of the model. As of yet there is little conclusive evidence from this approach as to whether executive compensation packages are correctly structured or conform to the principal-agent model. In this article we argue that this is due to two issues: The first is a measurement issue, and the second is that the exact prediction of the principal-agent model depends on many objects unobservable to the econometrician. We illustrate how using theory-based estimation together with a model-motivated measure of total compensation can help overcome these issues. We conclude that using a model-consistent measure of compensation and theory-based estimation shows that executive compensation broadly conforms to the principal-agent theory; however, each situation and the variables used have to be carefully modeled, identified, and estimated.

The rest of the article is organized as follows. Section 2 takes up the issue of the measurement of total compensation, which is consistent with the principal-agent model and shows how this measurement differs significantly from what is used in most of the literature. It then applies this measurement concept to two distinct datasets covering over 60 years and documents how each component has changed. Section 3 takes up the issue of the unobservability of important aspects of the basic moral hazard model (called a pure moral hazard model from now on) and shows how theory-based estimation can be used to obtain measures of these important concepts. The pure moral hazard model is then estimated and used to answer this question: Why has executive compensation risen 10 times as fast as the pay of the average worker over the past 60 years? Section 4 illustrates that some aspects of executive compensation that seem to contradict the basic model can be easily reconciled with the more-general theory and hence the issue of marginal versus joint distribution of variables needs to be considered more carefully when choosing a model. Section 5 concludes and gives some direction for future research.

2 HOW IS EXECUTIVE COMPENSATION MEASURED?

The cost to shareholders of employing a manager, called direct compensation, is the sum of salary and bonuses, the value of restricted stocks and options granted, and the value of retirement and long-term compensation schemes. The discounted sum of these direct-compensation items measures the reduction in the firm's value from outlays to management. Total compensation to a manager is defined as direct compensation plus changes in wealth from holding firm options and changes in wealth from holding firm stock. To compute the remaining two components in total compensation, one must address where managers would place this wealth if it were not held in their firms' financial securities. We assume that managers would hold a well-diversified portfolio instead, an implication of our model. When forming their portfolio of real and financial assets, managers recognize that part of the return from their firm-denominated securities should be attributed to aggregate factors, so they reduce their holdings of other stocks to neutralize those factors. Hence, the change in wealth from holding their firms' stock is the value of the stock at the beginning of the period multiplied by the abnormal return.

The principal-agent model implies that changes in wealth from holding firm options and changes in wealth from holding firm stock both have mean zero. An efficient contractual arrangement would not induce a risk-averse agent to hold more risk than is absolutely necessary, because any additional risk held by the risk-averse agent would have to be compensated for by a risk premium. Therefore, all risk beyond the agent's control should be netted out of the compensation. This can be done by allowing the agent to hold a well-diversified market portfolio, which is equivalent to netting out the market portfolio and any predictable component of the firm's securities. Therefore, both from the manager's and firm's perspectives, the netting out of the return on the market portfolio is desirable and, hence, the expected values of the change in wealth from holding firm options and change in wealth from holding firm stock are zero. This implies that direct and total compensation have the same expected value. Therefore, whether risk-neutral shareholders minimize expected total compensation or expected direct compensation is moot. However, changes in wealth from holding firm stock and options reflect the costs a manager incurs from not being able to fully diversify his wealth portfolio because of restrictions on stock and option sales. Consequently, managers care about total compensation, not direct compensation, because the former determines how their wealth changes from period to period when they optimally smooth their consumption over the life cycle and make optimal portfolio choices.

A third measure of compensation, called constrained compensation, is the sum of cash, bonuses, and the value of restricted stock and option grants, plus the change in the value of restricted stock and grant holdings. Constrained compensation exposes the manager to aggregate risk to the degree that the firm's share price fluctuates with the market. Rational managers would neutralize their market risk by reducing their holdings of the market portfolio to compensate for the additional market risk that holding restricted stock entails. Suppose managers held no diversified stock after receiving their compensation and were prevented from selling futures in the market portfolio (maturing when their firm-specific securities can be redeemed

through sales). Then we might conclude compensation is based on market returns if cash and bonus payments were not sufficiently countercyclical to offset the manager's aggregate risk of holding a portfolio of his firm's financial securities. We are unaware of any evidence showing that the wealth portfolio of a manager is constrained by his own shareholders to hold more market risk than he voluntarily chooses. This explains why the measure of compensation most consistent with the principal-agent model is total compensation rather than constrained compensation.

2.1 The Income-Equivalent Measure of Total Compensation

This section presents techniques for estimating a current income-equivalent measure of total compensation, which follow Antle and Smith (1985, 1986), Hall and Liebman (1998), and Margiotta and Miller (2000). The current income equivalent is defined to be the amount of before-tax dollars that an executive would require to offset exactly the value of the compensation package received in a given year. The term "compensation package" refers to the before-tax value of salaries, short-term bonuses, deferred-to-retirement bonuses, stockholdings, stock bonuses, stock options, dividend units, phantom shares, pension benefits, savings-plan contributions, long-term performance plans, and any other special items (such as a loan to the executive made at a below-market rate). In the absence of evidence to the contrary, the following assumptions underlie the estimation procedures: (i) executives remain with their firms until retirement at age 80; (ii) all non-contingent, deferred compensation is sure to be received; (iii) salary levels are not expected to fall; and (iv) an executive does not possess inside information regarding future stock prices or the probability that he or she will die in any given year.

2.1.1 Data Construction Details. In this article, we use data from two sources. The first dataset covers the years 1944-78, and the details of how it is constructed can be found in Antle and Smith (1985, 1986) and Margiotta and Miller (2000). The second dataset covers the years 1993-2009. Below we provide some essential details on its construction.

Firm type is defined as a combination of the industrial sector and firm characteristics for each firm in each year. The data used to measure firm characteristics are from Compustat. First, we classify the whole sample into three industrial sectors according to the Global Industry Classification Standard (GICS) code. The primary sector includes firms in the energy (GICS code 1010), materials (GICS code 1510), industrials (GICS codes 2010, 2020, 2030), and utilities (GICS code 5510) sectors. The consumer goods sector includes firms in the consumer discretionary (GICS codes 2510, 2520, 2530, 2540, 2550) and consumer staples (GICS codes 3010, 3020, 3030) sectors. The services sector includes firms in the health care (GICS codes 3510, 3520), financial (GICS codes 4010, 4020, 4030, 4040), and information technology and telecommunication services (GICS codes 4510, 4520, 5010) sectors.

We use raw stock prices and adjustment factors from the Compustat PDE dataset. For each firm in the sample, we calculate monthly compounded returns adjusted for splitting and repurchasing for each fiscal year; we then subtract the return to a value-weighted market portfolio (NYSE/NASDAQ/AMEX) from this raw return to determine the net excess return for the firm's corresponding fiscal year. We drop firm-year observations if the firm changed

its fiscal year end such that all compensation and stock returns are based on 12 months and consequently comparable with each other. The excess return is obtained by adding the total compensation in the fiscal year (scaled by the firm's value at the beginning of the fiscal year) to the net excess returns in the same firm year.

2.1.1.1 Compensation. In addition to the total compensation included in Compustat ExecuComp data, we also calculate the holding value of firm-specific equities. Due to data limitations, we cannot observe for each sample year all the inputs of the Black-Scholes formula for grants carried from before 1993, the beginning year of our sample. Compustat ExecuComp provides the valuation information only for those options newly granted after 1993, including the number of underlying stock shares, exercised prices, expiration dates, and issue dates. However, we need to know these Black-Scholes inputs for options granted before 1993 to completely value the wealth change of CEOs by estimating the value of unexercised options and updating them each year. To facilitate the calculation, we assume that (i) all options are exercised on their expiration dates, (ii) stock options granted before 1993 are exercised in a first-in first-out fashion, and (iii) each CEO holds his own stock options granted before 1993 for a period of the average length of the holding period across all years when he is in the sample. Consequently, we can back out the issue dates and exercised prices for options granted before 1993 for each CEO. The same routine applies to nonzero options granted before the CEO entered our sample. We then apply the dividend-adjusted Black-Scholes formula to reevaluate the call options for each CEO in each year. The dividend-adjusted Black-Scholes formula used is as follows: Let c denote the call option value, K the exercise price, T_m the time to maturity (in years), S the underlying security price, q the dividend yield, r the risk-free rate, and σ the implied volatility. Let $N(\cdot)$ denote the standard normal cumulative distribution function. Then the call option value is given by

$$(1) \quad c = Se^{-qT_m}N(d_1) - Ke^{-rT_m}N(d_2),$$

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

and

$$(3) \quad d_2 = d_1 - \sigma\sqrt{T_m}.$$

Following the concept of income-equivalent total compensation defined above, we construct the total compensation by adding the change in wealth from options held and stock held to the other components of compensation included in ExecuComp.

2.1.2 Documentation of the Changes in Components of Three Different Samples.

Table 1 summarizes and compares the distribution of the five main compensation components among three samples. The components include salary and bonuses, the value of options granted, the value of restricted stock granted, the change in wealth from options held, and the change in wealth from stock held. The remaining unlisted components include retirement and long-term compensation. The "Old" sample covers the years 1944-78. The other two

Table 1**Cross-Sectional Information on Components of Compensation in Thousands of US\$ (2000)**

Variable	Rank	Compensation		
		Old	New Restricted	New All
Salary and bonuses	All	151 (68)	672 (576)	707 (1,036)
	CEO	151 (67)	1,199 (833)	1,176 (1,674)
	Non-CEO	146 (75)	530 (373)	584 (739)
Value of options granted	All	29 (104)	2,170 (7,184)	2886 (12,198)
	CEO	29 (105)	5,015 (12,432)	5,967 (18,263)
	Non-CEO	29 (93)	1,402 (4,593)	2,079 (9,861)
Value of restricted stock granted	All	0.0078 (0.0679)	242 (720)	306 (1,622)
	CEO	0.0085 (0.0708)	551 (1,310)	637 (2,097)
	Non-CEO	0.0001 (0.0006)	159 (404)	219 (1,460)
Change in wealth from options held	All	10 (284)	141 (6,131)	-235 (13,040)
	CEO	12 (286)	414 (10,503)	-479 (21,028)
	Non-CEO	-17 (257)	68 (4,239)	-171 (9,937)
Change in wealth from stock held	All	12 (896)	211 (12,144)	21 (20,170)
	CEO	0.7 (826)	632 (21,741)	109 (34,720)
	Non-CEO	142 (1,484)	98 (7,733)	-3 (14,055)

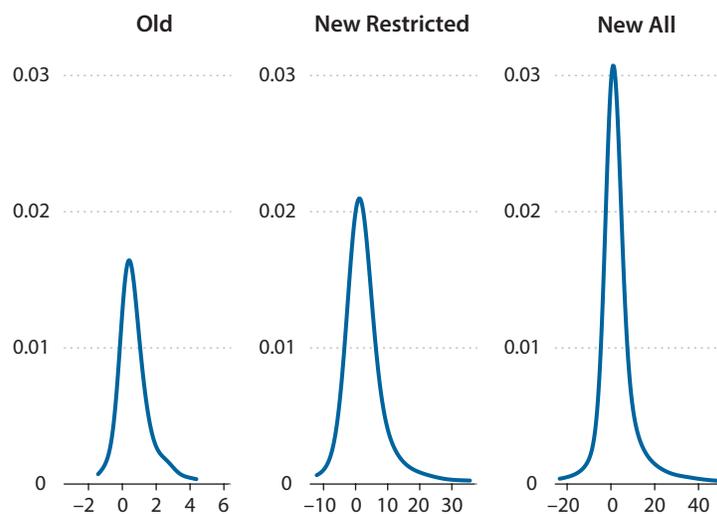
NOTE: Standard deviations are in parentheses.

samples cover the years 1993-2009. The Old sample and the “New Restricted” sample include the same industries, that is, aerospace, chemicals, and electronics. The “New All” sample includes all industries in the dataset from merged ExecComp, COMPUSTAT Fundamentals Annual, and CRSP⁶ monthly data.

Two persistent patterns arise from the comparison. Regardless of the time period and executive rank, the level of total compensation mainly depends on the first three components explicitly specified in compensation contracts, including salary and bonuses, the value of options granted, and the value of restricted stock granted. The wealth change, either from

Figure 1

Kernal Estimates of the Density of Total Compensation



holding options or from holding stock, contributes less. However, the variation in total compensation is mainly driven by the last two components, which are based on wealth changes from holding firm-specific equity.

In addition, some time-series variations across the three samples are worth noting. Both the absolute level and the relative level of the compensation components change over time. First, all the components increase from the old period to the new period. For the Old and New Restricted samples, which cover the same three industries, Table 1 shows that the increase in total compensation is dominated by the equity-based components. Salary and bonuses increase almost four and half times in the three industries, which is the smallest increase among the five components. Second, the relative weights of these components change over time as well. We observe a dramatic increase in the importance of equity-based compensation. In the Old sample, cash-based compensation (salary and bonuses) is almost three times the size of equity-based compensation (the sum of the value of options granted and the value of stock granted) in the three industries, but it becomes only about one-quarter of the latter in the New Restricted sample. Thus, total compensation has increased much faster than salary and bonuses in the three industries. The component contributing the most to this dramatic shift is the options granted to managers, valued using the Black-Scholes formula. In both the restricted and unrestricted samples, the value of options granted is the biggest component of managerial compensation. In addition, the growth of stock compensation outperforms that of options compensation, even though it accounts for a smaller portion of total compensation. The value of options granted increases by more than 170 times for CEOs and by about 50 times for non-CEOs, and the value of restricted stock granted increases from almost nothing for all executives to \$551,000 for CEOs and \$159,000 for non-CEOs on average.

The third pattern is about the change in dispersion. The five components become more dispersed from the old period to the new period. Holding financial securities in their own firms rather than a well-diversified market portfolio exposes managers to considerable uncertainty. Table 1 shows that changes in wealth from holding stock and changes in wealth from holding options are more dispersed than any other component. The standard deviation is higher than for cash and bonuses, options granted, and stock granted. Note that the standard deviations of these components have dramatically increased—wealth changes in stocks and options by more than 100 fold. The two components account for a considerable amount of the increase in the volatility of total compensation. The value of options granted also contributes to a significant degree to the volatility of total compensation in the new period.

Figure 1 illustrates the distributional differences among the three samples. First, the dispersion of total compensation increases over time. The standard deviation of total compensation in the new period is several times as much as that in the old period. What's more, a significant portion of CEOs have negative compensation, even though the distribution presents a longer right tail. The negative compensation mainly stems from the change in wealth from stock held and the change in wealth from options held. To summarize, managerial compensation has substantially increased in real terms and become more dispersed. This has been accomplished by a dramatic increase in stock option grants.

3 WHAT IS THE RELATIONSHIP BETWEEN FIRM SIZE AND EXECUTIVE COMPENSATION?

The dramatic increase in both the level of CEO compensation and its sensitivity to firm performance over the past 50 years is widely documented.⁷ These studies show that, of all the components making up executive pay—including cash, bonuses, stock grants, and retirement benefits—the biggest increases have been in option grants. Thus, much of the increase in managerial compensation is attributable to increases in asset grants whose value is explicitly tied to the value of the firm. Since moral hazard explains why managerial compensation and firm performance should be connected, it is tempting to suggest that changes in the nature of moral hazard might have triggered these trends.

The theory of moral hazard provides a plausible transmission mechanism for connecting the compensation paid to a firm's executives with the returns on their firm's assets. There are two channels for inducing secular changes in managerial compensation within the principal-agent paradigm. First, contracts reflect heterogeneity across firms, such as their size, their capital-to-labor ratios, the sectors they belong to, and the dispersion of their financial returns. Consequently, changing the heterogeneity across firms induces changes in the aggregate level and variability of compensation. Second, the optimal contract is a function of the preferences and risk attitudes of managers. Changing those preferences also affects the probability distribution of compensation across executives. This section summarizes the results from Gayle and Miller (2009a), who estimate a model of moral hazard with data spanning a 60-year period in order to investigate how well these two channels explain secular changes in managerial compensation and to assess their relative importance. We then contrast their findings with

others papers in the literature. In this section, we demonstrate, using a simple moral hazard model and the data above, that the change in firm size is responsible for most of these changes observed over time.

3.1 The Relationship Between Firm Size and the Different Components of Total Compensation

The positive relationship between firm size and pay for ordinary workers is one of the most robust empirical findings in labor economics (Idson and Oi, 1999). As documented by Gayle, Golan, and Miller (2015), this is also true in the executive labor market. However, executive compensation has many more components than the pay of ordinary workers. But which component of executive compensation is responsible for the positive correlation between total compensation and firm size? Can the increase in firm size over time explain the increase in compensation over time?⁸

Table 2 presents the results of measures of firm size, total assets, and the number of employees on three of the basic components of total compensation—salary and bonuses, the value of options granted, and the value of restricted stock granted. The results are presented for three different samples and for CEOs and non-CEOs separately. For the New All sample, we observe positive and significant relationships between both measures of firm size and all three components of compensation for both CEOs and non-CEOs. When the sample is restricted to the industries in the Old sample, the same positive and significant relationships are observed, with the exception of the relationship between the number of employees and the value of restricted shares granted to CEOs, which is positive but not statistically significant. However, the positive and statistically significant relationships between firm size and the components of total compensation are not ubiquitously present in the Old sample. This gives reason to pause when considering the conclusion that the increases in the level of executive compensation over these periods are driven by an increase in firm size over time.

The most fundamental prediction of the principal-agent model for executive compensation is that in order to align shareholders' interests with the interests of the executive, the executive's compensation should be tied to the output of the firm. In practice, the change in wealth of the manager from holding firm-denominated securities is the main instrument for achieving this goal. Therefore, Table 3 presents the results of some basic regressions of the empirical measures of the change in the wealth of executives from holding options and restricted stocks of the firm on excess returns on the firm's stocks, firm size measures, and interactions of excess returns and these measures of firm size. The results show that the basic prediction of the principal-agent model is borne out by the data, that is, that there is a positive relationship between firm performance and the change in executive wealth from holding firm-denominated securities. However, like the results in Table 2, the positive relationship between firm size and the sensitivity of executive wealth to firm performance is only robust in the New All sample. This could be for a number of reasons. First, it could be that the other sample sizes are just too small to draw any conclusion. Second, it could be that the regression does not properly control for all the elements the theory predicts. Below we will use a fully specified principal-agent model to see whether we can provide more conclusive evidence from the samples we have.

Table 2
Regression of Compensation Components on Firm Characteristics

		A: CEO					
		New Restricted			Old		
	New All	Value of options granted	Salary and bonuses	Value of restricted stock granted	Value of options granted	Salary and bonuses	Value of restricted stock granted
Total assets	0.006 (0.0002)	0.037 (0.003)	0.018 (0.005)	0.008 (0.0003)	0.396 (0.077)	-0.002 (0.004)	3.28e-10 (4.7e-09)
Number of employees	4.236 (0.239)	37.79 (2.63)	9.953 (1.163)	4.46 (0.29)	70.03 (19.39)	1.623 (110.6)	-4.0e-04 (1.0e-04)
Observations	19,599	19,599	1,000	19,599	1,000	753	753

		B: Non-CEO					
		New Restricted			Old		
	New All	Value of options granted	Salary and bonuses	Value of restricted stock granted	Value of options granted	Salary and bonuses	Value of restricted stock granted
Total assets	0.004 (5.1e-05)	0.022 (0.001)	0.014 (0.001)	0.004 (1.1e-04)	0.090 (0.017)	-0.016 (0.012)	5.71e-11 (1.76e-11)
Number of employees	1.685 (0.052)	12.15 (0.764)	2.330 (0.292)	1.872 (0.116)	30.32 (4.28)	2,235 (571)	-9.1e-06 (8.4e-06)
Observations	75,379	75,379	3,693	75,379	3,693	68	68

NOTE: Standard errors are in parentheses. All regressions are controlled for the industry, year fixed effect, and debt-to-equity ratio.

Table 3**Regression of Wealth Change on Firm Characteristics and Excess Returns**

	New All		New Restricted		Old	
	Change in wealth from options held	Change in wealth from stock held	Change in wealth from options held	Change in wealth from stock held	Change in wealth from options held	Change in wealth from stock held
A. CEO						
Total assets	0.013 (0.003)	0.004 (0.005)	0.184 (0.084)	0.017 (0.142)	1.81e-06 (2.2e-05)	5.03e-06 (6.0e-05)
Number of employees	14.55 (3.06)	-1.474 (4.837)	-1.932 (21.15)	-7.699 (35.43)	-1.122 (0.642)	-1.214 (1.758)
Excess return	12,427 (335)	21,280 (530)	8,877 (1,181)	16,243 (1,979)	438 (79.5)	1,893 (218)
Excess return sq.	-817 (45.17)	-1161 (71.42)	-3448 (1,581)	9,419 (2,650)	-285 (105)	-467 (292)
Excess return × total assets	0.082 (0.011)	0.038 (0.017)	0.487 (0.337)	-0.671 (0.565)	7.6e-06 (8.6e-05)	-4.3e-04 (2.4e-05)
Excess return sq. × total assets	0.025 (0.011)	0.004 (0.017)	0.153 (0.813)	-2.453 (1.363)	-4.1e-05 (1.0e-04)	1.89e-04 (2.8e-04)
Excess return × No. emp.	158 (10.17)	128 (16.08)	167 (76.73)	143 (128.6)	1.143 (2.46)	13.82 (6.75)
Excess return sq. × No. emp.	-26.08 (5.43)	-8.696 (8.581)	19.02 (186)	605 (312)	17.19 (2.951)	5.676 (8.084)
Observations	19,599	19,599	1,000	1,000	753	753
B. Non-CEO						
Total assets	0.006 (0.000756)	0.0001 (0.00113)	0.003 (0.0174)	0.016 (0.0322)	-3.9e-5 (0.000116)	6.12e-5 (0.000314)
Number of employees	3.700 (0.726)	-0.370 (1.083)	9.958 (4.446)	-7.072 (8.202)	0.891 (4.687)	3.770 (12.72)
Excess return	4,289 (78.85)	5,017 (117.5)	2,593 (237.0)	2,559 (437.1)	182 (505.1)	511 (1,370.6)
Excess return sq.	-339 (13.37)	-148 (19.92)	-976 (320)	1,363 (591)	-302 (1,149)	1,715 (3,118)
Excess return × Total assets	0.040 (0.003)	0.027 (0.004)	0.427 (0.068)	-0.041 (0.126)	3.1e-04 (3.2e-04)	0.003 (8.7e-04)
Excess return sq. × Total assets	-0.002 (2.0e-04)	-0.002 (3.0e-04)	0.300 (0.164)	-0.535 (0.302)	4.4e-04 (0.002)	0.005 (0.004)
Excess return × No. emp.	41.47 (2.36)	29.47 (3.52)	-55.03 (16.31)	72.72 (30.08)	-12.59 (15.47)	-51.33 (41.98)
Excess return sq. × No. emp.	-1.936 (0.761)	1.094 (1.134)	-76.01 (39.86)	235.3 (73.53)	1.020 (69.66)	-87.70 (189)
Observations	75,379	75,379	3,693	3,693	68	68

NOTE: No. emp., number of employees. sq., squared. Standard errors are in parentheses. The regressions also include the debt-to-equity ratio interacted with total assets and the number of employees.

3.2 A Basic Model for Inference

The importance of moral hazard can be characterized three ways: the gross loss shareholders would incur (before accounting for managerial compensation) from the manager tending his own interests, the benefits accrued to the manager from tending his own interests instead of shareholder interests, and how much the shareholders are willing to pay to eliminate the problem of moral hazard altogether.

The first measure, denoted τ_1 , is the expected gross-output loss to the firm from switching from the distribution of abnormal returns for the diligent work to the distribution for shirking, that is, the difference between the expected firm output from the manager pursuing the firm's goals versus his own before netting out expected managerial compensation. Let v denote the value of the firm at the beginning of the period, and let x denote the firm's abnormal returns realized at the end of the period. Following the convention in the economic literature, we describe a manager who pursues the interests of the firm as "working" and a manager who pursues his own interests, when compensation is independent of firm performance, as "shirking." Then

$$\begin{aligned}\tau_1 &= E[x | \text{manager works}]v - E[x | \text{manager shirks}]v \\ &= -E[x | \text{manager shirks}]v,\end{aligned}$$

where the second equality exploits the identity that the expected value of abnormal returns is zero when the manager is working (pursuing the interests of the firm).

The second measure, τ_2 , is the nonpecuniary benefits to the manager from shirking, that is, pursuing his own goals within the firm. Let w_2 denote the manager's reservation wage to work under perfect monitoring or if there were no moral hazard problem, and let w_1 denote the manager's reservation wage from shirking. Then τ_2 , the compensating differential for these two activities, can be expressed as the difference

$$\tau_2 = w_2 - w_1.$$

We also estimate the maximum amount shareholders are willing pay to eliminate the moral hazard problem—the value of a perfect monitor. Absent moral hazard, the firm would pay the manager the fixed wage, w_2 , instead of according to compensation, $w(x)$. The firm's willingness to pay for eliminating the moral hazard problem, denoted τ_3 , is accordingly defined as

$$(4) \quad \tau_3 = E[w(x)] - w_2.$$

This measure is actually a lower bound on the shareholders' willingness to pay for a perfect monitor because it is based on asking the manager to perform the same tasks. If, however, the manager's actions could be monitored perfectly, it is plausible that shareholders would modify the manager's job description to better exploit the monitoring technology for the benefit of the firm, an issue analyzed in Prendergast (2002).

Against the output reduction from shirking, τ_1 , is the savings in managerial compensation coming from two terms: the shadow value of a perfect monitor and the cost of inducing the

manager to work diligently when a perfect monitor is removed. Subtracting from τ_1 the sum of τ_2 and τ_3 , we obtain the net income loss a firm would sustain from signing a shirking contract with a manager. This net amount represents the value of preventing the manager from undoing contracts that align his incentives with the firm's by dealing with a lender who does not recognize the folly of allowing the manager to insure himself against poor firm performance and is unaware of public disclosure laws that require the manager to report his holdings of firm-related securities.

3.2.1 A Model of Pure Moral Hazard. This section lays out a theoretical principal-agent framework on which our empirical analysis is based. At each time period t , there are three activities in which a person can be engaged: working as the firm manager in the shareholders' interests, being employed as a manager at the firm but pursuing interests different from the shareholders', or not being engaged by the firm. Let $l_t \equiv (l_{0t}, l_{1t}, l_{2t})$ denote the three possible activities, where $l_{jt} \in \{0,1\}$ is an indicator for choice $j \in \{0,1,2\}$ and

$$\sum_{j=0}^{j=2} l_{jt} = 1.$$

The indicator $l_{0t} = 1$ denotes that the manager is not employed by the firm, $l_{1t} = 1$ denotes shirking, and $l_{2t} = 1$ denotes working diligently. While l_{0t} is common knowledge, the values of (l_{1t}, l_{2t}) are hidden from the shareholders. Apart from choosing his activity, the manager also chooses his consumption for the period. Let c_t denote the manager's consumption in period t . We assume that preferences over consumption and work are parameterized by a utility function exhibiting absolute risk aversion that is additively separable over periods and multiplicatively separable with respect to consumption and work activity within periods. In the model we estimate, lifetime utility can be expressed as

$$-\sum_{t=0}^{\infty} \sum_{j=0}^3 \alpha_j \beta^t l_{jt} \exp(-\rho c_t),$$

where β is the constant subjective discount factor, α_j are utility parameters associated with setting $l_{jnt} = 1$, and ρ is the constant absolute level of risk aversion. We set $\alpha_0 = 1$ as a normalization, since behavior is invariant to linear transformation of the utility function under the independence axiom. We assume that $\alpha_2 > \alpha_1$, or that diligence is more distasteful than shirking. This assumption is the vehicle by which the manager's preferences are not aligned with shareholders' interests. We are not suggesting that managers are inherently lazy, merely that their personal goals do not motivate them to maximize the value of the firm if their compensation is independent of the firm's performance. Finally, we require $\alpha_1 > 0$ to ensure utility is increasing in consumption.

In the optimal contract, shareholders induce their manager to bear risk on only that part of the return whose probability distribution is affected by his actions. Since managers are risk averse (an assumption we test empirically), his certainty equivalent for a risk-bearing security is less than the expected value of the security, so shareholders would diversify among themselves every firm security whose returns are independent of the manager's activities, rather than use it to pay the manager. We define the abnormal returns of the firm as the residual component of returns that cannot be priced by aggregate factors the manager does not control.

In an optimal contract, compensation to the manager might depend on this residual in order to provide him with appropriate incentives, but it should not depend on changes in stochastic factors that originate outside the firm, which in any event can be neutralized by adjustments within his wealth portfolio through the other stocks and bonds he holds.

More specifically, let w_t denote the overall compensation received by the manager at the end of period t as compensation for work done during the period and v_t the value of the firm at that point in time. Then the gross abnormal returns attributable to the manager's actions is the residual

$$x_t \equiv \frac{v_t + w_t - v_{t-1}}{v_{t-1}} - \pi_t - z_t \gamma,$$

where π_t is the difference between the return on the market portfolio in period t and the return on the firm's stock, and $z_t \gamma$ is a linear combination of some risk factors, denoted z_p , that lead to systematic deviations between the expected return on the firm's shares and the market portfolio. This study assumes that x_t is a random variable that depends on the manager's activity choice in the previous period but, conditional on (l_{1t}, l_{2t}) , is independently and identically distributed across both firms and periods. Given $l_{jt} = 1$, for $j \in \{1, 2\}$ we denote the probability density function of x_t by $f_j(x_t)$.

The measures of moral hazard described in the previous section can be derived as functions of the parameters defining this framework. The expected loss per period to the firm from the manager pursuing his own interests rather than value maximization is

$$\tau_1 = -v \int x f_1(x) dx,$$

where v is the value of the firm in the previous period. The compensating differential to the manager from pursuing his own interests within the firm compared with working diligently is derived directly from the manager's utility function:

$$\tau_2 = \rho^{-1} \log \left(\frac{\alpha_2}{\alpha_1} \right).$$

In contrast to the other two measures, the welfare cost of moral hazard depends on the optimal contract. It is the expected value of managerial compensation, less its certainty equivalent:

$$\tau_3 = \int w(x) f_2(x) dx - \rho^{-1} \log \left(\frac{\alpha_2}{\alpha_0} \right).$$

The value of being able to offer a contract that creates the manager's incentive to work, as opposed to paying him a fixed wage, is thus

$$\tau_1 - \tau_2 - \tau_3 = \rho^{-1} \log \left(\frac{\alpha_1}{\alpha_0} \right) - v \int x f_1(x) dx - \int w(x) f_2(x) dx.$$

Within this model there are five parameters that might account for differences in executive compensation, that is, apart from the firm's abnormal returns: (i) the probability distribution

of abnormal returns conditional on working, (ii) the probability distribution of abnormal returns conditional on shirking, (iii) the risk-aversion parameter, (iv) the nonpecuniary benefit from shirking versus working, and (v) the nonpecuniary benefit of working versus retiring or accepting employment outside the firm. The first two production parameters, $f_2(x)$ and $f_1(x)$, determine τ_1 ; three of the taste parameters, ρ and α_2/α_1 , are used to define τ_2 ; and as our brief discussion of the optimal contract shows below, all the parameters affect τ_3 . Our empirical analysis allows each parameter to differ across firm type and executive position. We also consider the possibility that the five parameters have changed over time and that they depend on underlying factors whose values have changed. In this way we seek to discover why managerial compensation has increased and become more diffuse over the past 60 years.

3.2.2 Estimation. All three measures of moral hazard require us to compute a counterfactual. In the case of τ_1 , we must impute the firm's value before compensation is paid if the manager shirks. The manager's utility from shirking is required for τ_2 , and in the case of τ_3 , what the firm would have paid if there were no moral hazard problem. To identify the parameters of the model, we make the behavioral assumption that shareholders contract with the manager to minimize his expected compensation subject to two weak inequality constraints that induce the manager (i) not to quit the firm (participation) and (ii) to pursue the shareholders' interests rather than his own (incentive compatibility).

The two constraints are satisfied by the optimal contract with strict equality. In our framework, the participation constraint is

$$\alpha_2^{1/(1-b_t)} = E \left[\exp \left(\frac{-\rho w_t}{b_{t+1}} \right) \right],$$

where b_t is the price of a bond in period t that pays a unit of consumption per period forever. The incentive-compatibility constraint is

$$E \left\{ \exp \left(\frac{-\rho w_t}{b_{t+1}} \right) \left[g(x_t) - \left(\frac{\alpha_2}{\alpha_1} \right)^{1/(b_t-1)} \right] \right\} = 0,$$

where

$$g(x_t) \equiv \frac{f_1(x_t)}{f_2(x_t)}$$

is the ratio of the two probability density functions for shirking and working, respectively. Notice the range of $g(x_t)$ is nonnegative and that its expectation under $f_2(x_t)$ is 1. We interpret $g(x_t)$ as the signal shareholders receive about the manager's effort choice. If the realized value of the signal is zero, they conclude that the manager must have worked diligently, but the greater the realized value of the signal, the less confident they are.

The optimal cost-minimizing contract that implements diligent behavior in this setting can be written as

$$w_t = \frac{b_{t+1}}{\rho(b_t - 1)} \ln(\alpha_2) + \frac{b_{t+1}}{\rho} \ln \left[1 + \eta_t \left(\frac{\alpha_2}{\alpha_1} \right)^{1/(b_t - 1)} - \eta_t g(x_t) \right],$$

where η_t is the unique strictly positive solution to the equation

$$\int \left[\eta(\alpha_2/\alpha_1)^{1/(b_t - 1)} - \eta g(x_t) + 1 \right]^{-1} f_2(x) dx = 1.$$

Optimal compensation is the sum of two pieces. The second expression determines how compensation varies with abnormal returns through the slope of the signal function, $g(x_t)$. If moral hazard was not a factor because managerial effort could be monitored, then a manager would be paid the flat rate $w_2 = \frac{b_{t+1}}{\rho(b_t - 1)} \ln(\alpha_2)$. The expected value of the other expression is τ_3 , the shadow value of moral hazard. Tracing out the contract as a function of abnormal returns, x_t , we recover the signal function, $g(x_t)$, up to a normalization. By definition $f_1(x_t) = g(x_t)f_2(x_t)$, and the probability density function for abnormal returns is identified from data on abnormal returns. Therefore we can estimate $f_1(x_t)$, the density abnormal returns in the absence of appropriate incentives, from a nonlinear regression of w_t on x_t .

To accommodate other factors that might affect compensation but are not included in our model of moral hazard, we assume that our observation on compensation, denoted \tilde{w}_t , is the sum of true compensation, denoted w_t , plus an independently distributed error ε_t , assumed orthogonal to the other variables of interest:

$$(5) \quad \tilde{w}_t = w_t + \varepsilon_t.$$

These four equations form the basis for the estimation.

Gayle and Miller (2015) provide regularity conditions for identifying and estimating, from cross-sectional or time-series data on (w_t, x_t, r_t, p_t) , the production functions $f_1(x)$ and $f_2(x)$ along with taste parameters $(\rho, \alpha_2, \alpha_1)$. In this analysis, we parameterize $f_1(x)$ and $f_2(x)$, the distributions of abnormal returns under shirking and working, respectively, as truncated normal with support bounded below by ψ , setting

$$(6) \quad f_j(x) = \left[\Phi \left(\frac{\mu_j - \psi}{\sigma} \right) \sigma \sqrt{2\pi} \right]^{-1} \exp \left[-\frac{(x - \mu_j)^2}{2\sigma^2} \right],$$

where $j \in \{1, 2\}$ denotes shirking and working, respectively, Φ is the standard normal distribution function, and (μ_j, σ^2) denotes the mean and variance of the parent normal distribution.

As indicated in the previous section, we cannot reject the null hypothesis of restricting the mean of abnormal returns conditional on working to zero conditional on the data. We impose this restriction in the estimation of the parameter μ_2 , which implies that μ_2 is determined as an implicit function of the parameters of the truncated normal distribution under work. Denoting by ϕ the standard normal probability density function, the implicit function for μ_2 is given by

$$(7) \quad 0 = E(x_t | l_{2t} = 1) = \mu_2 + \frac{\sigma \phi[(\psi - \mu_2)/\sigma]}{1 - \Phi[(\psi - \mu_2)/\sigma]}.$$

This leaves the following to be estimated: the bankruptcy return, ψ ; the mean of the parent normal distribution under shirking, μ_1 ; the common variance of the parent normal, σ ; the risk aversion parameter, ρ ; the ratio of nonpecuniary benefits from working to shirking, α_2/α_1 ; and the ratio of nonpecuniary benefits from working to quitting, α_2/α_0 .

The parameters of the distribution of returns are estimated separately for each sector. For each sector, the production parameters μ_1 and σ^2 are specified as functions of the number of employees in the firm, the firm's assets-to-equity ratio, and an aggregate economic condition—annual gross domestic product. Denoting the controls for observed heterogeneity by z_{1t} , we assume

$$\mu_1 = u_1' z_{1t}$$

and

$$\sigma^2 = \exp(s' z_{1t}).$$

The taste parameters α_2/α_1 and α_2 are specified as linear mappings of executive rank, firm sector, the number of employees in the firm, and the total assets of the firm. Denoting this vector of controls by z_{2t} , we assume

$$(8) \quad \alpha_2/\alpha_1 = a_1' z_{2t}$$

and

$$(9) \quad \alpha_2 = a_2' z_{2t}.$$

The parameter estimates and their asymptotic standard are obtained in three steps. In the first step, maximum-likelihood estimates of the parameter vector determining the distribution of abnormal returns, (ψ, s) are obtained using data on abnormal returns over time and across companies. In the second step, we used data on the abnormal returns and managerial compensation to form a generalized methods-of-moments estimator from the participation constraint, the incentive-compatibility constraint, and the managerial compensation schedule and thus the remaining parameter (ρ, u_1, a_1, a_2) . The third step corrects the estimated standard errors in the second step to account for the pre-estimation in the first step (see Gayle and Miller, 2009a, for more details).

3.2.3 Results from the Estimated Model. Table 4 presents the estimated average loss over all firms (i.e., before compensation) from inducing the manager to shirk, both per year and as a net present-value calculation, by sector and for the two samples. The implied average losses have increased more than tenfold in the aerospace and electronics sectors and by a factor of about five in the chemicals sector. In aerospace and electronics, the mean return to

Table 4**Gross Losses to Firms from Shirking in Millions of US\$ (2000)**

Industry	Parameter τ_1	
	Old	New
Aerospace		
Per year	13.751 (29.522)	180.212 (261.294)
Present value	81.065 (177.132)	1,261.484 (1,829.058)
Chemicals		
Per year	33.392 (73.537)	160.038 (240.970)
Present value	200.352 (441.222)	1,120.266 (1,686.79)
Electronics		
Per year	16.650 (49.182)	230.566 (600.607)
Present value	99.907 (894.492)	1,613.962 (4,204.249)

NOTE: Standard deviations are in parentheses.

firms from the manager shirking has fallen and the size of the firms has increased. Both factors contribute to the larger expected losses. In the chemicals sector, the mean return from shirking, while negative, has increased and partly offsets the greater loss due to the fact that chemical firms are larger. Comparing the present value of the losses as a ratio of the total assets and the equity value of the firm, we see two measures of how much claimants on the firm, and in the latter case shareholders, would lose from not providing an incentive to managers. Controlling for sector, as a ratio of total assets, the implied losses are of the same order of magnitude in the two datasets, roughly one-ninth in aerospace, just under one-half in chemicals, and about two-thirds in electronics. As a fraction of assets, the losses that would be incurred by not providing an incentive to managers appear relatively stable in these three sectors. Since firms are more leveraged than before, the loss has increased as a fraction of equity value. This is most noticeable in two of the sectors (electronics and chemicals), where the average estimated present value of losses exceeds the average equity value in the new data but not the in old.

The dominant role of firm size in explaining the large increase in the cost of ignoring moral hazard is evident from expressing τ_1 as the negative of the product of firm size v and the expected value of the signal $g(x)$ when the manager works diligently. Differencing the estimates obtained for the two regimes, we obtain the decomposition

$$(10) \quad -\Delta\tau_1 = (\Delta v) \int xg(x)f_2(x)dx + v \int x[\Delta g(x)]f_2(x)dx + v \int xg(x)[\Delta f_2(x)]dx.$$

Table 5**Nonpecuniary Benefits of Shirking in Thousands of US\$ (2000)**

Industry	Parameter τ_2	
	Old	New
Aerospace		
CEO	2,380 (43)	4,000 (92)
Non-CEO	1,500 (72)	3,400 (78)
Chemicals		
CEO	920 (274)	3,800 (209)
Non-CEO	812 (321)	600 (451)
Electronics		
CEO	747 (432)	3,048 (387)
Non-CEO	436 (515)	2,070 (366)

NOTE: Standard deviations are in parentheses.

The first of the three expressions on the right-hand side, the change in the cost of moral hazard due to the increasing size of firms, is unambiguously positive. The second expression arises because of changes in $g(x)$. In two of the sectors, the signal has weakened, reducing the gap between $f_1(x)$ and $f_2(x)$ and thus mitigating the losses that would be incurred from encouraging the manager to pursue his own goals instead of expected-value maximization. The third expression captures the effects of the change in the distribution of abnormal returns. Noting that $f_2(x)$ has undergone a mean-preserving spread in two sectors and that $g(x)$ is a convex decreasing function, it follows that the third expression is positive for these sectors, thus reducing the loss incurred. In summary, the growth of firms increased the losses from shirking so much that it dominates the other two effects.

The two remaining measures of moral hazard, τ_2 and τ_3 , can now be computed from the estimated parameters. The nonpecuniary value of deviating from the incentive-based contract depends only on the preferences of the manager, not the distribution of the abnormal returns. For each observation, we compute a consistent estimator for τ_2 :

$$(11) \quad \tau_2 = [\rho(b_t - 1)]^{-1} b_{t+1} \ln(\alpha_2 / \alpha_1).$$

Table 5 reports, by sector and executive position, the average of the consistent estimators and consistent estimates of their respective standard deviations. The firm averages for each executive type by sector have increased in five of the six categories, by a factor of more than three for CEOs in two sectors. As a proportion of total compensation averaged over observa-

Table 6**Welfare Cost of Moral Hazard in Thousands of US\$ (2000)**

Industry	Parameter τ_3	
	Old	New
Aerospace		
CEO	500 (1,316)	10,350 (15,473)
Non-CEO	330 (1,413)	1,280 (10,501)
Chemicals		
CEO	490 (1,437)	2,973 (5,087)
Non-CEO	299 (206)	301 (1,678)
Electronics		
CEO	278 (1,257)	4,873 (17,285)
Non-CEO	67 (188)	1,206 (11,159)

NOTE: Standard deviations are in parentheses.

tions for each executive type by sector, the compensating differential to managers for pursuing their own interests has fallen in all six categories. A key factor contributing to this measure of importance, τ_2/w , is that changes in the supply and demand for managerial services has roughly doubled the compensation of managers of a firm with any given set of characteristics.

In both samples, the average τ_2 is tiny compared with the expected losses a firm would incur; our model predicts there are enormous gains from having managers act in the interest of shareholders. From the manager's perspective, however, τ_2 is quite substantial, and for a sizeable proportion of the sample population, exceeds actual and even expected compensation. This paradox is resolved by noting that the manager would be harshly penalized if the firm does poorly, which is of course more likely if he shirks. Perhaps the most striking feature of these results is how they compare with estimates of α'_1 as defined in equation (8). Table 5 averages the predicted α_2/α_1 from equation (8) over firms within each sector after taking logarithms and scaling by $\rho - 1$. Since ρ has not changed much and the estimated changes in α'_1 are for the most part insignificant or negative, we attribute the dramatic differences between the tables to the changing composition of firms within each sector. More specifically, the effects of the average growth in firm assets dominate the decline in employment and are largely responsible for the increased compensating differential to work for shareholders versus pursuing some other agenda.

The last measure of moral hazard, τ_3 , is the welfare cost of the moral hazard—the willingness of a firm to pay for a perfect monitor—thus eliminating moral hazard. From the definition of τ_3 and the solution for the optimal contract, it follows that the welfare cost may be expressed as

$$(12) \quad \tau_3 = b_{t+1} \rho^{-1} \int \ln \left[1 + \eta_t (\alpha_2 / \alpha_1)^{\frac{1}{(b_t-1)}} - \eta_t g(x_t) \right] f_2(x) dx.$$

Table 6 presents consistent estimates of the average of τ_3 in the two samples and three sectors, along with the consistent estimates of the standard deviations. The table shows that the increase in managerial compensation presented in Table 3 is mirrored in the increased cost of moral hazard. From the formula above and the formula for η_t , changes in τ_3 are ultimately attributable to changes in α_2/α_1 , $f_1(x)$, and $f_2(x)$ only. After adjusting for the general rise in living standards, the estimated model attributes practically all the increase in managerial compensation to moral hazard and hardly any of it to changes in the supply and demand for managers, as reflected in the participation condition and hence α_2/α_0 .

To further investigate the sharply increased cost of moral hazard, we first note that changes in b_{t+1}/ρ between the two samples are minimal and decompose $\Delta\tau_3$ into changes stemming from changes in $f_2(x)$ and changes in the integrand. Since $g(x_t)$ is a convex decreasing function,

$$(13) \quad \ln \left[1 + \eta_t (\alpha_2 / \alpha_1)^{1/(b_t-1)} - \eta_t g(x_t) \right]$$

is a concave increasing function. Noting again that $\Delta f_2(x)$ is a mean-preserving spread in chemicals and engineering but not in aerospace, it therefore follows that

$$(14) \quad b_{t+1} \rho^{-1} \int \ln \left[1 + \eta_t (\alpha_2 / \alpha_1)^{1/(b_t-1)} - \eta_t g(x_t) \right] \Delta f_2(x) dx$$

is positive in chemicals and engineering but negative in aerospace. Thus, changes in the distribution of abnormal returns cannot explain why the welfare cost of moral hazard increased in aerospace, the sector where the biggest increases occurred. The remaining component to explain $\Delta\tau_3$ is

$$b_{t+1} \rho^{-1} \int \left\{ \Delta \ln \left[1 + \eta_t (\alpha_2 / \alpha_1)^{1/(b_t-1)} - \eta_t g(x_t) \right] \right\} f_2(x) dx.$$

The predominant change is due to a sharp increase in α_2/α_1 averaged over firms.⁹ This component is the most important factor responsible for the increase in τ_3 . To recapitulate, increased firm assets exacerbated the conflict between managers and shareholders by creating new opportunities for managers to act against shareholder interests. These were resolved through the compensation schedule by placing greater weight on penalizing poor firm performance and rewarding superior abnormal firm returns, thus subjecting risk-averse managers to the vagaries of greater insider wealth and causing their expected compensation to rise at a rate much greater than that of national income per capita.

3.3 Summary

The welfare cost of moral hazard is a compensating differential paid to risk-averse managers to hold insider wealth and accept nondiversifiable risk that aligns their incentives to those of the stockholders, who do not price risk from an individual firm's abnormal returns because of their portfolio choices. Tables 5 and 6 show that the welfare cost of the moral hazard asso-

ciated with employing CEOs has increased by an estimated factor of more than 20 times in the aerospace and electronics sectors and sixfold in the chemicals sector. Subtracting the welfare costs of the moral hazard displayed in Table 6 from the expected compensation paid to top executives reported in Table 1, we obtain, for each of the six categories, the average certainty equivalent wage, which equates the supply and demand for managerial services for a given firm. The overall increase in the 60-year period is 2.3, the same as the increase in national income per capita. Therefore, our results attribute all the difference between the rate of increase in managerial compensation and the rate of increase in national income per capita to the rising welfare cost of moral hazard.

The cost of moral hazard depends on the preferences of managers, what shareholders observe about their behavior, the distribution of abnormal returns accruing to firms, and the characteristics of the firms managed. We do not attribute the steep increase in the welfare cost to changing tastes. Gayle and Miller (2009a) show that, if anything, the conflict between a firm with a given set of characteristics and its executives has declined. As documented in Table 1, there have been changes in the probability distributions of abnormal returns, but not all in the same direction. Gayle and Miller (2009a) show that managerial preferences for risk have remained stable in an economic sense and that the compensating differential of deviating from the goal of maximizing the expected value of the firm with a given set of characteristics has not increased. Nevertheless, Table 4 shows that if managers were paid a flat wage to prevent skimming, and if our model of moral hazard is correctly specified, then conflict between managerial and shareholder objectives would remain unresolved and the ensuing losses incurred by firms would be catastrophic and would have grown substantially over the past 60 years.

4 WHY ARE ACCOUNTING RETURNS RELATED TO EXECUTIVE COMPENSATION?

As an empirical matter, managerial compensation varies significantly with abnormal financial returns.¹⁰ The theory of pure moral hazard postulates that risk-averse managers should receive compensation that fluctuates with signals (most notably abnormal returns) that risk-neutral shareholders observe based on decisions their managers make. That is, when managers' nonpecuniary goals differ from maximizing shareholder wealth and the actions and decision of management are not monitored, managers need to be incentivized to align their goals with those of the shareholders. Although the dominant paradigm, this explanation for executive compensation has been challenged on several fronts. First, several empirical studies find that trading by corporate insiders appears profitable,¹¹ but in models of pure moral hazard, managers do not have private information about the firm's future prospects. Second, as we show below, managerial compensation depends on not only the financial returns of the firm, but also its accounting returns. In models of pure moral hazard, shareholders might use signals other than financial returns to determine optimal compensation, but the reporting of accounting income is subject to considerable discretion by the manager. In qualitative terms, these three anomalies for the pure moral hazard model can be rationalized by the hybrid model. In this section, we see how fast the hybrid principal-agent model can go in rationalizing these anomalies.

4.1 Accounting Return and Executive Compensation

For the study in this section, we use binary variables based on firm size and capital structure (the debt-to-equity ratio) to categorize firms into four types. Firm size is measured by the total assets on a firm's balance sheet (*AT*; variable names in parentheses hereafter) at the end of period t . The capital structure is reflected by the debt-to-equity ratio. The numerator of the ratio is the total liabilities (*LT*), and the denominator is the total common equity (*CEQ*). The book values of assets, liabilities, and equity are deflated to the base year 2006. We classify each firm by (i) whether its total assets averaged over years were less than or greater than the median of the average total assets of firms in the same sector and (ii) whether its average debt-to-equity ratio was less than or greater than the median of the average debt-to-equity ratio of firms in that sector. Therefore, firm type is measured by the coordinate pair (A, C), with each corresponding to whether that element is above (L) or below (S) the medians of the industry. For example, (S, L) denotes lower total assets and a higher debt-to-equity ratio than the median debt-to-equity ratio for firms in that sector. By doing so, one firm stays in the same firm category and sector for the entire sample period.

In the model presented earlier, after accepting the contractual arrangement, CEOs collect and convey their private information on the firm's prospects. We construct an empirical measure of the report by equity return evaluated at book value, which is consistent with the concept of comprehensive income in accounting practice. Accounting numbers feature the private state in the theoretical framework because many of the 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 accounting features used as an information system. The smoothing-over periods require information about the state of the firm, which may be unknown to shareholders, especially in modern firms where the control rights and ownership are separated. Based on estimation, the accounting numbers can convey private information to shareholders about prospects.

Specifically, we define the binary private state (good or bad), denoted as S_{nt} , conditional on the accounting return to equity that is measured by book value. The accounting return is denoted as r_{nt} and calculated as

$$(15) \quad r_{nt} = \frac{Asset_{nt} - Debt_{nt} + Dividend_{nt}}{Asset_{n,t-1} - Debt_{n,t-1}},$$

where for firm n in year t , *Asset* is the total assets (*AT*) at the end of year t , *Debt* is the total liability (*LT*) minus minority interest (*MIB*), *Dividend* is the dividend to common stock (*DVC*) plus the dividend to preferred stock (*DVP*). All variables are deflated to the base year 2006 before calculating the accounting return.

4.1.1 The Relationship Between Accounting Returns and Executive Compensation.

In Table 7, we summarize and contrast total compensation between the bad state and the good state for each type of firm in each sector. If shareholders did not exploit a proper compensation contract designed to solicit CEOs' private information, which we assume is embedded in accounting returns, it is unlikely that the distribution of total compensation would present a

Table 7**Total Compensation by Accounting Returns**

	Total Compensation		
	Overall	Bad state	Good state
Primary			
(S,S)	2,576 (12,787)	737 (9,331)	4,716 (15,625)
(S,L)	1,965 (8,759)	428 (7,113)	3,995 (10,206)
(L,S)	5,462 (12,957)	4,104 (10,997)	7,172 (14,903)
(L,L)	5,320 (12,734)	3,981 (11,429)	6,957 (13,997)
Consumer goods			
(S,S)	2,479 (20,991)	-1,285 (15,058)	7,351 (25,998)
(S,L)	1,858 (13,639)	-477 (10,663)	4,501 (15,974)
(L,S)	6,896 (31,409)	1,693 (23,671)	12,711 (37,427)
(L,L)	8,234 (27,373)	4,152 (22,382)	13,744 (32,131)
Services			
(S,S)	3,580 (20,116)	480 (14,521)	7,159 (24,591)
(S,L)	3,627 (16,985)	2,000 (13,124)	5,460 (20,342)
(L,S)	11,070 (37,636)	5,386 (30,669)	18,285 (43,933)
(L,L)	10,003 (26,144)	6,733 (22,103)	14,772 (30,497)

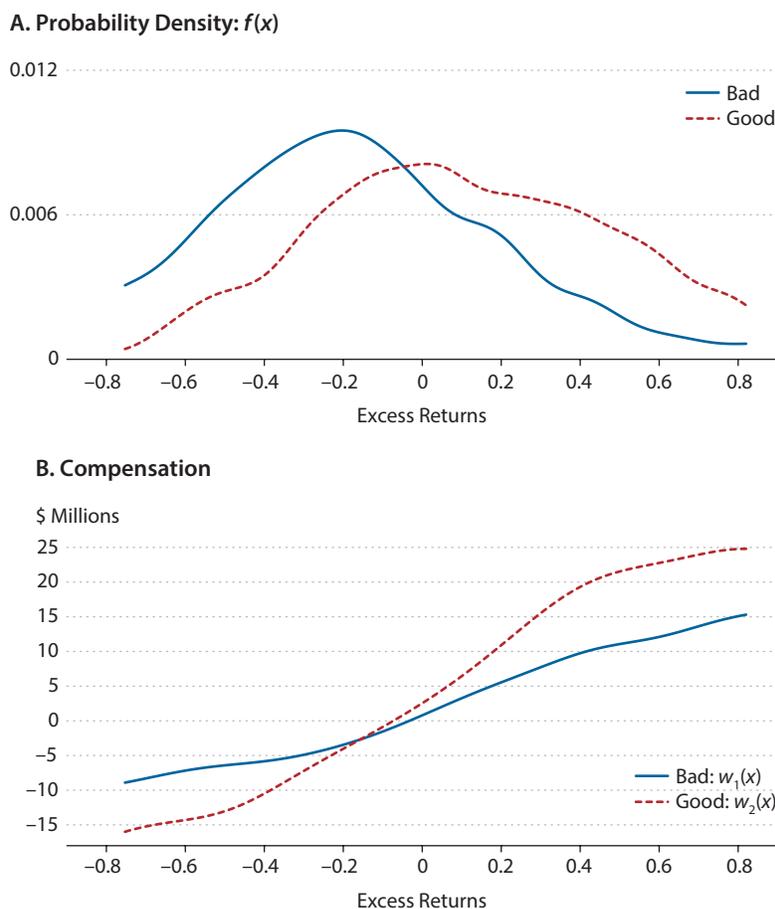
NOTE: Standard deviations are in parentheses. Firm type is measured by the coordinate-pair (A, C), where A is assets and C is the debt-to-equity ratio, with each corresponding to whether that element is above (L) or below (S) its industry median. Accounting returns are classified as "good (bad)" if they are greater (less) than the industry average. Assets (compensation) is measured in millions (thousands) of US\$.

systematic distinction between the two states. However, Table 7 presents the opposite. It reports the mean and standard deviations of total compensation conditional on firm size, capital structure, and industrial sector. The universal pattern is that total compensation always shows a lower level and smaller standard deviation in the bad state than in the good state. CEOs are paid more in the good state, but their pay is less concentrated.

In addition, total compensation is, not surprisingly, smaller in small firms than in large firms, regardless of the state. The relationship between capital structure and total compensation

Figure 2

Empirical Excess-Return Densities and the Total Compensation Schedule



NOTE: The panels present the non-parametrically estimated density of excess returns and the optimal compensation of firms with large size and high leverage in the primary sector. The compensation of both periods is anchored at bond prices equal to 16.5 (b_t) and 16.4 (b_{t+1}).

behaves differently between the two states. In the bad state, firms with a high debt-to-equity ratio are more likely to have higher compensation, except the firms in the primary sector. However, in the good state, this happens only in large firms in the consumer goods sector.

After some simple calculations, the smallest difference in total compensation between the two states is about \$3 million, for large firms in the primary sector with a high debt-to-equity ratio, and the largest difference is nearly \$12 million, for large firms in the services sector with a low debt-to-equity ratio. The between-state difference tends to be larger in larger firms than in smaller firms in the consumer goods sector and the services sector, but smaller in the primary sector. The difference is always smaller in firms with a high debt-to-equity ratio, regardless of size or sector.

Figure 2 graphically compares the distribution of total compensation and excess returns between the two states, taking large firms in the primary sector as an example. Panel A of Figure 2 presents the kernel density of excess returns for each of the two states. Excess returns are lower on average in the bad state than in the good state, indicating the lower compensation in the bad state, as Table 7 reports, which may reflect punishment of inferior performance too. Thus we need a more structured research design to separate the effect of productive performance and that of information rent on the level of total compensation.

In Panel B of Figure 2, the non-parametrically estimated compensation schedule is compared between the two states. The curve of the optimal contract in the good state is steeper than that in the bad state, indicating that in the good state, compensation is more sensitive to performance. The empirically estimated compensation schedule increases with excess returns and flattens at very high rates of excess returns. These features illustrate the agency problem and suggest that hidden information, not just hidden actions, may be a part of the agency problem.

4.2 A Hybrid Principal-Agent Model

To this end we now lay out a dynamic principal-agent model of optimal contracting between risk-neutral shareholders and a risk-averse CEO, based on Gayle and Miller (2015), in which the CEO has hidden information and also takes actions that cannot be directly observed by shareholders. An important feature of this model is that it treats accounting information as a non-verifiable statement by the CEO, whose credibility depends on the incentives that determine his payoff as a function of what he reports.

At the beginning of period t , the CEO is paid compensation, denoted by w_t , for his work the previous period, denominated in terms of period- t consumption units. He makes his consumption choice, a positive real number denoted by c_t , and the board proposes a new contract. The board announces how CEO compensation will be determined as a function of what he will disclose about the firm's prospects, denoted by $r_t \in \{1, 2\}$ ¹², and its subsequent performance, measured by excess returns x_{t+1} , revealed at the beginning of the next period. We denote this mapping by $w_{rt}(x)$, with the subscript t designating that the optimal compensation schedule may depend on current economic conditions, such as bond prices. Then the CEO chooses whether to be engaged by the firm. Denote this decision by the indicator $l_{t0} \in \{0, 1\}$, where $l_{t0} = 1$ if the CEO chooses to be engaged outside the firm and $l_{t0} = 0$ if he chooses to be engaged inside the firm.

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

The CEO then makes his unobserved labor-effort choice, denoted by $l_{stj} \in \{0, 1\}$ for $j \in \{1, 2\}$ for period t , which may depend on his private information about the state of the firm. There are two possibilities: to work and diligently pursue the shareholders' objectives of value maximization, thus setting $l_{st2} = 1$, or to shirk and accept employment with the firm but follow the objectives he would pursue if he were paid a fixed wage by setting $l_{st1} = 1$. Let $l_{st} \equiv (l_{st1}, l_{st2})$. Since leaving the firm, working, and shirking are mutually exclusive activities, $l_{st1} + l_{st2} = 1$.

At the beginning of period $t + 1$, excess returns for the firm, x_{t+1} , are drawn from a probability distribution that depends on the true state, s_t , and the CEO's action, l_{st} , in period t . We denote the probability density function for excess returns when the CEO works diligently and the state is s by $f_{st}(x)$. Similarly, let $f_{st}(x)g_{st}(x)$ denote the probability density function for excess returns in period t when the CEO shirks. Thus, for both states $s_t \in \{1, 2\}$:

$$(16) \quad \int x f_{st}(x) g_{st}(x) dx \equiv E_{st}[x g_{st}(x)] < E_{st}[x] \equiv \int x f_{st}(x) dx,$$

with the inequality reflecting the shareholders' preference for diligent work over shirking. Since $f_{st}(x)g_{st}(x)$ is a density, $g_{st}(x)$ is positive and integrating $f_{st}(x)g_{st}(x)$ with respect to x demonstrates $E_{st}[g_{st}(x)] = 1$. We assume the likelihood of shirking declines to zero as excess returns increase without bound:

$$(17) \quad \lim_{x \rightarrow \infty} [g_{st}(x)] = 0$$

for each $s \in \{1, 2\}$. We assume the weighted-likelihood ratio of the second state occurring relative to the first given any observed value of excess returns, $x \in R$, converges to an upper finite limit as x increases such that

$$(18) \quad \lim_{x \rightarrow \infty} [\varphi_{2t} f_{2t}(x) / \varphi_{1t} f_{1t}(x)] \equiv \lim_{x \rightarrow \infty} [h_t(x)] = \sup_{x \in R} [h_t(x)] \equiv \bar{h}_t < \infty.$$

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

$$(19) \quad -\sum_{t=0}^{\infty} \sum_{j=0}^2 \beta^t \alpha_{jt} l_{tj} \exp(-\gamma_t c_t),$$

where β is the constant subjective discount factor, γ_t is the constant absolute level of risk aversion, and α_{jt} is a utility parameter that measures the distaste from working at level $j \in \{0, 1, 2\}$. We assume working is more distasteful than shirking, meaning $\alpha_{2t} > \alpha_{1t}$, and normalize $\alpha_{0t} = 1$.

In this framework, there are no gains from a long-term arrangement between shareholders and the CEO: The optimal long-term contract between shareholders and the CEO decentralizes to a sequence of short-term one-period contracts. Therefore, the model can be solved in two steps. First we solve for the optimal consumption and savings plan for a CEO about to retire. It can be proved in this model that given the CEO's reporting about the state of the firm and the true state of the firm, his employment and effort choices depend on his preference parameters $(\alpha_{1t}, \alpha_{2t}, \gamma_t)$, the distribution of excess returns when he shirks $f_{st}(x)g_{st}(x)$ and when he works $f_{st}(x)$, and aggregate economic conditions as reflected in bond prices (b_t, b_{t+1}) . However, the employment and effort choices do not depend on his current (outside) wealth. Let $r_t(s)$ denote the CEO's disclosure rule about the state when the true state is $s_t \in \{1, 2\}$.

If the CEO, offered a contract of $w_{rt}(x)$ for announcing r , retires in period t or $t + 1$ by setting $(1 - l_{t0})(1 - l_{t+1,0}) = 0$, upon observing the state s and reporting $r_t(s)$, he optimally chooses $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$ to minimize

$$(20) \quad \sum_{s=1}^2 \varphi_{st} \left\{ l_{t0} + (\alpha_{1t} l_{st1} + \alpha_{2t} l_{st2})^{\frac{1}{(b_t-1)}} E_{st} \left[\exp \left(-\frac{\gamma_t w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_{st}(x) l_{st1} + l_{st2}] \right] \right\}.$$

The optimal short-term contract for shareholders is found by minimizing the expected compensation subject to four constraints the CEO prefers: (i) working for a period rather than leaving the firm, (ii) being truthful rather than lying, (iii) working instead of shirking, or (iv) being truthful and working diligently rather than lying and shirking. Suppressing for expositional convenience the bond price b_{t+1} and recalling our assumption that b_{t+1} is known at period t , we now let $v_{st}(x)$ measure how (the negative of) utility is scaled up by $w_{st}(x)$:

$$(21) \quad v_{st}(x) \equiv \exp \left(-\frac{\gamma_t w_{st}(x)}{b_{t+1}} \right).$$

First, to induce an honest, diligent CEO to participate, his expected utility from employment must exceed the utility he would obtain from retirement. Setting $(l_{t2}, r_t) = (1, s_t)$ in (20) and substituting in $v_{st}(x)$, the participation constraint is thus

$$(22) \quad \sum_{s=1}^2 \int \varphi_{st} v_{st}(x) f_{st}(x) dx \leq \alpha_{2t}^{\frac{1}{(b_t-1)}}.$$

Second, given his decision to stay with the firm one more period and to truthfully reveal the state, the incentive-compatibility constraint induces the CEO to prefer working to shirking for $s_t \in \{1, 2\}$. Substituting the definition of $v_{st}(x)$ into (20) and comparing the expected utility obtained from setting $l_{t1} = 1$ with the expected utility obtained from setting $l_{t2} = 1$ for any given state, we obtain the incentive-compatibility constraint for work:

$$(23) \quad 0 \leq \int \left(g_{st}(x) - \left(\frac{\alpha_{2t}}{\alpha_{1t}} \right)^{\frac{1}{(b_t-1)}} \right) v_{st}(x) f_{st}(x) dx.$$

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

$$(24) \quad 0 \leq \int [v_{1t}(x) - v_{2t}(x)] f_{2t}(x) dx.$$

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

$$(25) \quad 0 \leq \int \left(\left(\frac{\alpha_{1t}}{\alpha_{2t}} \right)^{\frac{1}{(b_t-1)}} v_{1t}(x) g_{2t}(x) - v_{2t}(x) \right) f_{2t}(x) dx,$$

where $(\alpha_{1t}/\alpha_{2t})^{1/(b_t-1)} v_{1t}(x)$ is proportional to the utility obtained from shirking and announcing the first state and $f_{2t}(x) g_{2t}(x)$ is the probability density function associated with shirking when the second state occurs. Minimizing expected compensation amounts to choosing $v_{st}(x)$ that maximizes

$$(26) \quad \sum_{s=1}^2 \int \varphi_{st} \ln[v_{st}(x)] f_{st}(x) dx.$$

Noting $\ln v_{st}$ is concave, increasing in v_{st} , the expectation operator preserves concavity, so the objective function is concave in $v_{st}(x)$ for each x . Each constraint is a convex set and its intersection is too. Therefore, we can appeal to the Kuhn-Tucker theorem, which guarantees there is a unique positive solution to the equation system formed from the first-order conditions augmented by the complementary-slackness conditions.

4.3 Comparing the Pure and Hybrid Model Contracts

The optimal contract for a parameterization of the hybrid model is plotted in Panel A of Figure 3. This parameterization follows Margiotta and Miller (2000) in assuming that excess returns are drawn from a truncated distribution, with a common lower bound for all states and independent of the effort level.¹³ For comparison purposes, we also plot in Panel B optimal compensation for the analogous two-state pure moral hazard model (where there are hidden actions but the state is known), by $y_{st}(x)$.

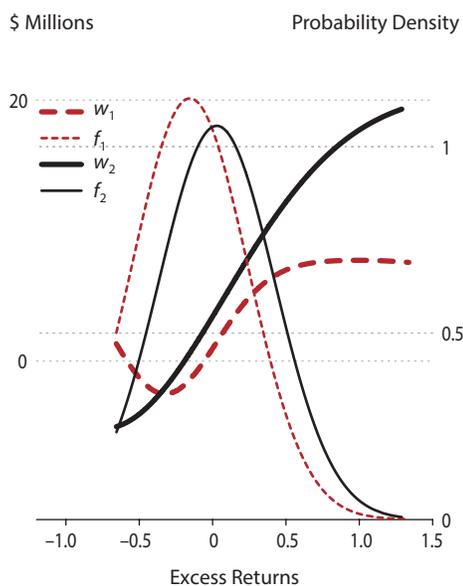
To derive $y_{st}(x)$, the optimal compensation in the analogous two-state pure moral hazard model, we drop the truth-telling and sincerity constraints, replace the single participation constraint with one for each state, retain both incentive-compatibility constraints, minimize the modified objective function, use the participation constraints to substitute out their associated Kuhn-Tucker multiplier, and rearrange the first-order conditions to obtain

$$(27) \quad y_{st}(x) = \gamma^{-1} \frac{b_{t+1}}{b_t - 1} \ln \alpha_2 + \gamma^{-1} b_{t+1} \ln \left[1 + \eta_{st}^p (\alpha_2 / \alpha_1)^{\frac{1}{b_t-1}} - \eta_{st}^p g_{st}(x) \right],$$

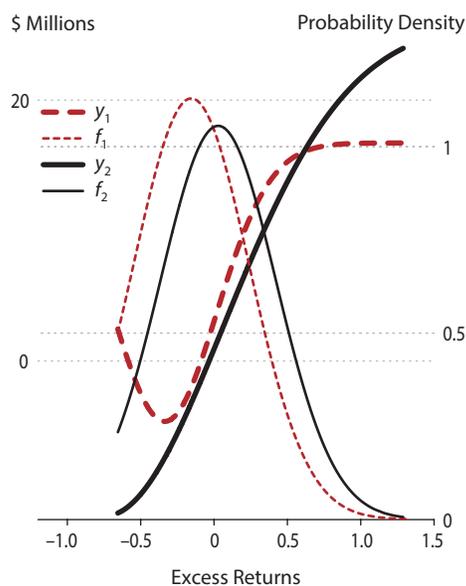
Figure 3

Optimal Compensation Schedules

A. Hybrid Moral Hazard



B. Pure Moral Hazard



NOTE: The plots use the return and optimal compensation of firms with small size and low leverage in the primary sector in the 1993-2001 period. The risk aversion parameter γ is equal to 0.08. The effort cost coefficient of shirking (α_1) equals 0.96, and the effort cost coefficient of working (α_2) equals 1.20. Bond prices are 16.5 (b_t) and 16.4 (b_{t+1}). The excess return is approximated by one-side truncated normal distribution $TN(a, \mu, \sigma)$ with truncated points on the left (a), mean (μ), and standard deviation (σ) as follows: working in the bad state: $TN(-0.66, -0.16, 0.39)$; working in the good state: $TN(-0.66, 0.03, 0.39)$; shirking in bad state: $TN(-0.66, -0.25, 0.27)$; and shirking in good state: $TN(-0.66, -0.11, 0.36)$. The probability of the bad state is 0.54, and the probability of the good state is 0.46.

where η_{st}^p is the unique positive solution to

$$(28) \quad \int_{\bar{x}}^{\infty} \frac{g_{st}(x) - \left[\frac{\alpha_2}{\alpha_1} \right]^{\frac{1}{b_t-1}}}{1 + \eta_{st}^p \left[\frac{\alpha_2}{\alpha_1} \right]^{\frac{1}{b_t-1}} - \eta_{st}^p g_{st}(x)} f_s(x) dx = 0.$$

We approximate the integral (28), accounting for the singularity problem that occurs when the denominator of the integrand is either zero or ∞ . First, we perform a grid search to detect the singularity points in the range of x . These singularity points divide the entire range of x into a number of subintervals. The integral (28) is approximated for a given η_{st}^p by first being approximated on each subinterval and then summed over the entire range. Then we numerically solve for the optimal value of η_{st}^p that satisfies (28) based on this approximated integral.

Figure 3 illustrates four important features. Since compensation in both models is a function of the likelihood ratio between the densities of the excess returns for working and shirking, not excess returns alone, the wage contract is not necessarily monotonically increasing in excess returns. For example, in the bad states of both models of the illustrated parameterization, pay optimally declines with marginal increments to excess returns when they are less than -0.5 . The same explanation applies to compensation leveling out at high levels of excess returns; the likelihood ratio converges to a constant, zero, under the assumption of a truncated normal distribution.

The other two noteworthy features relate to differences between the pure and hybrid contracts. The slope of the hybrid compensation schedule is everywhere greater in the good state than in the bad state, whereas in the pure moral hazard model the slope in the bad state is greater than in the good state over the intermediate range, where much of the probability mass of both excess distributions lie. Thus the point where the schedules cross is higher in the pure moral hazard model than in the hybrid model. The figure also illustrates two analytical results: In the hybrid model, expected utility of the agent is greater in the good state than the bad state; but in the pure moral hazard model, expected utilities are equalized across states. Intuitively, the argument is that in the hybrid model the principal induces the agent to truthfully reveal the good state by promising (i) more expected utility in the good state and (ii) a flatter compensation profile in the bad state.

Finally, because the constraints in the pure moral hazard optimization problem are not a subset of those in the hybrid model, there is no presumption that the expected compensation in the pure moral hazard case is lower than in the hybrid model. In other words, the principal may find it cheaper not to know the private information if he can optimally spread the utility the agent receives across both states, rather than meet the participation constraint in each state. Indeed, our parameterization illustrates an instance where the agency cost in the pure moral hazard model is greater than in its hybrid counterpart. Finally, a comparison of Figure 3, which is produced from the estimated hybrid model (see Gayle and Miller, 2015, and Gayle, Li and, Miller, 2016, for details of the estimation), and Figure 2, nonparametric estimates from the data, reveals a strict similarity between the optimal contract from the hybrid model and the empirical contract observed in the data.

4.4 Summary

If every piece of information a manager knows about his firm is codified and independently verifiable in a court of law, managers can be compelled to reveal all their private information through the firm's accounting records. In this case, a basic pure moral hazard model would apply conditional on verifiable information. Within the current legal system, however, managers exercise considerable discretion about how much information they release describing the state of their own firms. If the penal code for accounting protocol were augmented by incentives embedded in managerial compensation designed to elicit truthful revelation, a hybrid model of moral hazard would apply. Therefore, the fact that non-verifiable information is used in compensating executives does not reject the principal-agent model.

Gayle and Miller (2015) and Gayle, Li, and Miller (2016) use a large panel dataset measuring compensation of chief executive officers, financial and accounting returns, and size

and sector background characteristics of publicly traded firms. They investigate whether the hybrid model can reconcile the fact that executives are paid based on accounting returns. In the pure moral hazard models estimated and tested by Gayle and Miller (2015), managers do not have discretion about how they report accounting returns. In the hybrid model, estimated by Gayle and Miller (2015) and Gayle, Li, and Miller (2016), they interpret data on accounting returns as information reported by the CEO that cannot be fully corroborated by shareholders. Thus our empirical study compares and contrasts the role of these alternative information assumptions about accounting returns within competing models of moral hazard.

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

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

5 CONCLUSION

In this article we illustrate how using theory-based estimation together with a model-motivated measure of total compensation can help overcome the problem that the exact prediction of the principal-agent model (as well as most models of imperfect information) depends on many objects unobservable to the econometrician. The article concludes that using a model-consistent measure of compensation and theory-based estimation shows that executive compensation broadly conforms to the principal-agent theory; however, each situation and the variables used have to be carefully modeled, identified, and estimated.

Across all the different specifications of the principal-agent models summarized in this article, two robust facts emerge. First, more than 80 percent of total executive pay is from the risk premium paid to resolve the agency problem (see Gayle and Miller, 2009a,b, and Gayle, Golan, and, Miller, 2015). Second, the size of this risk premium is explained mostly by firm size (see Gayle and Miller, 2009a, and Gayle, Golan, and Miller, 2015). A risk premium, rationalized in the principal-agent model by incentive contracts to deter shirking, accounts for approximately 80 percent of the firm-size pay premium. More specifically, the estimated risk premium is \$1.6 million for small firms, \$2.6 million for medium-sized firms, and \$4.9 million for large firms. These findings are consistent with explanations that suggest large firms pay large efficiency wages to prevent shirking. Therefore, in order to understand the reasons for the increase in executive pay and the increased top income shares over time, researchers need to examine the reasons behind the increase in firm size over time. ■

NOTES

- ¹ See Piketty and Saez (2003) for more details on these trends.
- ² See Gayle and Miller (2009a) for more details.
- ³ For example, in the insurance market, the basic prediction of agency theory is that there is a correlation between insurance coverage and risk. If this prediction is correct, then policyholders who are known to themselves (but not to their insurer) to be high risk will tend to choose higher insurance coverage (lower deductibles); thus, coverage and risk are expected to be positively correlated. This coverage-risk correlation has been the major focus of empirical work in this area, where risk is measured by the likelihood of an accident.
- ⁴ In the managerial compensation setting, the positive correlation test boils down to testing the prediction of a positive correlation between executive compensation and the performance of the firm, that is, pay for performance sensitivity.
- ⁵ In this context, efficiency means a contractual arrangement with the minimum cost of balancing the incentive provision with the insurance needs of the executives.
- ⁶ Calculated based on data from the CRSP US Stock Databases ©2010 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business.
- ⁷ See Hall and Liebman (1998), Murphy (1999), Gayle and Miller (2009a), and the data analysis in Section 2.
- ⁸ See Gabaix and Landier (2008), Treviö (2008), and Gayle and Miller (2009a).
- ⁹ See Gayle and Miller (2009a) for more details.
- ¹⁰ See Antle and Smith (1985, 1986), Hall and Liebman (1998), and Gayle and Miller (2009a), who find that about half the total variation in compensation can be explained by a nonlinear regression on excess returns, industry effects, and bond prices.
- ¹¹ See Lorie and Niederhoffer (1968), Jaffe (1974), Finnerty (1976), and Seyhun (1986), who find that insiders tend to buy before an abnormal rise in stock prices and sell before an abnormal decline. Seyhun (1992a,b) presents evidence showing that insiders earn over 5 percent abnormal returns on average and determines that insider trades predict up to 60 percent of the total variation in one-year-ahead returns. Gayle and Miller (2009b) construct a simple self-financing dynamic portfolio strategy based on changes in asset holdings by managers that significantly outperforms the market portfolio, realizing over 90 percent of the gains that could have been achieved with perfect foresight.
- ¹² $r_t = 1$ if the private state is bad, and $r_t = 2$ if it is good.
- ¹³ If the lower bound depends on whether the agent works or shirks, a first-best solution is attained by imposing a sufficiently harsh penalty on the agent when abnormal returns can be attained only by shirking or otherwise paying the agent the first-best fixed wage. See Mirrlees (1975).

REFERENCES

- Aggarwal, Rajesh K. and Samwick, Andrew A. "Executive Compensation, Strategic Competition, and Relative Performance Evaluation: Theory and Evidence." *Journal of Finance*, December 1999, 54(6), pp. 1999-2043; <https://doi.org/10.1111/0022-1082.00180>.
- Antle, Rick and Smith, Abbie. "Measuring Executive Compensation: Methods and an Application." *Journal of Accounting Research*, Spring 1985, 23(1), pp. 296-325; <https://doi.org/10.2307/2490920>.
- Antle, Rick and Smith, Abbie. "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives." *Journal of Accounting Research*, Spring 1986, 24(1), pp. 1-39; <https://doi.org/10.2307/2490802>.
- Cardon, James H. and Hendel, Igal. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics*, February 2001, 32(3), pp. 408-27; <https://doi.org/10.2307/2696362>.
- Chiappori, Pierre-Andre and Salanie, Bernard. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, February 2000, 108(1), pp. 56-78; <https://doi.org/10.1086/262111>.
- Cohen, Alma. "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market." *Review of Economics and Statistics*, May 2005, 87(2), pp. 197-207; <https://doi.org/10.1162/0034653053970294>.
- Finnerty, Joseph E. "Insiders and Market Efficiency." *Journal of Finance*, September 1976, 31(4), pp. 1141-48; <https://doi.org/10.1111/j.1540-6261.1976.tb01965.x>.
- Gabaix, Xavier and Landier, Augustin. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, February 2008, 123(1), pp. 49-100; <https://doi.org/10.1162/qjec.2008.123.1.49>.
- Gayle, George-Levi; Golan, Limor and Miller, Robert A. "Promotion, Turnover, and Compensation in the Executive Labor Market." *Econometrica*, November 2015, 83(6), pp. 2293-369; <https://doi.org/10.3982/ECTA11020>.
- Gayle, George-Levi; Li, Chen and Miller, Robert A. "Was Sarbanes-Oxley Costly? Evidence from Optimal Contracting on CEO Compensation." Working Paper 2015-017B, Federal Reserve Bank of Saint Louis, 2016; <https://research.stlouisfed.org/wp/more/2015-017>.
- Gayle, George-Levi and Miller, Robert A. "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review*, 2009a, 99(5), pp. 1740-769; <https://doi.org/10.1257/aer.99.5.1740>.
- Gayle, George-Levi and Miller, Robert A. "Insider Information and Performance Pay." *CESifo Economic Studies*, 2009b, 55(3-4), pp. 515-41; <https://doi.org/10.1093/cesifo/ifp010>.
- Gayle, George-Levi and Miller, Robert A. "Identifying and Testing Models of Managerial Compensation." *Review of Economic Studies*, January 2015, 82(3), pp. 1074-118; <https://doi.org/10.1093/restud/rdv004>.
- Hall, Brian J. and Liebman, Jeffrey B. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, August 1998, 103(3), pp. 653-91; <https://doi.org/10.1162/003355398555702>.
- Idson, Todd L. and Oi, Walter Y. "Workers Are More Productive in Large Firms." *American Economic Review*, May 1999, 89(2), pp. 104-08; <https://doi.org/10.1257/aer.89.2.104>.
- Jaffe, Jeffrey F. "Special Information and Insider Trading." *Journal of Business*, July 1974, 47(3), pp. 410-28; <https://doi.org/10.1086/295655>.
- Jensen, Michael C. and Murphy, Kevin J. "Performance Pay and Top-Management Incentives." *Journal of Political Economy*, April 1990, 98(2), pp. 225-64; <https://doi.org/10.1086/261677>.
- Lorie, James H. and Niederhoffer, Victor. "Predictive and Statistical Properties of Insider Trading." *Journal of Law and Economics*, April 1968, 11(1), pp. 35-53; <https://doi.org/10.1086/466642>.
- Margiotta, Mary M. and Miller, Robert A. "Managerial Compensation and the Cost of Moral Hazard." *International Economic Review*, August 2000, 41(3), pp. 669-719; <https://doi.org/10.1111/1468-2354.00080>.
- Mirrlees, James A. "The Theory of Moral Hazard and Unobservable Behaviour." Unpublished manuscript, October 1975, Nuffield College, Oxford; https://www.jstor.org/stable/2566946?seq=1#page_scan_tab_contents.
- Murphy, Kevin J. "Executive Compensation," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*. Volume 3B, Chap. 38. Elsevier, 1999, pp. 2485-563; [https://doi.org/10.1016/S1573-4463\(99\)30024-9](https://doi.org/10.1016/S1573-4463(99)30024-9).

Gayle, Li, Miller

- Piketty, Thomas and Saez, Emmanuel. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics*, February 2003, 118(1), pp. 1-41; <https://doi.org/10.1162/00335530360535135>.
- Prendergast, Canice. "The Tenuous Trade-Off between Risk and Incentives." *Journal of Political Economy*, October 2002, 110(5), pp. 1071-102; <https://doi.org/10.1086/341874>.
- Seyhun, H. Nejat. "Insiders' Profits, Costs of Trading, and Market Efficiency." *Journal of Financial Economics*, June 1986, 16(2), pp. 189-212; [https://doi.org/10.1016/0304-405X\(86\)90060-7](https://doi.org/10.1016/0304-405X(86)90060-7).
- Seyhun, H. Nejat. "The Effectiveness of the Insider-Trading Sanctions." *Journal of Law and Economics*, April 1992a, 35(1), pp. 149-82; <https://doi.org/10.1086/467248>.
- Seyhun, H. Nejat. "Why Does Aggregate Insider Trading Predict Future Stock Returns?" *Quarterly Journal of Economics*, November 1992b, 107(4), pp. 1303-31; <https://doi.org/10.2307/2118390>.
- Terviö, Marko. "The Difference that CEOs Make: An Assignment Model Approach." *American Economic Review*, June 2008, 98(3), pp. 642-68; <https://doi.org/10.1257/aer.98.3.642>.