ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Was Sarbanes-Oxley Costly? Evidence from Optimal Contracting on CEO Compensation.

Authors	George-Levi Gayle, Chen Li, and Robert A. Miller
Working Paper Number	2015-017B
Revision Date	April 2016
Citable Link	https://doi.org/10.20955/wp.2015.017
Suggested Citation	Gayle, G.-L., Li, C., Miller, R.A., 2016; Was Sarbanes-Oxley Costly? Evidence from Optimal Contracting on CEO Compensation., Federal Reserve Bank of St. Louis Working Paper 2015-017. URL https://doi.org/10.20955/wp.2015.017

Was Sarbanes-Oxley Costly? Evidence from Optimal Contracting on CEO Compensation[1]

George-Levi Gayle
Department of Economics, Washington University in St. Louis
and Federal Reserve Bank of St. Louis

Chen Li
Zicklin School of Business, Baruch College, CUNY

Robert A. Miller
Tepper School of Business, Carnegie Mellon University

April 25, 2016

[1]We thank participants for their comments at the 2015 Jr. Accounting Theory Conference, and in seminars given at Baruch College CUNY, City University of London, the University of Calgary, the University of Nottingham, the University of Minnesota, and the Wharton School of the University of Pennsylvania. Li acknowledges financial support from a PSC-CUNY Award, funded by The Professional Staff Congress and The City University of New York. Email addresses: ggayle@wustl.edu (Gayle), chen.li@baruch.cuny.edu (Li), ramiller@andrew.cmu.edu (Miller).

Abstract

This paper investigates the effects of the Sarbanes-Oxley Act (SOX) on CEO compensation, using panel data constructed for the S&P 1500 firms on CEO compensation, financial returns, and reported accounting income. Empirically SOX (i) changes the relationship between a firm's abnormal returns and CEO compensation, (ii) changes the underlying distribution of abnormal returns, and (iii) significantly raises the expected CEO compensation in the primary sector. We develop and estimate a dynamic principal agent model of hidden information and hidden actions to explain these regularities. We find that SOX (i) increased the administrative burden of compliance in the primary sector, but reduce this burden in the service sector, (ii) increased agency costs in most categories of the firms, and (iii) reduced the off-equilibrium loss from the CEO shirking. (JEL C10, C12, C13, J30, J33, M50, M52, M55)

1 Introduction

This article is an empirical investigation of the changes in chief executive officer (CEO) compensation resulting from the passage of the Sarbanes-Oxley Act (SOX), a legislative response enacted in 2002 by the U.S. government after a wave of corporate governance failures at many prominent companies. Many studies have investigated how SOX has affected firm behavior, including switching earnings management methods,[1] reducing investment,[2] and delisting.[3] Several studies attempt to quantify the net benefit of SOX by investigating the stock market reaction to the approval of the SOX provisions by the securities and exchange commission (SEC), but the evidence is mixed and varies across firm type.[4] Yet how CEO compensation has been modified by shareholders in response to this regulation change is underexplored.[5] SOX has changed the environment that CEOs confront, and compensation is the crucial mechanism for exercising corporate governance to mitigate agency problems that arise from CEOs' hidden actions and information. For example, SOX includes a clawback provision, Section 304, requiring the CEO and chief financial officer (CFO) to return to shareholders performance-based components of their compensation when financial information and reports of the firm do not meet the requirements of federal securities laws; several cases were successfully prosecuted under this provision.[6] The consequences of SOX for CEO compensation are thus an important factor in any overall

[1] Cohen et al. (2008) find that accrual-based earnings management declined after the passage of SOX but real earnings management increased at the same time.

[2] Bargeron et al. (2010) find that, compared with non-U.S. firms, U.S. firms reduced investment in reseach and development and capital. Kang et al. (2010) find that (i) overall firms apply a higher rate to discount the payoff of investment projects and (ii) firms with good governance, a good credit rating, and early compliance with section 404 of SOX have become more cautious about investment.

[3] Engel et al. (2007) find that small firms chose to go private to avoid the cost of SOX. Leuz et al. (2008) show that the increased deregistration is driven mainly by firms that go dark, rather than private.

[4] Zhang (2007) finds a negative market reaction and Jain and Rezaee (2006) find a positive one. Livtak (2007) finds that the decline in the stock price of foreign cross-listed firms was greater than for the U.S. market index, cross-listed foreign firms not subject to SOX, and foreign firms not cross-listed. Hochberg et al. (2009) find that firms that had lobbied against SOX experienced positive abnormal returns.

[5] Cohen et al. (2013) documents a decline in pay-for-performance sensitivity after SOX. Carter et al. (2009) find the following: The weight of earnings increased for CEO bonuses; upward earnings management decreased, and the cash salary components decreased in the total compensation after SOX. Nekipelov (2010) attributes an increase in post-SOX salary and bonuses to increased risk aversion.

[6] See Appendix B of Ang, Cheng, and Fulmer (2013).

evaluation of SOX. This article investigates along which dimensions, and to what extent, the SOX regulatory changes exacerbated or mitigated the agency problems pertaining to executive management in different types of firms.

To evaluate the effects of SOX, we estimate a dynamic principal agent model of optimal contracting. The advantage of this approach is that changes in CEO compensation and agency costs can be attributed explicitly to changes in the fundamentals defining the primitives of the model. The framework also provides several measures of welfare costs that can be used to evaluate SOX. To understand how shareholders modify CEO compensation contracts in response to SOX, we estimate the changes of agency costs embedded in CEO compensation from the pre-SOX era to the post-SOX era. These costs are due to two fundamental frictions in the agency relationships between shareholders and CEOs — that is, CEOs' hidden action (moral hazard) and their hidden information about firms' prospects.

Our model has four key features motivated by previous work.[7] First, the model is based on hidden actions that create moral hazard, now widely acknowledged as the prime force explaining why the wealth of a CEO fluctuates with the value of the firm he or she manages. Second, the model also explicitly treats private information from which CEOs directly benefit through their holdings of financial securities in their own firms. This stylized fact is not controversial; for example, Gayle and Miller (2009a) show that following a simple portfolio strategy based on compensation schemes would have netted investors an extra 10 percent over and above holding the market portfolio. Third, accounting information is interpreted within the model as a signal that reveals the CEO's private information, reflecting a belief within the accounting profession that (i) executive management exercises considerable discretion in how they report on the firm's financial standing and (ii) nevertheless, accounting reports do indeed convey information about the firm.[8] A fourth key feature of the framework we develop is that optimal contracting can be implemented as a sequence of short-term contracts, a property consistent with the claim by Holmstrom and Kaplan (2003) that corporate governance in the United States of America reacts quickly to legislative innovation.

[7]See Murphy (1999, 2012) for empirical surveys of managerial compensation.

[8]In principle, it is possible to treat hidden information as part of a pure moral hazard model, either because shareholders deter managers from misreporting by verifying their reports or because shareholders do not fully optimize over the contract space. See, for example, Peng and Röell (2008). However, Gayle and Miller (2015) find estimates of a pure moral hazard model that incorporates accounting information yield counter-intuitive results. For example, the model implies that in bad accounting states managers would be willing to pay shareholders to be employed.

The optimality of short-term contracts implies there is no adjustment period between adjacent regimes regulating governance, a hypothesis we test to check the robustness of our findings.

Our analysis is organized as follows. Section 2 describes the dataset, constructed from financial and accounting returns plus CEO compensation of firms in the Standard and Poor's (S&P) 1500 index. As a precursor to the main analysis, Section 3 tests whether the distribution of financial returns and the distribution of CEO compensation both changed in a statistically significant sense after SOX was introduced. The results from the nonparametric tests we develop show that this is indeed the case. These twin findings motivate our model of CEO compensation, presented in Section 4, a dynamic model of optimal contracting between a risk-neutral principal (the shareholders) and a risk-averse agent (the CEO) where there are both hidden actions (of the CEO) and hidden information (about the future prospects of the firm) when at the aggregate level, interest rates vary over time. Section 5 defines the welfare measures used to evaluate SOX. In Section 6, we explain the equilibrium for the model and show how it is related to the welfare measures we wish to compute. Identification and estimation are discussed in Section 7. In Section 8, we report our structural estimates of the welfare costs and summarises our main findings in Section 9.

2 Data

Financial and accounting data on the S&P 1500 were extracted from Compustat, whereas data on executive compensation were taken from ExecuComp. Bond prices were constructed from the yield curve using data from the Federal Reserve Economic dataset. Supplementary Appendix A explains how the data were assembled. For the purposes of the study, we classified each firm in the S&P 1500 over the 13-year period 1993 through 2005 into one of three sectors: primary, consumer goods, and service. Figure 1 charts the timeline of reform following the bankruptcy of Enron in December 2001 through the passage of SOX in mid-2002 to the approval by the SEC of the proposals of the New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotations (NASDAQ) in November 2003 following internal reviews of their respective corporate governance requirements. To mitigate contamination of our results with events that occurred in this period, in the main text we omit data on the two years when legislation was in a flux (2002 and 2003) by defining the pre-SOX era as the years 1993 through 2001 and the post-SOX era as the years 2004 and 2005.

The top panel of Table 1 displays summary measures of assets, capital structure, and accounting returns by sector. Average total assets on the balance sheet of firm n at the end of annual period t, denoted by A_{nt}, are reported before and after SOX in the first two columns for each sector, along with their standard deviations. The third column for each sector shows the t-(for means) or F-statistic (for standard deviations) of the change between the two eras. Firms in the primary and consumer goods sectors are of comparable size, whereas those in the service sector are on average about four times as large but exhibit greater size variation. On average, A_{nt} grew significantly in every sector by roughly one-third, the most in the primary sector, and so did its dispersion (as measured by the standard deviation). We define the debt-to-equity ratio by $C_{nt} \equiv D_{nt}/\left(A_{nt} - D_{nt}\right)$, where D_{nt} denotes debt at the end of the period. Average C_{nt} is almost twice as large in the service sector as the other two, but there is no discernible common trend across sectors for the pre-and post-SOX eras. Accounting returns are defined by $r_{nt} \equiv \left(A_{nt} - D_{nt} + I_{nt}\right)/\left(A_{n,t-1} - D_{n,t-1}\right)$, where I_{nt} denotes the total value of dividends (and stock repurchases) paid throughout the preceding financial period. The dispersion of accounting returns declined in all three sectors after SOX, which is curious because executive management exerts considerable discretion when reporting accounting earnings.[9]

The bottom panel of Table 1 displays average compensation and their standard deviations for the pre-SOX and post-SOX eras by firm type, further partitioning them by accounting state, along with the t- or F-statistics for testing a change between the two eras (as in the top panel, in the third column). To facilitate comparisons of pre- and post-SOX compensation on total expected CEO compensation, all firm-year observations are grouped according to how they fit within the pre-SOX population of firm-years. Specifically, we classify each firm by whether its total assets averaged in the pre-SOX era were less than or greater than the median of the averaged total assets for firms in the same sector and whether its debt–to-equity ratio averaged in the pre-SOX era was less than or greater than the median of the averaged debt–to-

[9]For example, the reporting of accruals, defined as the difference between realized cash flow and reported earnings, is one area where management may exercise considerable discretion. See Table A1 for more details on these statistics.

equity ratio for firms in the same sector in the pre-SOX era. Therefore, firm type is measured by the coordinate pair (A, C) with each corresponding to whether that element is above (L) or below (S) its median of that industry in the pre-SOX era. For example, (S, L) denotes lower total assets and a higher debt–to-equity ratio than the median total assets and debt–to-equity ratio for firms in that sector. Likewise, Bad_{nt} means the accounting return r_{nt} is lower than the average for all firms within the same sector, size, and capital structure categories, and $Good_{nt}$ means the reverse. Following Antle and Smith (1985, 1986), Hall and Liebman (1998), and Margiotta and Miller (2000), our measure of total compensation includes salary, bonus, options, promised retirement benefits and restricted stocks, as well as the change in wealth attributable to holding financial securities in the firm rather than a fully diversified portfolio. In this way, executive compensation depends directly on the excess returns of the firms they manage, which we denote by $x_{nt} \equiv \pi_{nt} - \pi_t$, where π_{nt} denotes financial returns on equity in firm n at t, and π_t is the financial return from holding the market portfolio. On average, CEO compensation is highest in the service sector and lowest in the primary sector.

Estimated mean CEO compensation significantly increased in all firm types within the primary sector after the SOX legislation was introduced when conditioning on accounting state. However, with one exception, estimated mean CEO compensation did not change significantly in the other two sectors. The dispersion of compensation, as measured by its standard deviation, fell in 14 of the 24 sectors and did not change significantly in the remaining 10. Broadly speaking, SOX compressed managerial compensation. Table 1 also shows that accounting states matter: CEO compensation depends in part on what they themselves report, conditional on firm type. This is evident in two respects. Controlling for firm characteristics, average compensation is substantially lower in bad states than that in good states; these states are in part defined by CEOs exercising their considerable discretion about reporting unverifiable events. Moreover, compensation exhibits more variation in good states than bad states, as measured by their estimated standard deviations.

3 Testing for Structural Change

Mean CEO total compensation in every type of firm classification within the primary sector significantly increased after SOX was introduced, but it did not significantly change in any of the firm types in the other sectors (with one exceptions: service (L, S)). However, this does not imply CEO compensation in the consumer goods and service sectors was unaffected by SOX. CEO compensation depends on excess returns. Therefore, a structural shift occurs if the distribution of excess returns changes and/or the relationship between excess returns and CEO compensation changes. Here we test for equality, between the pre- and post-SOX eras, of the probability density functions for excess returns and shape of the compensation schedule.

Change in the distribution of excess returns Denote the set of 24 categorical variables (formed from 3 sectors, 2 firm sizes, 2 capital structures, and 2 accounting states) by Z, and let $f_{pre}(x_{nt}|z_{nt})$ denote the probability density function of excess returns in the pre-SOX era conditional on $z_{nt} \in Z$. Also define $f_{post}(x_{nt}|z_{nt})$ in a similar manner. Under the null hypothesis of no change, $f_{pre}(x|z) = f_{post}(x|z)$ for all $(x, z) \in \mathcal{R} \times Z$. Li and Racine (2007, page 363) propose a one-sided test for the null, in which the test statistic is asymptotically distributed standard normal. Panel A in Table 2 reports the test outcome for the 24 cases. (Supplementary Appendix B provides a detailed explanation of both tests conducted in this section.) Aside from the bad state of (S, L) in the consumer goods sector, the values of the statistic lie above the critical value of the 1 percent confidence level (2.33). Consequently, for practically all firm types in both accounting states, we reject the null hypothesis of no change in the excess returns density from the pre-SOX to post-SOX eras.

Change in the shape of the contract Let $w_{pre}(x_{nt}, z_{nt})$ denote CEO compensation as a function of (x_{nt}, z_{nt}) in the pre-SOX era and similarly define $w_{post}(x_{nt}, z_{nt})$ in the post-SOX era. A straightforward way of testing whether the two mappings are equal is to include an indicator variable for the post-SOX regime in nonparametric regressions of compensation on the excess return x_{nt} for each z_{nt}. The one-sided test of the null hypothesis of equality is asymptotically standard normal. Panel B in Table 2 reports the test statistics for a change in the shape of the compensation schedule for each of the 24 cases. In all but two cases, the value of the statistic exceeds 1.64, the 5 percent level, implying the null hypothesis of no change in the compensation contract shape is rejected. Moreover, in these two exceptions, Panel A shows we reject the null hypothesis that the excess returns density function was unaffected, which implies that the probability distribution of managerial compensation in those cases did change when SOX was implemented.

Illustrating the differences To convey some sense of what lies behind rejecting the null hypothesis of no change, Figure 2 shows how the shape of the excess returns probability density function and the estimated compensation schedule adjusts for small, low-leveraged firms in the consumer goods sector, controlling for the state of the firm (bad versus good) and the two eras (pre-SOX versus post-SOX). The two top panels show that in both states density for excess returns shifted to the right and became more concentrated about the mean after SOX. Comparing Panel A with B, mean returns are not surprisingly higher in the good state. The bottom panels show that in both eras the compensation schedule is steeper in the good state than the bad. In addition, both plots in the post-SOX era (Panel D) tend to be flatter than in the pre-SOX era (Panel C). The overall effect of concentrating the excess returns distribution and flattening the extremes of the compensation schedule is to reduce the dispersion of compensation between the pre- and post-SOX eras, as reported in Table 1.

4 Model

The results from the first test show SOX had an impact on excess returns in all three sectors, over and above a common displacement effect on the returns to all firms. Conducting the second test showed that executive compensation committees also reacted to the SOX changes. But these tests cannot be used to decide whether the reaction was simply in response to the new distribution of excess returns or whether CEO functions changed. Answering that question requires a model of CEO compensation incorporating information asymmetries between the CEO and the firm's shareholders, with primitives as parameters that might change with the implementation of SOX.

SOX was enacted as a reaction to a number of major corporate and accounting scandals, including those affecting Enron, Tyco International, Adelphia, Peregrine Systems, and WorldCom. Broadly speaking, SOX (i) required top management to individually certify the accuracy of financial information, (ii) increased the oversight role of boards of directors and the independence of the outside auditors who review the accuracy of corporate financial statements, and (iii) penalized fraudulent financial activities more severely than previously. The provisions of SOX make abundantly clear that its purpose is not to provide legal infrastructure undergirding long-term contracting, but rather to penalize executives who make statements that are falsified soon afterward. For example, the clawback provision of Section 304 referred to in the introduction applies to compensation received up to a year after the alleged offense. Thus, modeling the effects of SOX does not demand a long-term contracting framework but should certainly leave open the possibility that executive management might lie to shareholders about the state of the firm.

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 nonverifiable statement by the CEO, whose credibility depends on the incentives that determine his or her payoff as a function of what the CEO reports.

At the beginning of period t, the CEO is paid compensation denoted by w_t for work during the previous period, denominated in terms of period-t consumption units. The CEO makes 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 $r_t \in \{1, 2\}$[10], and its subsequent performance, measured by excess returns x_{t+1}, revealed at the beginning of the next period. We denote this mapping by $w_{rt}(x)$, where the subscript t designates 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 or not. 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 $\varphi_{st} \in (0, 1)$. 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 $s_t = 1$,

[10] $r_t = 1$ if the private state is bad and $r_t = 2$ if it is good.

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 $w_{2t}(x)$, in that case.

The CEO then makes an 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. There are two possibilities: to diligently pursue the shareholders objectives of value maximization by working, thus setting $l_{st2} = 1$, or to accept employment with the firm but follow the objectives he would pursue if he were paid a fixed wage by setting $l_{st1} = 1$, called shirking. Let $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$. Since leaving the firm, working and shirking are mutually exclusive activities, $l_{t0} + l_{st1} + l_{st2} = 1$.

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 $s_t \in \{1, 2\}$:

$$\int x f_{st}(x) g_{st}(x) \mathrm{d}x \equiv E_{st}[x g_{st}(x) < E_{st}[x] \equiv \int x f_{st}(x) \mathrm{d}x, \tag{1}$$

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

$$\lim_{x \to \infty} \left[g_{st}(x)\right] = 0 \tag{2}$$

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

$$\lim_{x \to \infty} \left[\varphi_{2t} f_{2t}(x) / \varphi_{1t} f_{1t}(x)\right] \equiv \lim_{x \to \infty} [h_t(x)] = \sup_{x \in R} [h_t(x)] \equiv \overline{h}_t < \infty. \tag{3}$$

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

$$-\sum_{t=0}^{\infty}\sum_{j=0}^{2}\beta^t\alpha_{jt}l_{tj}\exp\left(-\gamma_t c_t\right), \tag{4}$$

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 $\alpha_{0t} = 1$.

Finally, aggregate shocks in the model arise from fluctuations in the stock market index, from which abnormal returns are calculated, and through anticipated changes in bond prices. SOX was enacted at roughly the same time as the Omnibus Budget Reconciliation Act of 1993 (OBRA), which raised the income tax rate for high earners.[11] In our model making the marginal tax schedule steeper effectively increases the risk premium required to compensate a CEO for uncertain compensation.

5 Welfare measures

The catalyst for SOX was a failure in corporate governance that led to the dismissal of executives and in some cases, subsequent prosecution for fraud, conviction and imprisonment. These executives violated legal constraints that were subject to auditing. SOX was not confined to, or even primarily directed towards, realigning the incentives of law abiding managers. After its enactment, bringing greater accountability to financial statements, enforcing property rights in governance more rigorously, and increasing the penalties for fraud, might have reduced white collar crime. Implementing SOX changed firm value because of its effects on the willingness of managers to break the law, as well as its effects on the twin agency costs of motivating man-

[11] OBRA was introduced in the house May 1993 and put into effect August 1993. It created new tax brackets for individual income, raising the rate from top rate from 31 percent to 38 percent.

agers to act in the firm's interest rather engage in legal activities they prefer, and to accurately disseminate unverifiable financial information to the board. Our welfare analysis focuses on these agency issues, which are in turn intimately related to CEO compensation.[12]

Figure 3 is a schema for the welfare measures we investigate. Total expected compensation in the pre-SOX era, defined as $\sum_{s=1}^{2}\varphi_{st}E_{st}\left[w_{st}(x)\right]$, can be decomposed into administrative costs in the pre-SOX era, denoted by τ_{1t}, and agency costs, τ_{2t}. Agency costs are further divided into τ_{3t}, which arises from pure moral hazard or the costs of hidden actions, and τ_{4t}, the extra cost from hidden information when there is moral hazard. Changes in τ_{it} from the pre- to post-SOX eras are denoted by $\Delta\tau_{it}$. We now explain how each of these measures appears in our model and why SOX might affect their values.

Absent agency considerations, shareholders would pay the CEO in the pre-SOX era an amount τ_{1t}, which we interpret as an administrative wage to work for the firm instead of pursuing an outside option — in other words, the certainty equivalent of being employed as a CEO. These costs are broadly interpreted within our model and include the legal jeopardy executives were exposed to following the enactment of the legislation. Formally, $\tau_{1t} \equiv b_{t+1}\left[(b_t - 1)\,\gamma_t\right]^{-1}\ln\alpha_{2t}$, where b_t denotes the bond price in period t. SOX imposed additional responsibilities on executive management that make the job more onerous. For example, Section 302 of SOX holds the principal executive officer(s) and the principal financial officer(s) responsible for establishing and maintaining internal controls. Denoting the change in the administrative wage from the pre-SOX to post-SOX eras, by $\Delta\tau_{1t}$, the increased regulations lead us to speculate that $\Delta\tau_{1t} > 0$.

The difference between expected total compensation and the administrative wage a CEO would receive, τ_2 in the pre-SOX era, is the risk premium of accepting employment that pays uncertain compensation rather than a fixed wage, which shareholders pay because of agency problems, and as such represents the amount shareholders are willing to pay for perfect monitoring. We define $\tau_{2t} \equiv \sum_{s=1}^{2}\varphi_{st}E_{st}\left[w_{st}(x)\right] - \tau_{1t}$. The SOX provisions induced the firm to be more transparent about its future profitability. For example, Section 302 requires the principal executive officer(s) and the principal

[12] Chhaochharia and Grinstein (2007) conducted an empirical analysis of the effect of SOX on firm value, and found it increased the value of firms that were less compliant with SOX relative to those which more compliant.

financial officer(s) to certify in each annual or quarterly report filed or submitted that the financial statements and other financial information fairly present financial conditions and results and refrain from making misleading statements. Using legal machinery to enforce the truthful revelation of financial conditions may remove or ease the burden of the compensation committee in designing incentives that resolve the agency issues. For these reasons we might expect $\Delta\tau_{2t} < 0$.

The component of agency costs solely attributable to pure moral hazard — or the amount shareholders would pay to eliminate hidden action in the absence of private information — is the difference between the expected compensation in the pure moral hazard case, which for the pre-SOX era we denote by $y_{st}(x)$, and the certainty equivalent of being employed as a CEO. Thus, $\tau_{3t} \equiv \sum_{s=1}^{2} \varphi_{st} E_{st}\left[y_{st}(x)\right] - \tau_{1t}$. Because SOX increased the penalties associated with fraudulent financial reporting, we might predict that the benefits from shirking relative to working declined, and hence $\Delta\tau_{3t} < 0$. However, many of the SOX mandated requirements apply whether CEOs pursue their own objectives subject to their legal obligations or receive compensation that induces them to work in the interests of shareholders, and consequently would not affect the difference in utility from shirking versus working.

The component of agency costs solely attributable to private information — or the amount shareholders would pay to eliminate private information — is the difference between the expected compensation under the current optimal contract and expected compensation in the pure moral hazard case, which for the pre-SOX era we denote by $y_{st}(x)$. Thus, $\tau_{4t} \equiv \sum_{s=1}^{2} \varphi_{st} E_{st}\left[w_{st}(x) - y_{st}(x)\right]$. By construction $\tau_{4t} \equiv \tau_{2t} - \tau_{3t}$. If shareholders could observe the CEO's effort, then a first-best constant-wage contract would be paid regardless of whether the CEO had private information or not (Gayle and Miller, 2009b). Therefore, the only reason hidden information might be costly, meaning $\tau_{4t} > 0$, is that it exacerbates rather than ameliorates the pure moral hazard model, which we show is an empirical question as theory does not give a decisive answer. One purpose of SOX was to enhance the independence of auditors and boards conducting monitoring functions, making shareholders more informed, presumably to reduce the role of hidden information. For example, Section 304 of SOX requires the CEO and CFO to reimburse the firm for any compensation received during the 12-month period following equity issue filing if there was misconduct in filling a financial statement for that equity issue. This regulation makes CEO compensation less liquid and so can mitigate the CEO's incentives to take opportunistic advantage by misrepresenting financial states and hence enforcing the truthful revelation of financial condition. Consequently, we might expect $\Delta\tau_{4t} < 0$.

The remaining symbols, ρ_{1t} through ρ_{3t}, are summary measures of the channels SOX flowed through, changing the values of the primitives in our model to affect the welfare measures. Specifically, $\rho_{1t} \equiv E_{st}[x - xg_{st}(x)]$ is the loss shareholders would incur from a CEO shirking instead of working; to the extent SOX provided more protection to shareholders from shirking managers we would predict that $\Delta\rho_{1t} < 0$. The difference between a CEO's pecuniary cost of working and that of shirking is measured by $\rho_{2t} \equiv b_{t+1}\left[(b_t - 1)\gamma_t\right]^{-1}\ln(\alpha_{2t}/\alpha_{1t})$. If the penalties imposed by SOX diminished the incentives to shirk without imposing administrative burdens on working managers, then $\rho_{2t} < 0$. Finally, ρ_{3t} measures how much the loss from pure moral hazard changed because of the shift in the signal. It's change is hard to predict because the signal $g_{st}(x)$ depends on the likelihood, and hence the distribution of abnormal returns, when the manager shirks versus works.

6 Equilibrium

The welfare measures, τ_{1t} through τ_{4t}, and the summary measures of the driving forces of the agency problem, ρ_{1t} through ρ_{3t}, are functions of the utility parameters and the parameters determining the distribution of excess returns. Yet the state s and a sample analog to φ_{st} (the probability of each state) are not directly observed, and the parameters defining utility, γ_t, α_{1t}, and α_{2t}, and the likelihood ratio $g_{st}(x)$ cannot be estimated for either state $s \in \{1, 2\}$ without making behavioral assumptions about shareholders and CEOs. We now assume shareholders have diversified portfolios and are expected value maximizers, whereas CEOs are expected utility maximizers.

6.1 Optimization

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. (Both lemmas in this section are proved in the Appendix.)

Lemma 1 *Denote by $\overline{\varsigma}$ the date the CEO retires. The optimal long-term contract can be implemented by a $\overline{\varsigma}$-period replication of the optimal short-term contract.*

The next lemma solves the optimal consumption and savings plan for a CEO about to retire. It proves that in our model, 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 the bond prices (b_t, b_{t+1}). However, the employment and effort choices do not depend on his current (outside) wealth. To state the lemma, let $r_t(s)$ denote the CEO's disclosure rule about the state when the true state is $s_t \in \{1, 2\}$.

Lemma 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*

$$\sum_{s=1}^{2} \varphi_{st} \left\{ l_{t0} + (\alpha_{1t} l_{st1} + \alpha_{2t} l_{st2})^{1/(b_t - 1)} E_{st} \left[\exp\left(-\frac{\gamma_t w_{r_t(s)t}(x)}{b_{t+1}} \right) [g_{st}(x) l_{st1} + l_{st2}] \right] \right\}. \tag{5}$$

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

$$v_{st}(x) \equiv \exp\left(-\gamma_t w_{st}(x) / b_{t+1}\right). \tag{6}$$

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 (5) and substituting in $v_{st}(x)$, the participation constraint is thus

$$\sum_{s=1}^{2} \int \varphi_{st} v_{st}(x) f_{st}(x) \mathrm{d}x \leq \alpha_{2t}^{-1/(b_t - 1)}. \tag{7}$$

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 (5) 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:

$$0 \leq \int \left(g_{st}(x) - (\alpha_{2t}/\alpha_{1t})^{1/(b_t-1)} \right) v_{st}(x) f_{st}(x) \mathrm{d}x. \tag{8}$$

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:

$$0 \leq \int \left[v_{1t}(x) - v_{2t}(x) \right] f_{2t}(x) \mathrm{d}x. \tag{9}$$

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

$$0 \leq \int \left[(\alpha_{1t}/\alpha_{2t})^{\frac{1}{b_t-1}} v_{1t}(x) g_{2t}(x) - v_{2t}(x) \right] f_{2t}(x) \mathrm{d}x, \tag{10}$$

where $(\alpha_{1t}/\alpha_{2t})^{1/(b_t-1)} v_{1t}(x)$ is proportional to the utility obtained from shirking and announcing truthfully in 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

$$\textstyle\sum_{s=1}^{2} \int \varphi_{st} \ln \left[v_{st}(x) \right] f_{st}(x) \mathrm{d}x. \tag{11}$$

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 their 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.

6.2 Comparing the pure and hybrid model contracts

The optimal contract for a parameterization of the hybrid model is plotted in the left panel of Figure 4. 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, the right panel plots the optimal compensation for the analogous two-state pure moral hazard model (where there are hidden actions but the state is known), denoted by $y_{st}(x)$. Details explaining the solution and computation of $y_{st}(x)$ are provided in supplementary Appendix C.

Figure 4 illustrates four important features. As compensation in both models is a function of the likelihood ratio between the densities of the excess return for working and shirking, not the excess return itself, 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, 0, 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 greater everywhere in the good state than the bad, whereas in the pure moral hazard model the slope in the bad state is greater than in the good over the intermediate range where much of the probability mass of both excess return distributions lies. Thus, the point where the schedules cross is higher in the pure moral hazard model than in the hybrid model. Figure 4 also illustrates two analytical results: In the hybrid model, expected utility of the agent is greater in the good state than the bad, 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

[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, and otherwise paying the agent the first-best fixed wage. See Mirrlees (1975).

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.[14] Indeed, our parameterization illustrates an instance where the agency cost in the pure moral hazard model is greater than in its hybrid counterpart. The parameterization demonstrates a paradox: To the extent it succeeds in making financial disclosure more transparent, SOX may have perverse consequences in some sectors.

7 Identification and estimation

The parameters defining the model are characterized by $f_{st}(x)$ and $g_{st}(x)$ for $s_t \in \{1, 2\}$, which together define the probability density functions for revenue in each state, and φ_{st}, the probability of each state occurring. CEO preferences are defined, relative to the normalized utility from taking the outside option, by their distaste for working, α_{2t}, and shirking, α_{1t}, as well as their risk aversion parameter, γ_t. In equilibrium, CEOs truthfully reveal the state, implying $s_t = r_t(s)$, so $r_t = s_t$ is observed in the data. Hence, the reported state is sampled from a Bernoulli distribution with parameter φ_{st}, and the data on returns are generated by $f_{st}(x)$, implying those parameters are identified, the latter non-parametrically, along with $h_t(x)$. Aside from observing returns from working, we assume that compensation, $w_{st}(x)$, is also observed for different values of (x, s). This leaves only $g_{st}(x)$ plus $(\alpha_{1t}, \alpha_{2t}, \gamma_t)$ to identify from the first-order conditions, the complementary slackness conditions, plus a constraint that working is an optimal choice, from observations on (x_{nt}, s_{nt}, w_{nt}) generated from the CEO working.[15] This section explains the intuition supporting identification and estimation; the supporting technical details are relegated to supplementary Appendix D.

We motivate the identification of this model by comparing the equilibrium compensation schedule and the excess return density shown in Figure 4 with the sample

[14] There are assumptions guaranteeing expected compensation in the hybrid model is more expensive than in the pure moral hazard model. For example, if the two distributions for the good state are simply a shift of the distributions in the bad state by a constant amount to the right, the optimal contract in the pure moral hazard model depends only on the state through the translation parameter and is therefore cheaper than the optimal contract of the hybrid model.

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

estimates displayed Figure 2. Both the theoretical and the empirically estimated compensation schedules vary with excess returns, for the most part increasing, and flatten at very high rates of excess returns. These features illustrate the agency problem. Moreover, in the estimated schedules for both states and in the hybrid model, but not in the model of pure moral hazard, the schedules for the good state are everywhere steeper than for the bad state and also cross at negative excess returns. This suggests that hidden information, not just hidden actions, may be a part of the agency problem. Following Gayle and Miller (2015), we can separate the analysis of identification into two pieces: given $f_{st}(x)$ representing $g_{st}(x)$ and $(\alpha_{1t}, \alpha_{2t})$ as mappings of γ_t, and identifying the observationally equivalent values of γ_t. Estimation proceeds by forming a sample analog of the identified sets.

7.1 Mapping risk preferences into the remaining parameters

Extending the results of the static framework of Gayle and Miller (2015) to our dynamic setting, it follows directly from the first-order condition of the compensation contract, the participation and incentive compatibility constraints (both of which are binding), and the regularity conditions for $g_{st}(x)$ and $h_t(x)$ that for every period t;

$$\begin{aligned}
\alpha_{2t} &= E\left[v_{st}(x,\gamma_t)\right]^{1-b_t} \\
\alpha_{1t} &= \alpha_{2t}\left\{\overline{v}_{2t}(\gamma)^{-1} - E_{2t}\left[v_{2t}(x,\gamma_t)^{\ -1}\right]\right\}^{b_t-1}\left\{\overline{v}_{2t}(\gamma)^{-1} - E_{2t}\left[v_{2t}(x,\gamma_t)\right]^{-1}\right\}^{1-b_t} \\
g_{2t}(x) &= \frac{\overline{v}_{2t}(\gamma_t)^{-1} - v_{2t}(x,\gamma_t)^{-1}}{\overline{v}_{2t}(\gamma_t)^{-1} - E_{2t}\left[v_{2t}(x,\gamma_t)^{-1}\right]} \\
g_{1t}(x) &= \frac{\left(\frac{\alpha_{2t}}{\alpha_{1t}}\right)^{1-b_t}\left\{\overline{v}_{1t}(\gamma_t)^{-1} - v_{1t}(x,\gamma_t)^{-1} + \eta_{3t}\left[\overline{h}_t - h_t(x)\right]\right\} - \eta_{4t} g_{2t}(x) h_t(x)}{\overline{v}_{1t}(\gamma_t)^{-1} - E_t\left[v_{st}(x,\gamma_t)\right]^{-1} + \eta_{3t}\overline{h}_t},
\end{aligned} \tag{12}$$

where $\overline{v}_{st}(\gamma_t) \equiv \lim v_{st}(x, \gamma_t)$ as $x \to \infty$ and the Kuhn-Tucker multipliers η_{4t} and η_{3t} have the following representation[16]:

$$\eta_{4t} = \frac{\frac{E_{1t}[v_{1t}(x,\gamma_t)]}{E_t[v_{st}(x,\gamma_t)]} - E_{1t}\left[v_{1t}(x,\gamma_t)h_t(x)\right]\left\{E_{2t}\left[v_{2t}(x,\gamma_t)\right]^{-1} - E_t\left[v_{st}(x,\gamma_t)\right]^{-1}\right\} - 1}{\left(\frac{\alpha_{2t}}{\alpha_{1t}}\right)^{1-b_t} E_{1t}\left[v_{1t}(x,\gamma_t)g_{2t}(x)h_t(x)\right] - E_{1t}\left[v_{1t}(x,\gamma_t)h_t(x)\right]}$$

$$\eta_{3t} = E_{2t}\left[v_{2t}(x,\gamma_t)\right]^{-1} - \eta_{4t} - E_t\left[v_{st}(x,\gamma_t)\right]^{-1}. \tag{13}$$

Combining the nonparametric estimates of the density of excess return and relationship between excess return and compensation presented in Figure 2 with the formulas for the structural parameters in equation (12), one can glean the sources of variation in the data that identify the other structural parameters given a known γ_t. The first three equations fully apply to a pure moral hazard model in which the good state occurs with probability 1. First, α_{2t} is identified from the exante expected discounted utility derived from the compensation schedule.[17] The identification of α_{1t}/α_{2t}, and hence α_{1t}, comes from the concavity of the compensation schedule relative to the maximum compensation in the good state. The likelihood ratio in the (verifiable) good state, $g_{2t}(\cdot)$, is identified from the slope of the compensation schedule in the good state. The last equation identifies $g_{1t}(\cdot)$ given all the other parameters in the model. As with $g_{2t}(x)$, it also depends on the slope of the compensation schedule in the same state (unverifiable in this case), but $g_{1t}(x)$ also depends on the slope of the compensation schedule in the other state, as well as $h_t(x)$, the likelihood ratio of either state given x.

7.2 Set identification

Our model fully accounts for aggregate fluctuations through the volatility of bond prices, which in turn provides a source of identifying information about the risk-aversion parameter. To demonstrate this point, suppose that $b_t \neq b_{t'}$ for periods t and t', but that both periods fall within the same regime (pre-SOX or post-SOX), implying from our exclusion restrictions that $(\gamma_t, \alpha_{2t}) = (\gamma_{t'}, \alpha_{2t'})$. Differencing out α_{2t} in the first equation in (12) and taking logarithms we obtain $(1 - b_t)\log\{E\left[v_{st}(x, \gamma_t)\right]\} = (1 - b_{t'})\log\{E\left[v_{st}(x, \gamma_t)\right]\}$. Given data on compensation, the firm's state, and excess

[16] η_{3t} corresponds to the truth-telling constraint and η_{4t} corresponds to the sincerity constraint.

[17] This expectation is taken before the realization of the hidden information states to the manager.

returns, the solution(s) to γ_t yield a set of identifying restrictions. Because this equation is nonlinear in γ_t, there is no guarantee it has a unique solution, thereby ruling out a strict application of, but not the intuitive connection with, a difference-in-differences estimator.

There are also cross-sectional restrictions, in the form of equalities and inequalities implied by the model, that can be used to obtain bounds for admissible values of γ_t. At least one of the truth-telling constraints and the sincerity constraint bind. The three other Kuhn-Tucker multipliers are non-negative. Similarly, the complementary slackness conditions for the truth-telling and sincerity constraints yield two more equalities. We impose an exclusion restriction that α_{1t} does not depend on the private states. The likelihood for the bad state, $g_{1t}(\cdot)$, is positive with unit mass.[18] Value maximization implies another three inequalities reflecting that shareholders prefer the CEO working in both private states rather than shirking in either or both of them. Finally, we impose the restriction that the risk aversion does not depend on bond price. Formally, Gayle and Miller (2015) obtained sharp and tight bounds for the set of observational equivalent risk aversion parameters, and we can adopt their methods and results to our framework. Accordingly, let $\Gamma_t \equiv \{\gamma_t : Q_t(\gamma) = 0\}$ denote the Borel set of admissible values of γ_t for data generated in period t, where $Q_t(\gamma_t)$ is a quadratic form of the minus norm of equalities and inequalities implied by the model.[19] Thus, Γ_t denotes all the values of γ that are observationally equivalent given the probability distributions generating the data at period t. Imposing additional restrictions that arise from multiple time periods is straightforward. For example, if γ is time invariant across the first two periods, labeled 1 and 2, then the set of admissible risk aversion parameters is the intersection $\Gamma_1 \bigcap \Gamma_2$. We denote by Γ the identified set that arises from imposing all the relevant restrictions for the different time periods and its quadratic from by $Q(\gamma)$.

As γ is not identified pointwise but only up to the set Γ, it follows that the other taste parameters, the likelihood ratios, and the measures of agency cost are also only set identified. For example, we can write $\alpha_{1t}(\gamma)$ as the value of α_1 identified from data generated at t by the hybrid model when the agent's risk aversion parameter is γ. Thus, α_1 is identified up to the set $\{\alpha_{1t}(\gamma) : \gamma \in \Gamma\}$. Further restrictions obtained

[18] One can show that $g_{2t}(\cdot)$ is positive without imposing any restrictions on γ_t. See Gayle and Miller (2015).

[19] The minus norm of q, denoted $\|q\|_{-}$, is the norm of the maximun of $-q$ and 0, that is, $\|q\|_{-} = \|\max(-q, 0)\|$.

from the panel are imposed in the same way that admissible values of γ are restricted.

7.3 Estimation

We estimated a confidence region for Γ by exploiting the fact that approximations to $Q(\gamma)$ formed from the data deviate from 0 only because of differences between expectations (or, in some equations, population limits) and their sample analogs. Accordingly, let $Q^{(NT)}(\gamma)$ denote a sample analog to $Q(\gamma)$ and define $\Gamma^{(NT)} \equiv \left\{\gamma : Q^{(NT)}(\gamma) \leq c_{0.95}\right\}$, where $c_{0.95}$ is the critical value, below which $Q^{(NT)}(\gamma)$ falls 95 percent of the time under the null hypothesis that $\gamma \in \Gamma$. Once $\Gamma^{(NT)}$ has been numerically determined (by subsampling in our application), we can deduce that the estimated confidence region for $\{\alpha_{1t}(\gamma) : \gamma \in \Gamma\}$, for example, is $\left\{\alpha_{1t}(\gamma) : \gamma \in \Gamma^{(NT)}\right\}$. Estimated confidence regions for the other primitive parameters and the welfare measures are derived by following the same procedure.

8 Empirical findings from the model

Here we report estimates obtained by omitting the two years bordering on the SOX legislation, namely, 1993 through 2001 (16,894 observations) and 2004 through 2005 (3,781 observations). To account for heterogeneity in the data, the estimation is also conditional on the firm type (defined by sector, assets, and capital structure). As a robustness check, the supplementary Appendix reports estimates for the extended sample covering 1993 through 2002 for the pre-SOX era (containing 18,855 observations) and 2003 though 2005 for the post-SOX era (5,670 observations). The differences are minor, suggesting that a precise determination of the cutoff dates for the two regimes is empirically unimportant.

We did not reject the null hypothesis that φ_{st}, $f_{st}(x)$, α_{1t}, α_{2t}, and γ_t are time invariant within each era. Thus, bond prices and the stock market index (which differences out in our model) are sufficient to capture all aggregate variation within each era. To achieve comparability between the two eras we estimated the confidence region for each of the two bond prices that occurred in both eras, 16.4 and 16.8. This permits us to attribute all changes in social welfare costs to the changes in primitives rather than aggregate factors in the macroeconomy. Because the differences between the two sets of estimates are negligible, we report only those for 16.4.

Risk preferences The procedure used to obtain a confidence region for changes in the welfare measures depends somewhat on whether the risk aversion parameter γ changes between the pre- and post-SOX eras. If the null hypothesis — that the risk parameter was constant over this period — is maintained, then it is straightforward to compute confidence regions for $\Delta\tau_{it}$ for $i \in \{1, \ldots, 4\}$ by substituting those values of $\gamma_t \in \Gamma^{(NT)}$ into the formulas for $\Delta\tau_{it}$ (that come from evaluating τ_{it} in the pre- and post-SOX eras). If the confidence regions for $\Delta\tau_{it}$ contain only positive (negative) values, then we can reject the null hypothesis that the risk parameter was constant over the two periods. On the other hand, if the null hypothesis is rejected, the confidence region computed for $\Delta\tau_{it}$ is based on admissible values of two parameters γ_{pre} and γ_{post}, not just one, and a further component in the decomposition is introduced, which measures the contribution of the change in risk attitude to $\Delta\tau_{it}$.

For these reasons we first tested whether the null hypothesis — that risk aversion did not change — is rejected; only then did we construct the appropriate confidence regions for the welfare measures. The test is of independent interest. One concern raised by directors (Cohen et al., 2013) and bankers such as Alan Greenspan and William Donaldson (former SEC chairman) is that CEOs would overreact to SOX provisions and exercise undue caution in investment decisions, thus destroying shareholder value (see Coats and Srinivasan, 2014). Another concern was the possibility that OBRA might contaminate our analysis by having a significant effect on estimated risk preferences through increased income taxation at the upper levels.

The 95 percent confidence region of the risk aversion parameter for the (main) sample is common to both periods (0.0695, 0.6158); every observationally equivalent risk aversion parameter for one regime appears in the confidence region for the other.[20] Therefore, we do not reject the null hypothesis that no significant change in risk attitude occurred after SOX.[21]

To give economic meaning to our estimates of risk aversion, we also computed the

[20] The confidence region for the full sample of the post-SOX period covers a wider range (0.0616, 0.2335) than that of the pre-SOX period (0.0784, 0.2335), and the confidence region for risk aversion parameters for both periods in the full sample is a proper subset of the corresponding region in the restricted sample. Thus, adding the restrictions for the years 2002 and 2003 to the sample yields more precise results. See Table D2 in supplementary Appendix D for a more detailed report of these findings.

[21] These findings contrast with those of Nekipelov (2010), who finds the risk aversion of top executives in the retail apparel industry significantly increased after SOX was introduced. Three notable differences between his work and ours is that Nekipelov assumes the contract is linear, approximates compensation by salary and bonus, and, of course, estimates from a different sample population.

amount a CEO would pay to avoid an equiprobable gamble with losing or winning \$1,000,000. At the right boundary of the confidence region for the main sample in the pre-SOX period, the risk aversion parameter is 0.6158, implying the CEO would pay \$290,206 to avoid the gamble, but at the left boundary of 0.0695 would pay only \$34,722.[22] We conclude that changes to the distribution of excess returns and its mapping to CEO compensation documented in Section 3 did not arise because the implementing SOX legislation induced CEOs to think differently about risk and that OBRA did not have a significant effect. Other factors in our model caused the changes. We now investigate these other factors under the maintained hypothesis that the risk aversion parameter was constant over the entire sample period.

Administrative costs Administrative costs, denoted by τ_{1t}, are the premium that a CEO would be paid over the inclusive annuitized value of his outside option if there were no agency problems. The third column of Table 3 shows that these vary greatly by sector and firm type, but most of the variation is explained by the firm categories. For example, in the pre-SOX regime, the 95 percent confidence region for the administrative cost of (S, L) firms in the primary sector is covered by the interval ranging from \$0.9 to \$1.0 million. In (L, S) firms in the service sector, the corresponding region is covered by the interval ranging from \$7.9 to \$11.0 million. In both samples we cannot reject the hypothesis that $\triangle\tau_{1t} > 0$ in at least four of the categories and that $\triangle\tau_{1t} < 0$ in at least four. In both samples every firm category within the primary sector experienced increased administrative costs of between \$2.3 and \$4.6 million in the main sample. Our estimates from the main sample show that every category within the service sector experienced declines between \$0.5 and \$4.1 million. Both results broadly reflect our findings in Table 1, which shows that mean CEO compensation significantly increased in every subcategory within the primary sector following passage of SOX, but did not significantly increase in any subcategory of the service sector.

Agency costs Agency costs, τ_{2t}, measure the gross costs that shareholders would be willing to pay for perfect monitoring and thus avoid the penalties induced by the incentive compatibility and truth-telling constraints. Table 4 reports the 95 percent

[22] These estimates are in line with previously published work. See Gayle and Miller (2009a, 2009b, 2015) and Gayle et al. (2015)

confidence region for the observational equivalent values of τ_{2t} in the pre-SOX period and its change $\triangle\tau_{2t}$. Agency costs are small in some firm categories, as low as $22,000 per year in (S, L) firms within the primary sector, but within the service sector, these costs are much greater: between $105,000 and $3.425 million. SOX increased agency costs in 10 of 12 firm categories. For the most part, the absolute values of the changes are small to moderate, exceeding 1 million dollars only in the (L, S) consumer goods category. However, as a proportion of the levels, they are quite substantial; the estimated upper bound on $\triangle\tau_{2t}$ is at least as large as the lower bound of τ_{2t} in several categories.

Cost of hidden actions Shareholders' costs due to hidden actions, denoted by τ_{3t}, are the difference between the expected compensation that would have been paid if there were only a pure moral hazard problem and the certainty equivalent wage if CEOs could be perfectly monitored. Table 5 reports the estimated 95 percent confidence region for τ_{3t} in the pre-SOX era and the change it heralded, $\triangle\tau_{3t}$. The estimated bounds of the confidence intervals for the pre-SOX era range between $6,000 and $9.0 million depending on firm type, markedly lower in the primary sector than the other two. The effects of SOX on the estimated cost of hidden actions vary by sector and firm type; all the increases occur within the primary sector.

Cost of hidden information Recall that the costs of hidden information, τ_{4t}, are the difference between the expected compensation and what they would have been with hidden actions but not hidden information about the firm's state. Table 6 displays a property foreshadowed in Figure 2. With a single exception, hidden information ameliorates pure moral hazard. In these cases, adding truth-telling and sincerity constraints to the principal's optimization problem is less costly than adding the extra participation constraint that would arise in a pure moral hazard model. The net benefits range from -$29,000 in (L, L) firms in the primary sector to $6.7 million in (L, L) firms in the service sector. In the consumer goods sector there is evidence that $\triangle\tau_{4t} > 0$: The welfare costs of hidden information increased after SOX was introduced. Overall, the null hypothesis of no change is not rejected in half of the firm categories. Therefore, if anything, the clawback provision in Section 304 of the SOX made the hidden information problem more severe.

Gross loss to shareholders from CEO shirking The sources of the agency problem arise from conflicting objectives between the shareholders (who want the CEO to work to maximize their returns) and the CEO (who prefers to shirk). The expected gross loss to shareholders that would occur each year from the CEO shirking instead of working, denoted by ρ_{1t}, is reported as a percentage of market value in Table 7. Similar to estimates found in previous studies, they range from 5.0 to 20.2 percent per year.[23] As in previous tables, the variation explained by firm category far outweighs the indeterminacy from observational equivalence that arises from set, rather than point, identification. In particular, large firms tend to lose proportionately less than small firms from a shirking CEO. Given size and leverage, shareholders owning firms in the service sector have the most to lose when management objectives do no align with their own.

The most striking new result in Table 7 is that in both samples $\triangle\rho_{1t} < 0$ for 11 of the 12 categories. Overall, the effect of SOX was to limit the expected losses a CEO would impose on his firm by not pursuing a goal of expected value maximization.

Benefit to CEO from shirking The other side of the conflict driving the agency problem is the compensating differential the CEO is paid to work rather than shirk, measured by ρ_{2t}. Table 8 shows our estimates are a tiny fraction of the losses shareholders would incur with shirking, ranging between \$1.1 million and \$10.8 million annually. The conflict of interest faced by CEOs declines in half of the firm categories but is exacerbated in the remainder. The main new finding Table 8 is that when SOX was introduced, the differential mostly declined in the firm categories where it was relatively high. Thus, SOX had an equalizing effect on the working versus shirking compensating differential, rendering shirking a more homogeneous activity that does not depend as much on firm type. To some extent, then, greater regulation of management imposed by SOX, including the attendant legal responsibilities, enforcement, and penalties, channeled the type of shirking that occurs if and when CEOs lack proper incentives.

Signal quality The reason the different objectives cannot be resolved by fiat is that the signal shareholders use to evaluate the actions of the CEO, excess returns, is an

[23] A variety of different models and different estimators corroborate these estimates; see Margiotta and Miller (2000), Gayle and Miller (2009a, 2009b, 2015), and Gayle et al. (2015).

imperfect measure of CEO effort. To complete our empirical analysis we investigated whether SOX improved the signal. That is, holding the nonpecuniary benefits of the CEO constant at their pre-SOX level, what effect does changing the signal from the pre-SOX to the post-SOX regime have on the cost of pure moral hazard? Table 9 reports our results of ρ_{3t}. The first set of results shows the effect of substituting α_{1post} and α_{2post} for their pre-SOX values in the construction of τ_{3t}. In 5 of the 12 firm categories the change in the signal quality increases the cost of moral hazard; in 4 of the 12 firm categories it does not change the cost of moral hazard; and in the rest the change in signal quality reduces the cost of moral hazard. In the primary sector, the increase in the cost of moral hazard from a change in the quality of the signal tends to reinforce the increase in the cost of moral hazard from the increase in the benefit to the CEO from shirking.

9 Conclusion

SOX was a legislative response by the U.S. government to corporate governance failures at many prominent companies. This article describes an empirical analysis of its effects on CEO compensation using panel data constructed for the S&P 1500 firms on CEO compensation, financial returns, and reported accounting income. Our structural empirical analysis is motivated by the empirical facts that after SOX was enacted, there were significant changes in (i) the relation between a firm's excess returns and CEO compensation and (ii) the underlying distribution of excess returns. The net effect of these changes was to significantly raise expected CEO compensation in the primary sector but not in the consumer goods and service sectors. A third empirical regularity motivating our study is that, both before and after SOX, conditional on issuing a favorable accounting statement, CEOs receive compensation that is on average higher but also more volatile.

We develop a dynamic principal agent model to explain why this occurs. Each period a CEO agent has private information about his firm and takes hidden actions, neither of which is observed by the shareholder principal. In the model, accounting disclosures are treated as unverifiable discretionary messages sent by the agent to the principal about the state of the firm at the beginning of the period. Our data show that compensation practices quickly adapted to the new regulations, and our model reflects this feature: the optimal long-term contract can be implemented by a sequence of short-term contracts. The optimal contract does not base compensation on excess returns alone (as in a pure moral hazard model) but also incorporates accounting disclosures. In equilibrium, expected compensation is higher in the good accounting state than the bad one, and there is also greater variation in compensation outcomes in the good state — which are two of the empirical regularities mentioned above. In the model, CEOs are paid to reveal the good state with the promise of receiving very high compensation if the firm produces abnormally high returns. This prediction contrasts with those of a pure moral hazard model, which does not predict either empirical regularity.

We identify and estimate the model using data on compensation, excess returns, and accounting disclosure, controlling for different firm categories and aggregate conditions. The risk aversion parameter of the agent in our model is set identified, and the remaining parameters of the model are identified up to the value of the risk aversion parameter for each of the firm categories. Even though our model is semiparametric and does not impose sufficient functional form assumptions to achieve point identification, we find much of the variation in the data is explained by the primitives of the model and the returns process.

In summary, four main conclusions emerge from estimating the structural model. First, variation in our data can be accounted for without resorting to an explanation based on changing tastes. We do not find evidence that the preference for risk-taking by CEOs changed with SOX, contradicting concerns raised by directors (Cohen et al., 2013) and politicians such as Alan Greenspan and William Donaldson (former SEC chairman) that CEOs would overreact to provisions in SOX provisions and exercise undue caution in investment decisions, thus destroying shareholder value (see Coats and Srinivasan, 2014).

Second, the main impact of SOX was to increase the administrative burden of compliance in the primary sector but reduce this burden in the service sector. These findings of increased indirect costs from paying a higher compensating differential to CEOs complement those of Coats and Srinivasan (2014), who document the direct costs from control system expenditures incurred as a result of SOX's new requirements.

Third, despite the intention of SOX to make disclosure more transparent by reducing accounting manipulation, we find that SOX increased agency costs within most categories of all three sectors. In the primary sector this is mainly attributable to the higher cost of hidden actions, whereas in the consumer goods sector the cost of hidden information tended to increase. The latter finding is quite remarkable because the stated intention of SOX was to reduce the cost of obtaining private information by punishing the CEO and the CFO for financial misstatement. Evidently SOX exposed executive management to legal jeopardy from overstating their private information and thus exacerbated the incentive compatibility problem of inducing management to truthfully reveal good news that shareholders would use to help overcome the moral hazard issue of hidden actions.

Fourth, implementing SOX reduced the gross loss shareholders would bear if managers shirked, evidence that legislators were concerned with the potential for large losses rather than their expected value, which takes into account the probability of their occurrence. Ironically, these four summary findings suggest that laws introduced to improve corporate governance do not provide much evidence for the benevolent social planning view of legislative governance.

A Appendix

Proof of Lemma 1. In our model, the proof of Proposition 5 in Margiotta and Miller (2000) can be simply adapted to show that Theorem 3 of Fudenberg et al. (1990) applies, thus demonstrating that the long-term optimal contract can be sequentially implemented. An induction completes the proof by establishing that the sequential contract implementing the optimal long-term contract for a CEO who will retire in $\overline{\varsigma}$ periods replicates the one-period optimal contract. In the optimal short-term contract, the participation constraint is satisfied with strict equality, which implies that at the beginning of period $\overline{\varsigma} - 1$ the expected lifetime utility of the CEO is determined by setting $t = \overline{\varsigma} - 1$ in the equation

$$-b_t \exp\left(-\frac{a_t+\gamma_t e_t}{b_t}\right). \tag{A1}$$

Suppose that at the beginning of all periods $t \in \{\varsigma+1, \tau+2, \ldots, \overline{\varsigma}-1\}$; the expected lifetime utility of the CEO is given by equation (A1). We first show the expected lifetime utility of the CEO at ς is also given by Equation (A1). From Lemma 3.1 in the main text, the problem shareholders solve at ς is identical to the short-term optimization problem solved in the text. In the solution to each cost-minimization subproblem for the four (L_{1t}, L_{2t}) choices, the CEO's participation constraint is met with equality. Consequently, the CEO achieves the expected lifetime utility given by equation (A1), as claimed. Therefore, the problem of participating at time ς and possibly continuing with the firm for more than one period reduces to the problem of participating at time ς for one period at most, solved in Lemma 2. The induction step now follows. ■

Proof of Lemma 2. Let $\lambda_{t'}$ be the date-t price of a contingent claim made on a consumption unit at date t', implying the bond price is defined as $b_t \equiv E_t\left[\sum_{t'=t}^{\infty}\lambda_{t'}\right]$, and let q_t denote the date-t price of a security that pays off the random quantity $q_t \equiv E_t\left[\sum_{t'=t}^{\infty}\lambda_{t'}\left(\ln\lambda_{t'} - t'\ln\beta\right)\right]$. From equation (15) of Margiotta and Miller (2000, p. 680), the value to a CEO with current wealth endowment e_{nt} of announcing state $r_t(s)$ in period t when the true state is s and choosing effort level l_{st2} in anticipation of compensation $w_{r_t(s)t}(x)$ at the beginning of period $t+1$ when he retires one period later is

$$-b_t\alpha_{2t}^{1/b_t}\left\{E_t\left[\exp\left(-\tfrac{\gamma_t w_{r_t(s)t}(x)}{b_{t+1}}\right)\right]\right\}^{1-1/b_t}\exp\left(-\tfrac{q_t+\gamma_t e_{nt}}{b_{t+1}}\right).$$

The corresponding value from choosing effort level l_{st1} is

$$-b_t\alpha_{1t}^{1/b_t}\left\{E_t\left[\exp\left(-\tfrac{\gamma_t w_{r_t(s)t}(x)}{b_{t+1}}\right)\left[g_{st}(x)\right]\right]\right\}^{1-1/b_t}\exp\left(-\tfrac{q_t+\gamma_t e_{nt}}{b_{t+1}}\right),$$

whereas from equation (8) of Margiotta and Miller (2000, p. 678), the value from retiring immediately is $-b_t\exp\left(-\tfrac{q_t+\gamma_t e_{nt}}{b_{t+1}}\right)$. Dividing each expression through by the retirement utility, it immediately follows that the CEO chooses $l_{st} \equiv (l_{t0}, l_{st1}, l_{st2})$ to minimize the negative of expected utility:

$$l_{t0} + \left\{(\alpha_{1t}l_{st1} + \alpha_{2t}l_{st2})^{1/(b_t-1)}E_t\left[\exp\left(-\tfrac{\gamma_t w_{r_t(s)t}(x)}{b_{t+1}}\right)\left[g_{st}(x)l_{st1} + l_{st2}\right]\right]\right\}^{(b_t-1)/b_t}.$$

Because $l_{t0} \in \{0,1\}$ and $b_t > 1$, the solution to this optimization problem also solves

$$l_{t0} + (\alpha_1 l_{st1} + \alpha_2 l_{st2})^{1/(b_t-1)}E_t\left[\exp\left(-\tfrac{\gamma_t w_{r_t(s)t}(x)}{b_{t+1}}\right)\left[g_{st}(x)l_{st1} + l_{st2}\right]\right].$$

Summing over the two states $s \in \{1,2\}$ yields the minimand in Lemma 2. ■

References

ANG, J. S., CHENG, Y., AND FULMER, S. "Clawing Back Executive Compensation". Midwest Finance Association 2013 Annual Meeting Paper, 2013.

ANTLE, R. AND SMITH, A. "Measuring Executive Compensation: Methods and an Application." *Journal of Accounting Research*, Vol. 23 (1985), pp. 296–325.

ANTLE, R. AND SMITH, A. "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives." *Journal of Accounting Research*, Vol. 24 (1986), pp. 1–39.

BARGERON, L. L., LEHN, K. M., AND ZUTTER, C. J. "Sarbanes-Oxley and Corporate Risk-taking." *Journal of Accounting and Economics*, Vol. 49 (2010), pp. 34–52.

BOLTON, P. AND DEWATRIPONT, M. *Contract Theory.* Cambridge, MA: The MIT Press, 2005.

CARTER, M. E., LYNCH, L. J., AND ZECHMAN, S. L. C. "Changes in Bonus Contracts in the Post-Sarbanes-Oxley Era." *Review of Accounting Studies*, Vol. 14 (2009), pp. 480–506.

CENTER FOR AUDIT QUALITY (CAQ) "The CAQ's Fourth Annual Individual Investor Survey." (2010).

CHHAOCHHARIA, V., AND GRINSTEIN, Y. "Corporate Governance and Firm Value: The Impact of the 2002 Governance Rules." *Journal of Finance*, Vol. LXII (2007), pp.1789–1825.

COATES, J. C. AND SRINIVASAN, S. "SOX after Ten Years: A Multidisciplinary Review." *Accounting Horizons*, Vol. 28 (2014), pp. 627–671.

COHEN, D. A., DEY, A., AND LYS, T. Z. "Corporate Governance Reform and Executive Incentives: Implications for Investments and Risk Taking.", *Contemporary Accounting Research*, Vol.30 (2013), pp.1296–1332.

COHEN, D. A., DEY, A., AND LYS, T. Z. "Real and Accrual-Based Earnings Management in the Pre-and Post-Sarbanes-Oxley Periods." *Accounting Review*, Vol. 83 (2008), pp. 757–787.

COHEN, J. R., HAYES, C., KRISHNAMOORTHY, G., MONROE, G. S., AND WRIGHT, A. M. "The Effectiveness of SOX Regulation: An Interview Study of Corporate Directors." *Behavioral Research in Accounting*, Vol. 25 (2013), pp. 61–87.

ENGEL, E., HAYES, R. M., AND WANG, X. "The Sarbanes-Oxley Act and Firms' Going-private Decisions." *Journal of Accounting and Economics*, Vol. 44 (2007), pp. 116–145.

FUDENBERG, D., HOLMSTROM, B., AND MILGROM, P. "Short-term Contracts and Long-term Agency Relationships." *Journal of Economic Theory*, Vol. 51 (1990), pp. 1–31.

GAO, F., WU, J. S., AND ZIMMERMAN, J. "Unintended Consequences of Granting Small Firms Exemptions from Securities Regulation: Evidence from the Sarbanes-Oxley Act." *Journal of Accounting Research*, Vol. 47 (2009), pp. 459–506.

GAYLE, G.-L., GOLAN, L., AND MILLER, R. A. "Promotion, Turnover and

Compensation in the Executive Labor Market." *Econometrica*, Vol. 83 (2015), pp. 2293–2369.

GAYLE, G.-L. AND MILLER, R. A. "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review*, Vol. 99 (2009), pp. 1740–1769.

GAYLE, G.-L. AND MILLER, R. A. "Insider Information and Performance Pay." *CESifo Economic Studies*, Vol. 55 (2009), pp. 515–541.

GAYLE, G.-L. AND MILLER, R. A. "Identifying and Testing Models of Managerial Compensationn." *Review of Economic Studies*, Vol. 82 (2015), pp. 1074–1118.

HALL, B. J. AND LIEBMAN, J. B. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, Vol. 103 (1998), pp. 653–691.

HOCHBERG, Y. V., SAPIENZA, P., AND VISSING-JØRGENSEN, A. "A Lobbying Approach to Evaluating the Sarbanes-Oxley Act of 2002." *Journal of Accounting Research*, Vol. 47 (2009), pp. 519–583.

HOLMSTROM, B. AND KAPLAN, S. N. "The State of U.S. Corporate Governance." European Corporate Governance Institute Finance Working Paper, Vol. 23, 2003.

JAIN, P. K. AND REZAEE, Z. "The Sarbanes-Oxley Act of 2002 and Capital-Market Behavior: Early Evidence." *Contemporary Accounting Research*, Vol. 23 (2006), pp. 629–654.

KAMAR, E., KARACA-MANDIC, P., and TALLEY, E "Going-Private Decisions and the Sarbanes-Oxley Act of 2002: A Cross-Country Analysis." *Journal of Law, Economics, and Organization*, Vol. 25 (2009), pp. 107–133.

KANG, Q., LIU, Q., AND QI, R. "The Sarbanes-Oxley Act and Corporate Investment: A Structural Assessment." *Journal of Financial Economics*, Vol. 96 (2010), pp. 291–305.

LEUZ, C., TRIANTIS, A., AND WANG, T. Y. "Why Do Firms Go Dark? Causes and Economic Consequences of Voluntary SEC Deregistrations." *Journal of Accounting and Economics*, Vol. 45 (2008), pp. 181–208.

LI, Q. AND RACINE, J. S. *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ: Princeton University Press, 2007.

LITVAK, K. "The Effect of the Sarbanes-Oxley Act on Non-US Companies Cross-listed in the US." *Journal of Corporate Finance*, Vol. 13 (2007), pp. 195–228.

MARGIOTTA, M. M. AND MILLER, R. A. "Managerial Compensation and the Cost of Moral Hazard." *International Economic Review*, Vol. 41 (2000), pp. 669–719.

MURPHY, K. J. "Executive Compensation." In O. Ashenfelter and D. Card, eds., *The Handbook of Labor Economics*, Vol. 3, Amsterdam: Elsevier Science, 1999.

MURPHY, K. J. "Executive Compensation: Where We Are, and How We Got There." In G. Constantinides, M. Harris, and R. Stulz, eds., *Handbook of the Economics of Finance: Corporate Finance*, Amsterdam: Elsevier Science North Holland, 2013.

NEKIPELOV, D. "Empirical Content of a Continuous-Time Principal-Agent Model:

The Case of the Retail Apparel Industry." Working paper, University of California at Berkeley, 2010.

PENG, L. AND RÖELL, A. "Manipulation and Equity-based Compensation." *American Economic Review*, Vol. 98 (2008), pp. 285–90.

PROTIVITI. "Sarbanes-Oxley Compliance Survey." (2012).

ZHANG, I. X. "Economic Consequences of the Sarbanes-Oxley Act of 2002." *Journal of Accounting and Economics*, Vol. 44 (2007), pp. 74–115.

Table 1: Firm Characteristics and Compensation

	Primary Sector			Consumer Goods Sector			Service Sector		
	Pre	Post	t-/F-stat	Pre	Post	t-/F-statt	Pre	Post	t-/F-statt
Total assets	4460	6971	6.7	3460	5091	4.0	13678	20311	3.3
	(7436)	(11549)	(0.4)	(8242)	(11358)	(0.5)	(42619)	(78265)	(0.3)
Debt-to-equity	1.792	2.092	3.5	1.559	1.498	-0.9	3.651	2.798	-6.9
	(1.428)	(2.646)	(0.3)	(1.589)	(1.779)	(0.8)	(5.273)	(4.231)	(1.6)
Accounting return	1.110	1.135	3.2	1.117	1.070	-4.7	1.180	1.090	-11.7
	(0.233)	(0.231)	(1.0)	(0.283)	(0.258)	(1.2)	(0.342)	(0.258)	(1.8)

		Total Compensation								
		Overall			Bad			Good		
		Pre	Post	t-/F-stat	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat
Primary	(S, S)	1837	4704	3.6	387	2268	2.1	3593	8600	3.5
		(12126)	(13797)	(0.8)	(8132)	(12283)	(0.4)	(15480)	(15174)	(1.0)
	(S, L)	1099	4675	4.7	-131	3173	4.5	2818	6927	2.4
		(7970)	(8906)	(0.8)	(7520)	(5679)	(1.8)	(8267)	(11970)	(0.5)
	(L, S)	4300	10074	4.4	3225	8525	3.3	5772	11924	2.9
		(11983)	(16369)	(0.5)	(9827)	(15074)	(0.4)	(14305)	(17717)	(0.7)
	(L, L)	4204	8362	5.2	3147	7106	4.3	5595	10115	3.2
		(11571)	(14816)	(0.6)	(10662)	(12921)	(0.7)	(12540)	(17003)	(0.5)
Consumer goods	(S, S)	1425	2151	0.6	-1874	-2903	-0.9	6056	7875	0.9
		(21403)	(18474)	(1.3)	(15927)	(12315)	(1.7)	(26629)	(22279)	(1.4)
	(S, L)	1702	2355	0.4	-691	-1516	-0.4	4754	5120	0.1
		(14362)	(13605)	(1.1)	(11263)	(9682)	(1.4)	(17079)	(15316)	(1.2)
	(L, S)	6074	6793	0.2	1355	1399	0.0	12012	10903	-0.2
		(32090)	(31953)	(1.0)	(23206)	(24290)	(0.9)	(39848)	(36345)	(1.2)
	(L, L)	7297	9015	1.0	3460	4590	0.6	12710	13283	0.2
		(26565)	(28798)	(0.9)	(21908)	(22864)	(0.9)	(31228)	(33068)	(0.9)
Service	(S, S)	3757	2149	-1.9	590	-953	-1.8	7766	5700	-1.3
		(21304)	(18728)	(1.3)	(15543)	(14180)	(1.2)	(26350)	(22352)	(1.4)
	(S, L)	3311	4318	0.6	2041	1725	-0.2	4706	7579	0.9
		(16672)	(19656)	(0.7)	(13242)	(12581)	(1.1)	(19696)	(25703)	(0.6)
	(L, S)	11231	7135	-1.6	6065	205	-2.1	18877	14515	-1.0
		(38738)	(32155)	(1.5)	(31249)	(25182)	(1.5)	(46727)	(36918)	(1.6)
	(L, L)	9438	9185	-0.2	6383	6221	-0.1	14114	14365	0.1
		(26040)	(24173)	(1.2)	(22689)	(19051)	(1.4)	(29875)	(30536)	(1.0)

Note: In the columns "Pre" and "Post" indicating the pre- and post- SOX eras, standard deviation is listed in parentheses below the corresponding mean. The columns "t-/F-stat" report the statistics of a two-sided t-test on equal mean with critical value equal to 1.96 at the 5% confidence level, and the one-sided F-test on equal variance with critical value equal to 1. 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average. Assets (Compensation) is measured in millions (thousands) of 2006 U.S. dollars.

Table 2: Nonparametric Tests

A: Test on PDF of Abnormal Returns

Sector	Primary		Consumer goods		Service	
(A, C)	Bad	Good	Bad	Good	Bad	Good
(S, S)	24.16	27.82	15.52	14.67	23.65	23.76
(S, L)	8.31	6.85	-0.62	2.98	14.98	6.69
(L, S)	8.59	19.36	4.66	3.02	7.84	18.29
(L, L)	43.55	17.36	9.06	12.56	61.39	22.34

B: Test on Contract Shape

	Primary		Consumer Goods		Service	
(A, C)	Bad	Good	Bad	Good	Bad	Good
(S, S)	10.25	1.81	2.55	1.25	1.70	1.52
(S, L)	8.24	8.28	2.16	2.30	5.09	11.78
(L, S)	28.16	7.86	3.43	1.72	5.70	3.33
(L, L)	16.28	9.62	2.26	5.02	8.90	5.75

Note: 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average. Both tests are one-sided test and both statistics follow a standard normal distribution $N(0,1)$.

Table 3: Administrative Costs

(measured in thousands of 2006 US$)

Sector	(A, C)	τ_1	$\Delta\tau_1$	
Primary	(S, S)	(1440, 1860)	(2285, 2455)	+
	(S, L)	(872, 1043)	(3182, 3209)	+
	(L, S)	(3699, 4079)	(4113, 4648)	+
	(L, L)	(3727, 3994)	(2829, 3165)	+
Consumer goods	(S, S)	(-279, 1282)	(-437, 31)	=
	(S, L)	(931, 1407)	(-25, 110)	=
	(L, S)	(2467, 4560)	(-1041, 590)	=
	(L, L)	(4734, 6766)	(-767, -389)	-
Service	(S, S)	(2348, 3701)	(-1473, -1153)	-
	(S, L)	(1877, 2642)	(-462, -112)	-
	(L, S)	(7942, 10951)	(-4129, -3888)	-
	(L, L)	(7684, 9374)	(-1738, -1262)	-

Note: $\tau_1 \equiv \gamma^{-1}\frac{b_{t+1}}{b_t-1}\ln\alpha_{2,pre}$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table 4: Aggregate Agency Costs
(measured in thousands of 2006 US$)

Sector	(A, C)	τ_2	$\Delta\tau_2$	
Primary	(S, S)	(56, 477)	(20, 190)	+
	(S, L)	(22, 194)	(3, 30)	+
	(L, S)	(50, 430)	(76, 611)	+
	(L, L)	(35, 302)	(43, 379)	+
Consumer goods	(S, S)	(222, 1783)	(-527, -59)	-
	(S, L)	(65, 542)	(21, 156)	+
	(L, S)	(302, 2395)	(182, 1812)	+
	(L, L)	(290, 2323)	(81, 459)	+
Service	(S, S)	(187, 1540)	(-360, -41)	-
	(S, L)	(105, 869)	(45, 395)	+
	(L, S)	(416, 3425)	(113, 355)	+
	(L, L)	(233, 1924)	(53, 529)	+

Note: $\tau_2 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre}\left[w_{s,pre}(x)\right] - \tau_1$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table 5: Welfare Costs of Moral Hazard
(measured in thousands of 2006 US$)

Sector	(A, C)	τ_3	$\Delta\tau_3$	
Primary	(S, S)	(103, 1072)	(228, 1532)	+
	(S, L)	(73, 342)	(-39, 165)	=
	(L, S)	(91, 535)	(96, 1774)	+
	(L, L)	(6, 1149)	(265, 380)	+
Consumer goods	(S, S)	(1116, 7441)	(-4387, -600)	-
	(S, L)	(208, 1934)	(-817, -202)	-
	(L, S)	(1028, 5419)	(-3111, -649)	-
	(L, L)	(1067, 6876)	(-3848, -332)	-
Service	(S, S)	(492, 2612)	(-328, -150)	-
	(S, L)	(121, 1012)	(-399, 268)	=
	(L, S)	(1078, 9040)	(-6438, 433)	=
	(L, L)	(588, 8599)	(-2621, 479)	=

Note: $\tau_3 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre}[y_{s,pre}(x)] - \tau_1$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table 6: Welfare Costs of Hidden Information
(measured in thousands of 2006 US$)

Sector	(A, C)	τ_4	$\Delta\tau_4$	
Primary	(S, S)	(-596, -47)	(-1342, -208)	-
	(S, L)	(-149, -51)	(-135, 42)	=
	(L, S)	(-106, -41)	(-1163, -20)	-
	(L, L)	(-846, 29)	(-227, 87)	=
Consumer goods	(S, S)	(-5657, -893)	(540, 3860)	+
	(S, L)	(-1391, -143)	(223, 973)	+
	(L, S)	(-3024, -725)	(831, 4923)	+
	(L, L)	(-4553, -776)	(413, 4307)	+
Service	(S, S)	(-1072, -305)	(-32, 217)	=
	(S, L)	(-144, -16)	(-218, 795)	=
	(L, S)	(-5615, -663)	(-320, 6788)	=
	(L, L)	(-6675, -355)	(-348, 3150)	=

Note: $\tau_4 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre} \left[w_{s,pre}(x) - y_{s,pre}(x) \right]$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table 7: Gross Losses to Firms from CEO Shirking
(measured in percentage)

Sector	(A, C)	ρ_1	$\Delta\rho_1$	
Primary	(S, S)	(11.09, 11.31)	(-2.69, -1.96)	-
	(S, L)	(9.20, 11.70)	(-6.92, -4.75)	-
	(L, S)	(7.70, 9.67)	(-2.82, -2.10)	-
	(L, L)	(4.97, 5.70)	(-1.96, -1.95)	-
Consumer goods	(S, S)	(15.65, 16.28)	(-9.16, -8.72)	-
	(S, L)	(9.13, 13.15)	(2.12, 12.21)	+
	(L, S)	(6.60, 9.13)	(-0.40, 1.54)	-
	(L, L)	(5.46, 7.58)	(-2.68, -2.11)	-
Service	(S, S)	(19.64, 20.25)	(-8.93, -6.34)	-
	(S, L)	(10.48, 13.94)	(-3.02, -1.03)	-
	(L, S)	(17.25, 19.76)	(-16.59, -15.37)	-
	(L, L)	(7.63, 10.11)	(-5.97, -5.07)	-

Note: $\rho_1 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre} \left\{ x \left[1 - g_{s,pre}(x) \right] \right\}$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price (16.4).

Table 8: Compensating Differential from CEO Shirking versus Working

(measured in thousands of 2006 US$)

Sector	(A, C)	ρ_2	$\Delta\rho_2$	
Primary	(S, S)	(2262, 2879)	(122, 221)	+
	(S, L)	(1108, 1299)	(-57, -24)	-
	(L, S)	(1459, 1904)	(1716, 2125)	+
	(L, L)	(1395, 1665)	(100, 380)	+
Consumer goods	(S, S)	(5325, 7854)	(-3213, -2091)	-
	(S, L)	(1947, 2596)	(287, 476)	+
	(L, S)	(3314, 5727)	(18, 792)	+
	(L, L)	(2976, 5384)	(-1078, -654)	-
Service	(S, S)	(4024, 5728)	(-780, -487)	-
	(S, L)	(1549, 2455)	(67, 446)	+
	(L, S)	(6492, 10841)	(-7697, -5721)	-
	(L, L)	(4286, 6472)	(-2041, -1985)	-

Note: $\rho_2 \equiv b_{t+1}\left[(b_t - 1)\gamma\right]^{-1}\ln(\alpha_{2,pre}/\alpha_{1,pre})$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table 9: Change in Welfare Costs of Moral Hazard Caused by Signal Quality

(measured in thousands of 2006 US$)

Sector	(A, C)	$\tau_3(\alpha_{j,post}, g_{s,pre})$	ρ_3	
Primary	(S, S)	(115, 1343)	(216, 1261)	+
	(S,L)	(72, 309)	(-37, 198)	=
	(L, S)	(338, 2919)	(-617, -151)	-
	(L, L)	(46, 1382)	(59, 225)	+
Consumer Goods	(S, S)	(336, 1800)	(180, 1254)	+
	(S, L)	(278, 2692)	(-1575, -272)	-
	(L, S)	(1309, 5511)	(-3202, -930)	-
	(L, L)	(789, 2314)	(-224, 1181)	=
Service	(S, S)	(372, 1886)	(-130, 398)	=
	(S, L)	(172, 1116)	(-503, 212)	=
	(L, S)	(9, 92)	(1503, 3804)	+
	(L, L)	(226, 1509)	(724, 4470)	+

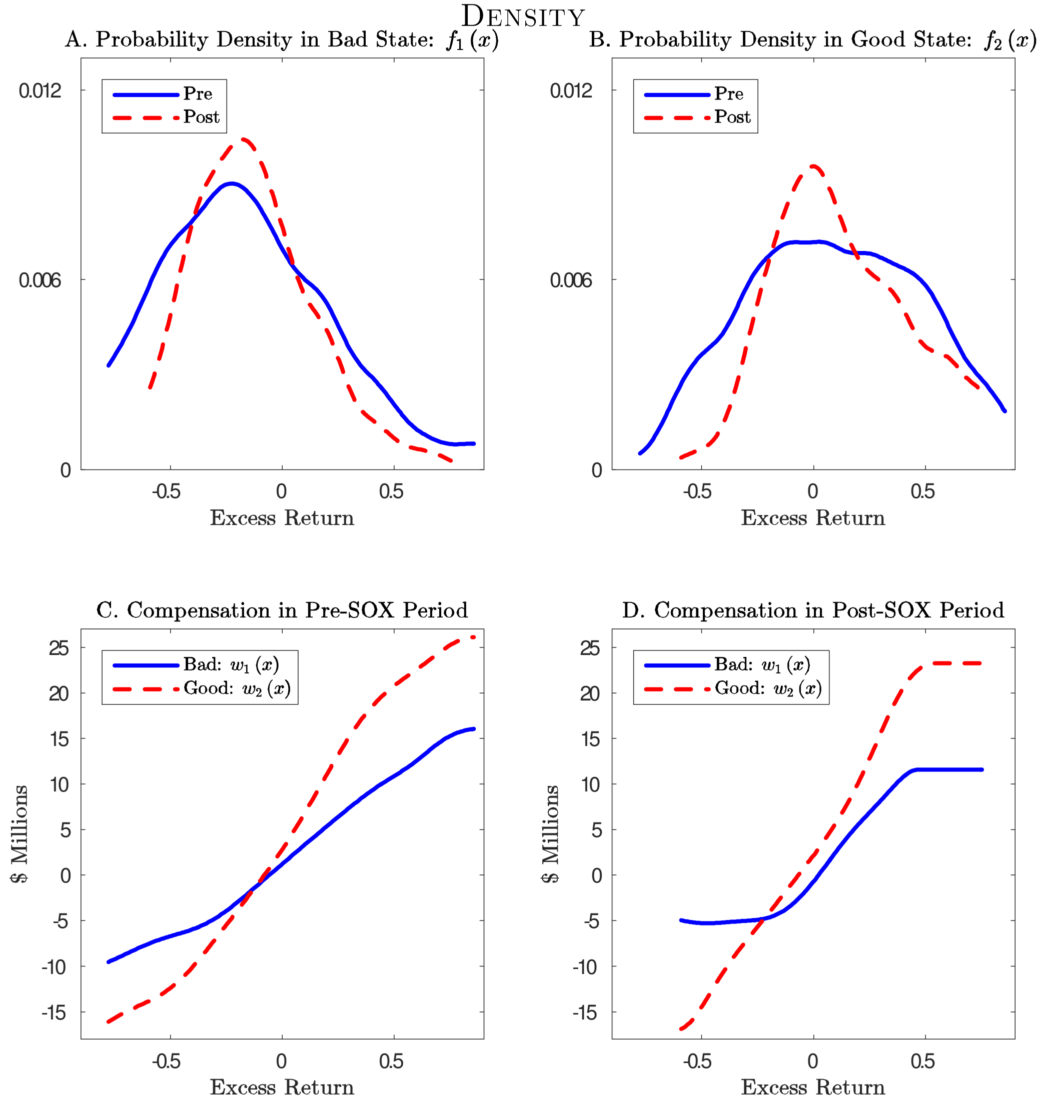
Note: $\rho_3 \equiv \tau_3(\alpha_{j,post}, g_{s,post}) - \tau_3(\alpha_{j,post}, g_{s,pre})$. Here "+" ("-") means that the change is positive (negative) and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Figure 1: Timeline of Events That Led to SOX and the Exchange Regulations

November 2001 11/8/2001 Enron files earnings restatement with the SEC.	December 2001 12/2/2001 Enron files for bankruptcy.	January-March 2002 Senate Banking committee hearings. House Financial Services committee hearings. 2/13/2002 SEC asks NYSE and NASDAQ to review their corporate governance requirements.	April 2002 4/12/2002 NASDAQ's Executive Committee approves first round of new corporate governance requirements. 4/24/2002 The House passes the bill of Senator Oxley to enhance corporate and auditing accountability, responsibility, and transparency.	May 2002 5/22/2002 NASDAQ's board approves the Executive Committee proposals.	June 2002 6/6/2002 NYSE committee unveils its new governance requirements. 6/25/2002 WorldCom announces that its profits had been inflated by $3.8 billion. 6/28/2002 NASDAQ's committee discusses 2nd round of proposals that would follow NYSE proposals.
July 2002 7/15/2002 The Senate passes the bill of Senator Sarbanes to enhance auditing related procedures, corporate responsibility, and financial disclosure. 7/24/2002 NASDAQ's board approves 2nd round of proposals. House and Senate reach a consensus on a joint SOX bill. 7/25/2002 House and Senate approve the SOX bill. 7/30/2002 President G.W. Bush signs the SOX into law.	August 2002 8/1/2002 NYSE board approves proposals. 8/16/2002 NYSE submits its proposed rules to the SEC.	September 2002	October 2002 10/9/2002 NASDAQ submits its proposed rules (2nd round) to the SEC.	....	November 2003 11/4/2003 SEC approves NASDAQ and NYSE proposals.

Note: From Chhaochharia and Grinstein (2007).

Figure 2: Empirical Compensation Schedule and Excess Return Density



Note: The plots 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. "Pre" and "Post" indicating the pre- and post- SOX eras. The compensation of both periods is anchored at bond prices equal to 16.5 (b_t) and 16.4 (b_{t+1}).

FIGURE 3: WELFARE COSTS SUMMARY

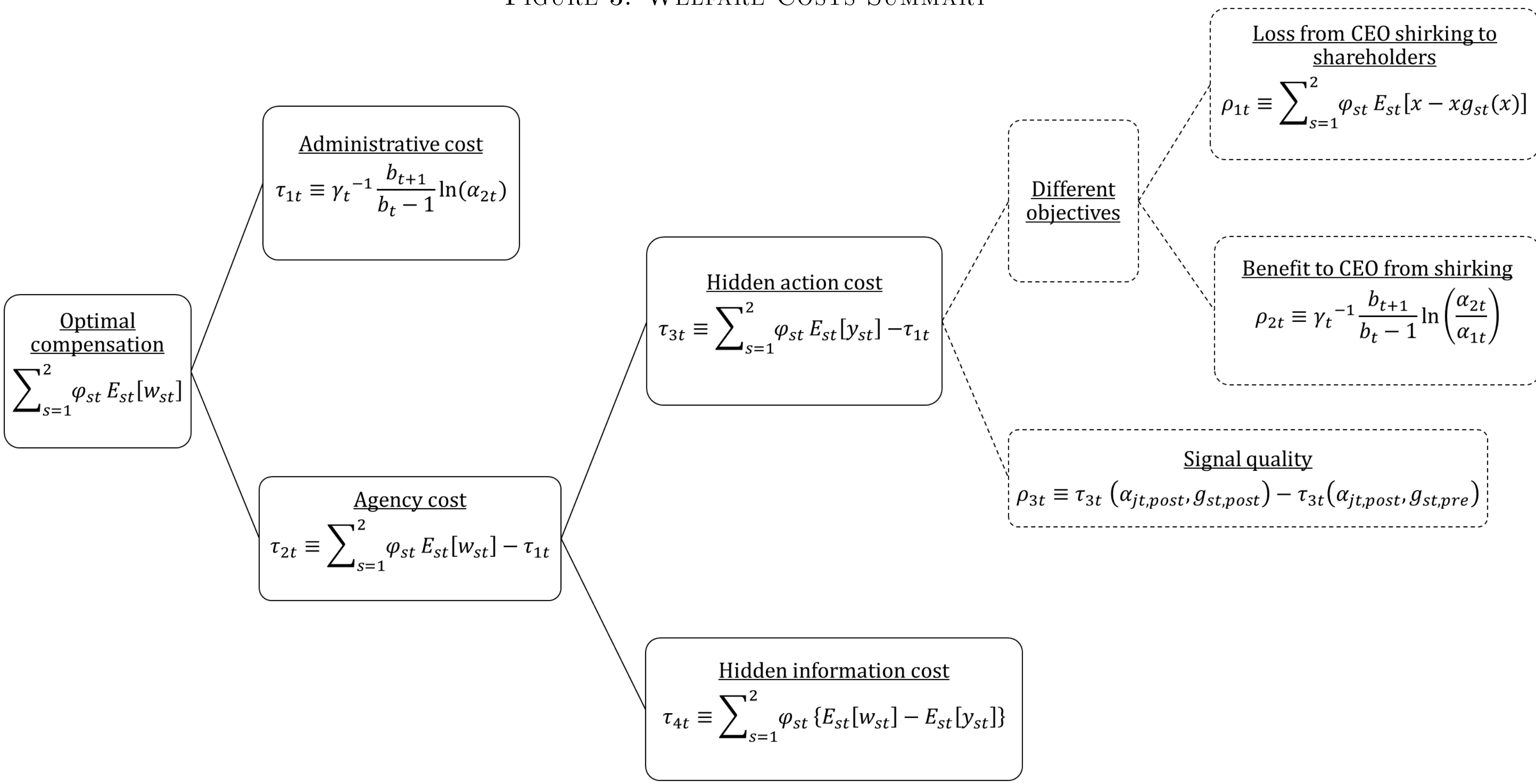
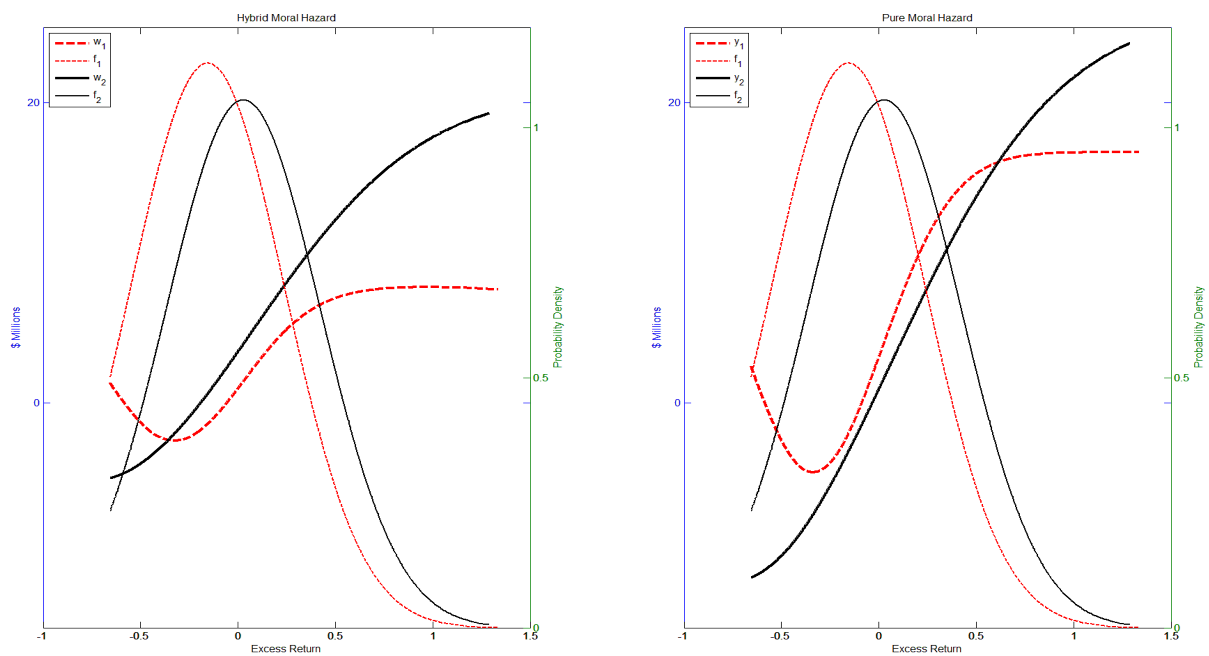


FIGURE 4: OPTIMAL COMPENSATION SCHEDULES



Note: The plots use the return and optimal compensation of firms with small size and low leverage in the Primary sector in the pre-SOX period. The risk aversion parameter γ equals 0.08. The cost of shirking (α_1) equals 0.96 and the effort cost 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 a one-side truncated normal distribution $TN(a,\mu,\sigma)$ with truncated point on the left (a), mean (μ), and standard deviation (σ) as follows. Working in bad state: $TN(-0.66,-0.16,0.39)$. Working in good state: $TN(-0.66,0.03,0.39)$. Shirking in bad state: $TN(-0.66,-0.25,0.27)$. 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.

Supplementary Appendices to: "Was Sarbanes-Oxley Costly? Evidence from Optimal Contracting on CEO Compensation"

George-Levi Gayle
Department of Economics, Washington University in St. Louis
and Federal Reserve Bank of St. Louis

Chen Li
Zicklin School of Business, Baruch College, CUNY

Robert A. Miller
Tepper School of Business, Carnegie Mellon University

April 25, 2016

Abstract

This supplementary appendix provides more details on the data construction (Appendix A), nonparametric tests for structural change (Appendix B), numeric solution of the pure hazard model (Appendix C) and identification and estimation (Appendix D). It also contains additional tables with data summary, intermediate results, and results from robustness check exercises.

A Data construction details

Firm type definition: Firm type is defined as a combination of industrial sector and firm characteristics for each firm in each era. 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 energy (GICS: 1010), materials (GICS: 1510), industrials (GICS: 2010, 2020, 2030), and utilities (GICS: 5510). The consumer goods sector includes firms in consumer discretionary (GICS: 2510, 2520, 2530, 2540, 2550) and consumer staples (GICS: 3010, 3020, 3030). The service sector includes firms in health care (GICS: 3510, 3520), financial (GICS: 4010, 4020, 4030, 4040), and information technology and telecommunication services (GICS: 4510, 4520, 5010). Firms that appear in different sectors over the sample period, they are classified into the sector in which they appear most frequently.

Second, we use binary variables based on firm size and capital structure (debt-to-equity ratio) to categorize firms into four types. The firm size is measured by the total assets on a firm's balance sheet (AT, variable name 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 whether its total assets in the pre-SOX era averaged over years were less than or greater than the median of the averaged total assets for firms in the same sector and whether its averaged debt-to-equity ratio was less than or greater than the median of the averaged debt-to-equity ratio for firms in that sector in the pre-SOX era. Therefore, firm type is measured by the coordinate pair (A, C) with each corresponding to whether that element is above (L) or below (S) its median of the industry in the pre-SOX era. 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 in both eras.

Accounting return definition In our model, after accepting the contractual arrangement, CEOs collect and convey their private information on the firm's prospects. We constructed 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 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 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 about prospects to shareholders.

Specifically, we define the binary private state, 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

$$r_{nt} = \frac{Asset_{nt} - Debt_{nt} + Dividend_{nt}}{Asset_{n,t-1} - Debt_{n,t-1}} \tag{A.1}$$

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 base year 2006 before calculating the accounting return.

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

Compensation In addition to the total compensation included in Compustat ExecuComp, 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 years 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, exercise prices, expiration dates, and issue dates. However, we need to know these Black-Scholes inputs for options granted before year 1993 to completely value the wealth change of CEOs by estimating the value of unexercised options and updating it each year. To facilitate the calculation, we assume that (1) all options are not exercised until expiration dates, (2) stock options granted before 1993 are exercised in a FIFO fashion, (3) 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 routines apply to those nonzero options granted before the CEO entered our sample. Then we apply the dividend-adjusted Black-Scholes formula to re-evaluate the call options for each CEO in each year. The dividend-adjusted Black-Scholes formula used is as follows. Let c denotes the call option value, K the exercise price, T_m

the time to maturity (in years), S the underlying security price, q the dividend yield, r the risk-free rate, and σ the implied volatility. Let $N(\cdot)$ denotes the standard normal cumulative distribution function. Then the call option value is given by

$$c = Se^{-qT_m}N(d_1) - Ke^{-rT_m}N(d_2), \tag{A.2}$$

$$d_1 = \frac{\ln(S/K) + (r - q + \sigma^2/2)T_m}{\sigma\sqrt{T_m}}, \tag{A.3}$$

and

$$d_2 = d_1 - \sigma\sqrt{T_m}. \tag{A.4}$$

Following the concept of income-equivalent total compensation adopted by Antle and Smith (1985, 1986), Hall and Liebman (1998), and Margiotta and Miller (2000), we construct the total compensation by adding change in wealth from options held and stocks held to the other components of compensation included in ExecuComp.

Additional summary tables Three additional tables with summary statistics of our dataset are included in this supplementary Appendix. Table A1 presents the time-series summary of the main firm characteristics. Table A2 presents the firm characteristics and compensation before and after SOX was enacted. It is a more detailed version of Table 1 in the main text. Table A3 presents estimates of the pre- and post-SOX probability distribution of accounting returns.

B Testing for structural change details

In the article, we conduct two nonparametric tests of structural change before and after SOX passage. We test for the inequality of the probability density functions for excess returns and for differences in the shape of the compensation schedule between the pre-and post-SOX eras.

Change in the density of abnormal returns Denote the set of categorical variables used to partition firms defined above by

$$Z \equiv \{primary, consumer\ goods, services\} \times \{S, L\} \times \{S, L\} \times \{Bad, Good\}. \tag{B.1}$$

Let $f_{pre}(x_{nt}|z_{nt})$ denote the probability density function of abnormal returns in the pre-SOX era conditional on $z_{nt} \in Z$ and define $f_{post}(x_{nt}|z_{nt})$ in a similar manner. For each $z_{nt} \in Z$ under the null hypothesis of no change $f_{pre}(x|z) = f_{post}(x|z)$ for all $(x, z) \in \mathcal{R} \times Z$. Let $N_{1,Z}$ and$N_{2,Z}$ denote, respectively, the number of observations in the pre- and post-SOX samples conditional on $z_{nt} \in Z$. Following Li and Racine (2007, p. 363), we calculate the statistics T_Z^{PDF} by

$$T_Z^{PDF} = (N_{1,z}N_{2,Z}h_Z^2)^{1/2}\frac{(I_{n,Z}^b - c_{n,b,Z})}{\widehat{\sigma_{b,Z}}}, \tag{B.2}$$

where

$$I_{n,Z}^b = \frac{1}{N_{1,Z}^2}\sum_{m=1}^{N_{1,Z}}\sum_{n=1}^{N_{1,Z}} K_{h,mn}^{pre,Z} + \frac{1}{N_{2,Z}^2}\sum_{m=1}^{N_{2,z}}\sum_{n=1}^{N_{2,z}} K_{h,mn}^{post,z} - \frac{2}{N_{1,z}N_{2,z}}\sum_{m=1}^{N_{1,Z}}\sum_{n=1}^{N_{2,Z}} K_{h,mn}^{pre,post,Z}, \tag{B.3}$$

$$c_{n,b,z} = \frac{1}{h_Z\sqrt{2\pi}\widehat{\sigma}_{x,z}}[\frac{1}{N_{1,Z}} + \frac{1}{N_{2,Z}}], \tag{B.4}$$

and

$$\begin{aligned}\widehat{\sigma_{b,Z}}^2 &= \frac{h_z}{N_{1,z}N_{2,Z}}\left\{\frac{N_{2,Z}}{N_{1,Z}}\left[\sum_{m=1}^{N_{1,Z}}\sum_{n=1}^{N_{1,Z}}(K_{h,mn}^{pre,Z})^2 + \sum_{m=1}^{N_{2,Z}}\sum_{n=1}^{N_{2,Z}}(N_{1,Z}/N_{2,Z})(K_{h,mn}^{post,Z})^2\right]\right.\\ &\left. +2\sum_{m=1}^{N_{1,Z}}\sum_{n=1}^{N_{2,Z}}(K_{h,mn}^{pre,post,Z})^2\right\}. \end{aligned} \tag{B.5}$$

The three kernel density functions are defined as

$$K_{h,mn}^{pre,Z} \equiv \frac{1}{h_Z\sqrt{2\pi}\widehat{\sigma}_{x,z}}\exp\left\{-\frac{1}{2}\left(\frac{x_{mt}^{pre} - x_{nt}^{pre}}{h_Z}\right)^2\right\}, \tag{B.6}$$

$$K_{h,mn}^{post,Z} \equiv \frac{1}{h_Z\sqrt{2\pi}\widehat{\sigma}_{x,z}}\exp\left\{-\frac{1}{2}\left(\frac{x_{mt}^{post} - x_{nt}^{post}}{h_Z}\right)^2\right\}, \tag{B.7}$$

and

$$K_{h,mn}^{pre,post,Z} \equiv \frac{1}{h_Z\sqrt{2\pi}\widehat{\sigma}_{x,z}}\exp\left\{-\frac{1}{2}\left(\frac{x_{mt}^{pre} - x_{nt}^{post}}{h_Z}\right)^2\right\}. \tag{B.8}$$

The bandwidth h_Z is Silverman's rule of thumb calculated conditional on Z for the combined sample periods. Similarly, $\widehat{\sigma}_{x,z}$ is the standard deviation of abnormal returns for the combined sample periods.

The test statistic T_Z^{PDF} is distributed normal with mean 0 and variance 1. It is a one-sided test and hence the null hypothesis is rejected at the (i) 1% significance level if $T_z^{PDF} > 2.33$, (ii) 5% significance level if $T_Z^{PDF} > 1.64$, and (iii) 10% significance level if $T_Z^{PDF} > 1.28$.

Change in the shape of the compensation schedule Let $w_{pre}(x_{nt}, z_{nt})$ denote CEO compensation as a function of (x_{nt}, z_{nt}) in the pre-SOX era, and similarly define $w_{post}(x_{nt}, z_{nt})$ in the post-SOX era. Our next test is based on the null hypothesis that there is no change is the shape of the compensation schedule and/or the level of compensation. This test is equivalent to a model specification test on the significance of a dummy variable in the standard Nadaraya-Watson kernel regression of observed compensation on the excess return conditional on $z_{nt} \in Z$. Let the dummy variable I^{SOX} equal 1 if the observation is from the post-SOX era and 0 otherwise. The null hypothesis can be stated as $\Pr\left[w\left(x, z, I^{SOX}\right) = W(x, z)\right] = 1$, where $W(x, z)$ is the compensation schedule when both eras are combined. This means that the shape of the compensation schedule is not significantly different between the pre- and post-SOX eras. If the null hypothesis is false, then the squared difference in nonparametric estimates of the functions $w\left(x, z, I^{SOX}\right)$ and $W(x, z)$ should be beyond certain critical values in the distribution of the test statistic, T_Z^W. This test statistic is defined as

$$T_z^W = \frac{1}{\sigma_{11,z}} \left[\sum_{n=1}^{N,z} \left\{ w_{n,z}^h - W_{n,Z}^H \right\}^2 \frac{A_{n,Z}}{N_Z} - \sum_{n=1}^{N} \frac{\sigma_{nh,Z}^2}{f_n^h} \frac{A_{n,Z}}{h_Z N 4\pi} - \sum_{n=1}^{N} \frac{\sigma_{nH,Z}^2}{f_n^H} \frac{h_Z \widetilde{A}_{n,Z}}{H_Z N 2\sqrt{\pi}} \right], \tag{B.9}$$

where $A_{n,Z}$ is a nonnegative N-dimension trimming vector whose element; corresponding to each observation. $A_{n,Z} = 1$ if x_n falls into the 2.5% to 97.5% range of excess returns otherwise, $A_n, z = 0$ for all $z_{nt} \in Z$ and $\widetilde{A}_{n,Z}$ is an estimate of the conditional expectation $A_{n,z}$ on x_n. The statistic T_z^W is a composite of the differences in the conditional mean ($w_{n,z}^h$ and $W_{n,Z}^H$) and variance ($\sigma_{nh,Z}^2$ and $\sigma_{nH,Z}^2$) between the post-SOX and the combined pre- and post-SOX eras. The kernel-based estimators of $w_{n,Z}^h$ and $W_{n,Z}^H$ are given by

$$w_{n,Z}^h = \sum_{m=1}^{N} w_m \left[\frac{I\{Z_m = Z\} I\{I_m^{SOX} = 1\} \exp\left\{ -\frac{1}{2} \left(\frac{x_m - x_n}{h_Z} \right)^2 \right\}}{\sum_{m=1}^{N} I\{Z_m = Z\} I\{I_m^{SOX} = 1\} \exp\left\{ -\frac{1}{2} \left(\frac{x_m - x_n}{h_Z} \right)^2 \right\}} \right] \tag{B.10}$$

and

$$W_{n,Z}^H = \sum_{m=1}^{N} w_m \left[\frac{I\{Z_m = Z\} \exp\left\{ -\frac{1}{2} \left(\frac{x_m - x_n}{H_Z} \right)^2 \right\}}{\sum_{m=1}^{N} I\{Z_m = Z\} \exp\left\{ -\frac{1}{2} \left(\frac{x_m - x_n}{H_Z} \right)^2 \right\}} \right], \tag{B.11}$$

where h_Z and H_Z are the bandwidths, respectively, for the post-SOX and combined sample periods.

The densities of abnormal returns, $f_{n,Z}^h$ and $f_{n,Z}^H$, are estimated by kernel density estimation and are given by

$$f_{n,Z}^{h}=\frac{\sum_{m=1}^{N}I\{Z_m{=}Z\}I\{I_m^{SOX}{=}1\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{h_Z}\right)^2\right\}}{N\sqrt{2\pi}\widehat{\sigma}_{x,z}} \tag{B.12}$$

and

$$f_{n,Z}^{H}=\frac{\sum_{m=1}^{N}I\{Z_m{=}Z\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{h_Z}\right)^2\right\}}{N\sqrt{2\pi}\widehat{\sigma}_{x,z}}. \tag{B.13}$$

The kernel-based estimator of the conditional expectation $A_{n,Z}$ on x_n, $\widetilde{A}_{n,Z}$, is given by

$$\widetilde{A}_{n,Z}=\sum_{m=1}^{N}A_{m,Z}\left[\frac{I\{Z_m{=}Z\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{H_Z}\right)^2\right\}}{\sum_{m=1}^{N}I\{Z_m{=}Z\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{H_Z}\right)^2\right\}}\right]. \tag{B.14}$$

Finally, the estimates of the conditional variance terms are given by

$$\sigma_{nh,Z}^{2}=\sum_{m=1}^{N}\left[w_m^2-\left(w_{n,Z}^{h}\right)^2\right]\left[\frac{I\{Z_m{=}Z\}I\{I_m^{SOX}{=}1\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{h_Z}\right)^2\right\}}{\sum_{m=1}^{N}I\{Z_m{=}Z\}I\{I_m^{SOX}{=}1\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{h_Z}\right)^2\right\}}\right], \tag{B.15}$$

$$\sigma_{nH,Z}^{2}=\sum_{m=1}^{N}\left[w_m^2-\left(W_n^{H}\right)^2\right]\left[\frac{I\{Z_m{=}Z\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{H_Z}\right)^2\right\}}{\sum_{m=1}^{N}I\{Z_m{=}Z\}\exp\left\{-\frac{1}{2}\left(\frac{x_m-x_n}{H_Z}\right)^2\right\}}\right], \tag{B.16}$$

and

$$\sigma_{11,z}^{2}=\frac{1}{N4\pi}\sum_{n=1}^{N}\frac{\left(\sigma_{nh,z}^{2}\right)^2A_{n},_z^2}{f_n^h}. \tag{B.17}$$

The test statistic T_Z^W is distributed normal with mean 0 and variance 1. It is a one-sided test and, hence, the null hypothesis is rejected at the (i) 1% significance level if $T_z^W>2.33$, (ii) 5% significance level if $T_Z^W>1.64$, and (iii) 10% significance level if $T_Z^W>1.28$. See Aït-Sahalia, Bickel, and Stoker (2001) for more details on this test.

C Numeric solution of the optimal contract in the pure moral hazard model

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 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

$$y_{st}(x) = \gamma^{-1}\frac{b_{t+1}}{b_t - 1}\ln\alpha_2 + \gamma^{-1}b_{t+1}\ln[1 + \eta^p_{st}\left(\frac{\alpha_2}{\alpha_1}\right)^{\frac{1}{b_t-1}} - \eta^p_{st}g_{st}(x)], \tag{C.1}$$

where η^p_{st} is the unique positive solution to

$$\int_{\overline{x}}^{\infty}\frac{g_{st}(x) - \left[\frac{\alpha_2}{\alpha_1}\right]^{\frac{1}{b_t-1}}}{1 + \eta^p_{st}\left[\frac{\alpha_2}{\alpha_1}\right]^{\frac{1}{b_t-1}} - \eta^p_{st}g_{st}(x)}f_s(x)dx = 0. \tag{C.2}$$

We approximate the integral (C.2), accounting for the singularity problem that occurs when the denominator of the integrand is either 0 or ∞. First, we performed 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 (C.2) is approximated for a given η^p_{st} by first being approximated on each subinterval and then summed over the entire range. Then we numerically solved for the optimal value of η^p_{st} that satisfies (C.2) based on this approximated integral.

D Identification and estimation details

This appendix presents the details about identification, estimation, and the counterfactual analysis of computing the decomposition.

D.1 Identification

This subsection establishes set identification of the risk aversion parameter, γ, and obtains sharp and tight bounds for the set.

Identifying tight and sharp bounds for γ The following set of restrictions places limits on the observationally equivalent values of γ. The model requires that at least one of the truth-telling constraint and the sincerity constraint should be binding. This implies that

$$\Psi_{3t}(\gamma) \equiv E_2\left[v_{1t}(x,\gamma) - v_{2t}(x,\gamma)\right] \geq 0, \tag{D.1}$$

$$\Psi_{4t}(\gamma) \equiv E_2\left[\alpha_{1t}(\gamma)^{1/(b_t-1)}v_{1t}(x,\gamma)g_{2t}(x,\gamma) - \alpha_{2t}(\gamma)^{1/(b_t-1)}v_{2t}(x,\gamma)\right] \geq 0, \tag{D.2}$$

and

$$\Psi_{5t}(\gamma^*) \equiv \Psi_{3t}(\gamma^*)\Psi_{4t}(\gamma^*) = 0. \tag{D.3}$$

To be Kuhn-Tucker multipliers, $\eta_{jt}(\gamma)$ for $j \in \{1,3,4\}$ need to be nonnegative. Meanwhile, the complementary slackness conditions for the truth-telling and sincerity constraints must be satisfied, which implies $\Psi_{6t}(\gamma) \equiv \Psi_{3t}(\gamma)\eta_{3t}(\gamma) = 0$ and $\Psi_{7t}(\gamma) \equiv \Psi_{4t}(\gamma)\eta_{4t}(\gamma) = 0$. Also, we impose another exclusion restriction that α_{1t} does not depend on the private state, yielding

$$\begin{aligned}\Psi_{1t}(\gamma) \quad \equiv \quad & E\left[v_{st}(x,\gamma)\right]^{-1} - E_1[v_{1t}(x,\gamma)]^{-1} - \eta_{3t}(\gamma)E_1[h(x)v_{1t}(x,\gamma)]E_1[v_{1t}(x,\gamma)]^{-1} \\ & -\eta_{4t}(\gamma)\left[\tfrac{\alpha_{1t}(\gamma)}{\alpha_{2t}(\gamma)}\right]^{1/(b_t-1)} E_1[g_{2t}(x,\gamma)h(x)v_{1t}(x,\gamma)]E_1[v_{1t}(x,\gamma)]^{-1} = 0. \end{aligned} \tag{D.4}$$

Besides, the likelihood g_{1t} should be positive with unit mass, implying that

$$\Psi_{2t}(\gamma) \equiv E_1\left[1\left\{g_{1t}(x,\gamma) > 0\right\} - 1\right] = 0. \tag{D.5}$$

The shareholders' profit maximization problem implies another three restrictions reflecting that they prefer the CEO to be working in both private states rather than shirking in any or both of them. Consequently, there are three inequality restrictions imposed on the data:

$$\Lambda_{1t}(\gamma) \equiv \sum\nolimits_{s=1}^{2} \varphi_s \left\{E_s[Vx - w_{st}(x)] - E_s\left[Vxg_{st}(x,\gamma) - \tfrac{b_{t+1}}{b_t-1}\gamma^{-1}\ln[\alpha_{1t}(\gamma)]\right]\right\} \geq 0 \tag{D.6}$$

where the compensation for a manager shirking in both states is $\frac{b_{t+1}}{b_t-1}\gamma^{-1}\ln[\alpha_{1t}(\gamma)]$,

$$\Lambda_{2t}(\gamma) \equiv \varphi_1 E_1[w_{1t}^{(1)}(x,\gamma) - w_{1t}(x)] + \varphi_2 E_2\left[x - w_{2t}(x) - g_{2t}(x,\gamma)\left[Vx - w_{2t}^{(1)}(x,\gamma)\right]\right] \geq 0, \tag{D.7}$$

and

$$\Lambda_{3t}(\gamma) \equiv \varphi_1 E_1\left[x - w_{1t}(x) - g_{1t}(x,\gamma)\left[Vx - w_{1t}^{(2)}(x,\gamma)\right]\right] + \varphi_2 E_2\left[w_{2t}^{(2)}(x,\gamma) - w_{2t}(x)\right] \geq 0. \tag{D.8}$$

The collection of these sets of restrictions defines Γ_{Ht}, a Borel set of risk aversion parameters, as

$$\Gamma_{\mathrm{H}t} \equiv \left\{ \gamma > 0 : \begin{array}{c} \Lambda_{it}(\gamma) \geq 0 \text{ for } i \in \{1,2,3\} \\ \eta_{jt}(\gamma) \geq 0 \text{ for } j \in \{1,3,4\} \\ \Psi_{jt}(\gamma) = 0 \text{ for } j \in \{1,2\} \text{ and } \Psi_{kt}(\gamma) \geq 0 \text{ for } k \in \{3,4\} \\ \Psi_{3t}(\gamma)\Psi_{4t}(\gamma) = \Psi_{3t}(\gamma)\eta_{3t}(\gamma) = \Psi_{4t}(\gamma)\eta_{4t}(\gamma) = 0 \end{array} \right\}. \tag{D.9}$$

In addition, we impose the restriction that the risk aversion does not depend on bond price. We take the intersection of the time-dependent sets to construct the identified set of risk aversion parameters as

$$\Gamma_{\mathrm{H}}(T) \equiv \bigcap_{t=1}^{T} \Gamma_{\mathrm{H}t} = \{\gamma > 0 : Q_{\mathrm{H}}(\gamma) = 0\} \tag{D.10}$$

where the criterion function is defined as

$$\begin{aligned} Q_{\mathrm{H}}(\gamma) \equiv & \sum_{t=1}^{T}\sum_{k=3}^{4} \min\left[0, \Psi_{kt}(\gamma)\right]^2 + \sum_{t=1}^{T}\sum_{j=1}^{2} \Psi_{jt}^2(\gamma) + \sum_{t=1}^{T}\sum_{j=5}^{7} \Psi_{jt}^2(\gamma) \\ & + \sum_{t=1}^{T}\sum_{k=1}^{3} \min\left[0, \Lambda_{kt}(\gamma)\right]^2 . \end{aligned} \tag{D.11}$$

If we further restrict the cost of effort to be stable over time, then the risk aversion parameter also needs to satisfy

$$\Psi_{8t}(\gamma^*) \equiv \alpha_{1t}(\gamma^*) - \alpha_{11}(\gamma^*) = 0, \forall t \tag{D.12}$$

and

$$\Psi_{9t}(\gamma^*) \equiv \alpha_{2t}(\gamma^*) - \alpha_{21}(\gamma^*) = 0, \ \forall t. \tag{D.13}$$

In this case, the criterion function can include another two quadratic terms as follows:

$$\begin{aligned} Q_{\mathrm{H}\alpha}(\gamma) \equiv & \sum_{t=1}^{T}\sum_{k=3}^{4} \min\left[0, \Psi_{kt}(\gamma)\right]^2 + \sum_{t=1}^{T}\sum_{j=1}^{2} \Psi_{jt}^2(\gamma) + \sum_{t=1}^{T}\sum_{j=5}^{7} \Psi_{jt}^2(\gamma) \\ & + \sum_{t=1}^{T}\sum_{k=1}^{3} \min\left[0, \Lambda_{kt}(\gamma)\right]^2 + \sum_{t=1}^{T}\sum_{j=8}^{9} \Psi_{jt}^2(\gamma). \end{aligned} \tag{D.14}$$

D.2 Estimation

Before we proceed to the estimation of the identified set for the risk aversion parameters and hence the remaining primitives — α_1 and α_2 are point identified, along with the likelihood ratios $g_{1t}(x)$ and $g_{2t}(x)$ — we first need to nonparametrically estimate the compensation schedule and the density of abnormal returns under working, $f_{st}(x)$.

Estimating optimal compensation and performance measures Our theoretical model implies that equity-based compensation is designed to align the interests of CEOs to those of shareholders and to incentivize the CEO to truthfully report his/her private information about the state of the firm. To empirically take this prediction to the data, we need to overcome two issues. First, the stock return that is used as a performance measure in the optimal contract should be closely tied to CEOs' efforts but eliminate stochastic variations that are out of their control. Second, the performance measure should reflect the outcome sharing between shareholders and CEOs; that is, it should reflect returns before compensation payment.

Taking into account these two points, we construct the performance measure, abnormal returns as called, in the following steps. First, we subtract the market portfolio return from the annual return to a firm stock in the same corresponding fiscal year. The residual captures the idiosyncratic components in firm stock returns. This non-diversifiable portion generates the incentive for the CEO to work rather than shirk. Given that neither the excess return nor the optimal compensation can be directly observed from the data, we construct consistent estimators of them as discussed below.

Let $\widetilde{x}_{nt}$ denote the net abnormal returns and $\widetilde{w}_{mt}$ denote the total compensation of firm n in year t observed in the dataset. First, we estimate the optimal compensation by running the following nonparametric regression:

$$\widehat{w}_{nt} = \sum_{m=1,m\neq n}^{N} \widetilde{w}_{mt} \left[\frac{I\{Z_{mt}=Z_{nt}\} K\left(\frac{\widetilde{x}_{mt}-\widetilde{x}_{nt}}{h_x}, \frac{v_{m,t-1}-v_{n,t-1}}{h_v}\right)}{\sum_{m=1,m\neq n}^{N} I\{Z_{mt}=Z_{nt}\} K\left(\frac{\widetilde{x}_{mt}-\widetilde{x}_{nt}}{h_x}, \frac{v_{m,t-1}-v_{n,t-1}}{h_v}\right)} \right] \tag{D.15}$$

where $v_{n,t-1}$ is the market value of firm n at the end of year $t-1$(See Gayle and Miller, 2015, for a formal justification of this procedure.).We used the multivariate standard normal kernel density function with Silverman's rule of thumb to choose the bandwidths as follows:

$$K\left(\frac{\widetilde{x}_{mt}-\widetilde{x}_{nt}}{h_x}, \frac{v_{m,t-1}-v_{n,t-1}}{h_v}\right) = \exp\left\{-\frac{1}{2}\left(\frac{\overline{x}_{mt}-\overline{x}_{nt}}{h_{x,Z}}\right)^2\right\} \exp\left\{-\frac{1}{2}\left(\frac{\overline{v}_{mt}-\overline{v}_{nt}}{h_{v,Z}}\right)^2\right\} \frac{|S_Z|^{-1/2}}{(2\pi)h_{x,Z}h_{v,Z}}, \tag{D.16}$$

where S_Z is the variance-covariance matrix of $\widetilde{x}$ and v is conditional on Z. The standardized version of $(\widetilde{x}_t, v_{t-1})$ (the net excess returns and raw one-year lagged market value) is defined as $(\overline{x}, \overline{v}) = (\widetilde{x}, v)S^{-1/2}$. The bandwidths are given by

$$h_{x,Z} = 1.06\sqrt{Var(x|Z)}(\textstyle\sum_{m=1,m\neq n}^{N} I\{Z_{mt} = Z_{nt}\})^{-1/5} \tag{D.17}$$

and

$$h_{v,Z} = 1.06\sqrt{Var(v|Z)}(\textstyle\sum_{m=1,m\neq n}^{N} I\{Z_{mt} = Z_{nt}\})^{-1/5}. \tag{D.18}$$

In the theoretical model, compensation is based on gross abnormal returns — that is, the abnormal return before compensation to the CEO. In the data, we observed net abnormal return — that is, the abnormal return after compensation to the CEO. To be internally consistent with the theory, the excess return is obtained by

$$x_{nt} \equiv \widetilde{x}_{nt} + \tfrac{\widehat{w}_{nt}}{v_{n,t-1}}. \tag{D.19}$$

Now the consistent estimate of optimal compensation conditional on $z \in Z$ is given by

$$w_t(x|Z) = \sum_{n=1}^{N} \widehat{w}_{nt} \left[\frac{I\{Z_{nt}=Z\} \exp\left\{-\frac{1}{2}\left(\frac{x_{nt}-x}{h_{x,Z}}\right)^2\right\}}{\sum_{n=1}^{N} I\{Z_{nt}=Z\} \exp\left\{-\frac{1}{2}\left(\frac{x_{nt}-x}{h_{x,Z}}\right)^2\right\}} \right]. \tag{D.20}$$

Finally, the probability density function of excess return, x_{nt}, is nonparametrically estimated by

$$f(x|Z) = \frac{1}{h_{x,Z}} \textstyle\sum_{n=1}^{N} \left[\dfrac{I\{Z_{nt}=Z\} \exp\left\{-\frac{1}{2}\left(\frac{x_{nt}-x}{h_{x,Z}}\right)^2\right\}}{\sum_{n=1}^{N} I\{Z_{nt}=Z\}} \right] \tag{D.21}$$

Estimating a confidence region for γ: Using the sample analog of the population components in the criterion functions in equation (D.14), we can construct the confidence region of the risk aversion parameter as

$$\Gamma_{\mathrm{H}\alpha}^{(N)}(T) \equiv \left\{\gamma > 0 : Q_{\mathrm{H}\alpha}(\gamma) \leq c_{\mathrm{H}\alpha\delta}^{(N)}\right\}, \tag{D.22}$$

where $c_{\mathrm{H}\delta}^{(N)}$ are consistent estimators for the critical values of the confidence regions associated with tests of size δ for each specification. In the actual implementation, the components $\sum_{t=1}^{T} \sum_{k=2}^{3} \min\left[0, \Lambda_{kt}^{(N)}(\gamma)\right]^2$ are not included for the same reason as in Gayle and Miller (2015).

We derive a confidence region that covers the identified set of observationally equivalent parameters, any element of which could have generated the data. First, we derive a confidence region for γ by exploiting the fact that approximations to $Q_{\mathrm{H}\alpha}(\gamma)$ formed from the data deviate from zero only because of differences between expectations and limits in the population and their sample analog. While the rate of convergence can be derived analytically ($\sqrt{NT}$ when x is bounded), subsampling proved to be the most practical way of determining a confidence region for any given critical value. The confidence region takes the form $\{\gamma : Q_{NT}(\gamma) \leq c_{\mathrm{H}\alpha,\delta}\}$, where $Q_{NT}(\gamma)$ is a sample analog to $Q_{\mathrm{H}\alpha}(\gamma)$. We then modify the subsampling procedure proposed by Chernozhukov, Hong and Tamer (2007) to estimate the critical value $c_{\mathrm{H}\alpha,\delta}$. Consider all subsets of the data with size $N_b < N$, where $N_b \longrightarrow \infty$, but $N_b/N \longrightarrow 0$, and denote the number of subsets by B_N. Define $c_{\mathrm{H}\alpha,\delta}$ and $\Gamma^{(N)}_{\mathrm{H}\alpha}(T)$ as

$$c_0 \equiv \inf_{\widetilde{\gamma} > \widetilde{\gamma}_N} \left[N^{1/3} Q_{NT}(\gamma)\right] + \kappa_N \tag{D.23}$$

$$\Gamma^{(N)}_{\mathrm{H}\alpha,0}(T) \equiv \{\gamma \geq \gamma_N : N^{1/3} Q_{NT}(\gamma) \leq c_0\}, \tag{D.24}$$

where $\kappa_N \propto \ln N$ and γ_N, a strictly positive sequence, converges to zero at a rate faster than N^a. For each subset $i \in \{1, ..., B_N\}$ of size N_b define

$$C^{(i,N_b)}_{\mathrm{H}\alpha} \equiv \sup_{\gamma \in \Gamma^{(N)}_{\mathrm{H}\alpha,0}} \left[(N_b)^{1/3} Q^{(i,N_b)}_{NT}(\gamma)\right], \tag{D.25}$$

and denote by $c^{(N)}_{\mathrm{H}\alpha\delta}$ the $\delta-$quantile of the sample $\left\{C^{(1,N_b)}_{\mathrm{H}\alpha}, \ldots, C^{(B_N,N_b)}_{\mathrm{H}\alpha}\right\}$.

To implement the subsampling procedure, we draw 100 subsamples from the original full sample, following the joint distribution of the public states and the private states. Each subsample contains 80% of the observations in the original sample. For each subsample, we calculate the value of the objective function and use these values to estimate the 95% critical value of the confidence region. The 95 percent confidence region of the risk aversion parameter in the CEO CARA utility function is displayed in Table D1, estimated for each phase separately and imposing a common value over both phases. The confidence regions in Panel A are obtained using the full sample. The Certainty Equivalent column in Table D2 gives economic meaning to the estimates of risk aversion in Table D1, where the amount a CEO would pay to avoid an equiprobable gamble with losing or winning \$1,000,000.

D.3 Counterfactual analysis

Nonparametric identification and estimation are useful in (i) exploring what variation in the data identifies which parameter in our model and (ii) guarding against rejection of our model because of functional form restrictions not necessary to obtain the theoretical results of the model. This explains why up until now we have maintained that $f_{st}(x)$ and $g_{st}(x)f_{st}(x)$ are nonparametrically specified and estimated. However, for counterfactual analysis, maintaining the nonparametric specification becomes problematic. First, nonparametric estimates are not always smooth and differentiable which makes numerical analysis (e.g. finding the root of equation (C.2) in the optimal contract) difficult. Second, working with data outside the range of the observed sample under different counterfactual regimes will call for extrapolation as nonparametric estimates are usually defined only over the data range observed in our sample. For these and other reasons, we approximate $f_{st}(x)$ and $g_{st}(x)f_{st}(x)$ by truncated normal distributions before calculating the counterfactual welfare costs. This subsection outlines the details of these approximation procedures and assesses their performance.

We assume that the distribution of gross returns when the CEO work, $f_{st}(x)$, is truncated normal with support bounded from below by x_L. Specifically, we assume

$$f_{zs}(x, x_L, \mu_{zs}^F, \sigma_{zs}^F) = \left[\Phi\left(\frac{\mu_{zs}^F - x_{zL}}{\sigma_{zs}^F}\right)\sigma_{zs}^F\sqrt{2\pi}\right]^{-1} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_{zs}^F}{\sigma_{zs}^F}\right)^2\right], \tag{D.26}$$

where Φ is the standard normal distribution function and $(\mu_{zs}^F, \sigma_{zs}^F)$ denote the mean and standard deviation of the parent normal distribution. The cutoff point of the support is estimated by a superconsistent estimator of the lower bound of the observed data. Then we obtain an equally spaced grid on $[x_{zL}, \infty)$ and use the kernel density estimator in equation (D.21) to obtain the nonparametric estimator of $f_s^{(N)}(x_j|Z)$ for $x_j \in [x_{zL}, \infty)$. Next we choose μ_{zs}^F and σ_{zs}^F to minimize the mean squared deviation between $f^{(N)}(x_j|Z)$ and $f_{zs}(x_j, x_{zL}, \mu_{zs}^F, \sigma_{zs}^F)$. Formally,

$$(\widehat{\mu}_{zs}^F, \widehat{\sigma}_{zs}^F) = \arg\min_{\mu_{zs}^F, \sigma_{zs}^F} \sum_{j=1}^{J} \left[f_s^{(N)}(x_j|Z) - f_{zs}^{TN}(x_j, x_{zL}, \mu_{zs}^F, \sigma_{zs}^F)\right]^2. \tag{D.27}$$

Table D3 presents the MSE for these approximations. It shows that the truncated normal distribution can approximate closely the distribution of excess return under working.

Similarly, we approximate $g_{st}(x)f_{st}(x)$ by a truncated normal distribution. However, there are several things that are different. First, these estimates depend on the risk-aversion parameter estimates and the bond prices. The risk aversion parameter is set identified and estimated; hence, the approximation has to be done for each point in the estimated set. So, we first estimate $\widetilde{g}_{z1t}(x_j, \gamma)$ and $\widetilde{g}_{z2t}(x_j, \gamma)$ for each $\gamma \in \Gamma_{\mathrm{H}}^{(N)}(T)$ and each bond price indexed by t in subscript (suppressed here for notational ease). That is, for each $x_j \in [x_{zL}, \infty)$ we compute the following:

$$\widetilde{g}_{z2}(x_j,\gamma) \equiv \frac{\overline{v}_{z2}(\gamma)^{-1} - v_{z2}(x_j,\gamma)^{-1}}{\overline{v}_{z2}(\gamma)^{-1} - E_{z2}\left[v_{z2}(x_j,\gamma)^{-1}\right]} \tag{D.28}$$

$$\begin{aligned}\widetilde{g}_{z1}(x_j,\gamma) \quad \equiv \quad & \frac{\overline{v}_{z1}(\gamma)^{-1} - v_{z1}(x_j,\gamma)^{-1}}{\eta_{z1}(\gamma)} \\ & + \frac{\eta_{z3}(\gamma)\left[\overline{h}_z - h_z(x_j)\right] - \eta_{z4}(\gamma) g_{z2}(x_j,\gamma) h_z(x) \frac{\widehat{\alpha}_{z1}(\gamma)}{\widehat{\alpha}_{z2}(\gamma)}}{\eta_{z1}(\gamma)}. \end{aligned} \tag{D.29}$$

Then we specify the truncated truncated normal distributions for each γ and each bond price t as

$$fg_{zs}^{TN}(x_j, x_L, \mu_{zs}^{FG}, \sigma_{zs}^{FG}) = \left[\Phi\left(\frac{\mu_{zs}^{FG} - x_{zL}}{\sigma_{zs}^{FG}}\right)\sigma_{zs}^{FG}\sqrt{2\pi}\right]^{-1} \exp\left[-\frac{1}{2}\left(\frac{x_{0j} - \mu_{zs}^{FG}}{\sigma_{zs}^{FG}}\right)^2\right], \tag{D.30}$$

where Φ is the standard normal distribution function and $(\mu_{zs}^{FG}, \sigma_{zs}^{FG})$ denote the mean and standard deviation of the parent normal distribution. We use the estimated value of x_{zL} as in $f_{zs}^{TN}(x_j, x_{zL}, \mu_{zs}^{F}, \sigma_{zs}^{F})$. This is done for theoretical reasons because if the support of $f_{st}(x)$ and $g_{st}(x)f_{st}(x)$ is not the same, then the first-best allocation can be achieved in the principal agent problem. The mean and standard deviation of the truncated normal distribution are then obtained by minimizing the MSE as follows:

$$(\widehat{\mu}_{zs}^{FG}(\gamma), \widehat{\sigma}_{zs}^{FG}(\gamma)) = \arg \min_{\mu_{zs}^{FG}(\gamma), \sigma_{zs}^{FG}(\gamma)} \sum_{j=1}^{J} \left\{ \left[f_s^{(N)}(x_j|Z)\widetilde{g}_{zs}(x_j,\gamma) - fg_{zs}^{TN}(x_j, \mu_{zs}^{FG}, \sigma_{zs}^{FG}) \right]^2 \right\}. \tag{D.31}$$

Table D3 presents the confidence interval of the MSE for these approximations. It shows that the truncated normal distribution can approximate closely the distribution of excess return under shirking.

References

[1] ANTLE, R. AND SMITH, A. "Measuring Executive Compensation: Methods and an Application." *Journal of Accounting Research*, Vol. 23 (1985), pp. 296-325.

[2] ANTLE, R. AND SMITH, A. "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives." *Journal of Accounting Research*, Vol. 24 (1986),

pp. 1-39.

[3] AÏT-SAHALIA, Y., BICKEL, P. J., AND STOKER, T. M. "Goodness-of-fit Tests for Kernel Regression with an Application to Option Implied Volatilities." *Journal of Econometrics*, Vol. 105 (2001), pp. 363–412.

[4] GAYLE, G.-L. AND MILLER, R. A. "Identifying and Testing Models of Managerial Compensationn." *Review of Economic Studies*, Vol. 82 (2015), pp. 1074–1118.

[5] HALL, B. J. AND LIEBMAN, J. B. "Are CEOs Really Paid Like Bureaucrats?" *The Quarterly Journal of Economics*, Vol. 103 (1998), pp. 653–691.

[6] LI, Q. AND RACINE, J. S. *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ: Princeton University Press, 2007.

[7] MARGIOTTA, M. M. AND MILLER, R. A. "Managerial Compensation and the Cost of Moral Hazard." *International Economic Review*, Vol. 41 (2000), pp. 669–719.

Table A1: Time-series Summary

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Asset	6647	6269	6449	6785	7056	7384	9229	10154	10589	10840	12077	12081	13275
	(19191)	(18153)	(19230)	(20944)	(22586)	(25147)	(34536)	(37564)	(41634)	(40157)	(50961)	(50965)	(57501)
Debt/Equity	2.604	2.413	2.548	2.429	2.314	2.391	2.515	2.562	2.447	2.415	2.466	2.354	2.188
	(3.925)	(3.380)	(3.738)	(3.543)	(3.677)	(3.719)	(3.576)	(3.717)	(3.534)	(3.518)	(3.640)	(3.522)	(3.206)
Accounting	1.155	1.139	1.169	1.147	1.177	1.144	1.127	1.111	1.086	1.048	1.104	1.128	1.068
Return	(0.276)	(0.255)	(0.285)	(0.285)	(0.313)	(0.295)	(0.284)	(0.332)	(0.311)	(0.291)	(0.229)	(0.266)	(0.232)
Net Excess	0.051	-0.029	-0.072	-0.046	-0.032	-0.160	-0.206	0.092	0.140	0.052	0.051	0.024	-0.011
Return	(0.324)	(0.280)	(0.368)	(0.335)	(0.382)	(0.402)	(0.415)	(0.468)	(0.369)	(0.340)	(0.351)	(0.307)	(0.289)
Observations	1647	1860	1850	1914	2030	2029	1837	1822	1881	1958	1785	1767	1697

Note: Assets are listed in millions of 2006 US$. Standard deviations are shown in parentheses. Net excess returns are firm stock returns net of a return to a value-weighted market portfolio.

Table A2: Firm Characteristics and Compensation (Extended Sample)

	Primary Sector			Consumer goods sector			Service sector		
	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat
Total assets	4602	6910	7.4	3637	5126	4.2	14009	19967	3.6
	(7713)	(11783)	(0.4)	(8872)	(12010)	(0.5)	(44521)	(76742)	(0.3)
Debt-to-equity	1.810	2.113	4.5	1.561	1.530	-0.5	3.580	2.913	-6.2
	(1.437)	(2.548)	(0.3)	(1.627)	(2.033)	(0.6)	(5.204)	(4.368)	(1.4)
Accounting	1.100	1.126	4.0	1.115	1.085	-3.6	1.164	1.092	-11.1
returns	(0.236)	(0.221)	(1.1)	(0.282)	(0.253)	(1.2)	(0.343)	(0.252)	(1.8)

		Overall			Bad			Good		
Sector	(A, C)	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat
Primary	(S, S)	2190	4047	2.8	521	1526	1.3	4096	7206	2.7
		(12452)	(13908)	(0.8)	(8244)	(12516)	(0.4)	(15743)	(14913)	(1.1)
	(S, L)	1467	4120	3.9	-29	2432	4.0	3453	6299	2.3
		(8678)	(8805)	(1.0)	(7356)	(5528)	(1.8)	(9835)	(11433)	(0.7)
	(L, S)	4771	8131	3.3	3429	6849	2.7	6501	9607	2.0
		(12224)	(15201)	(0.6)	(10015)	(14023)	(0.5)	(14415)	(16390)	(0.8)
	(L, L)	4546	8460	5.7	3368	6586	3.9	6015	10569	4.1
		(11860)	(15407)	(0.6)	(10847)	(13337)	(0.7)	(12867)	(17228)	(0.6)
Consumer goods	(S, S)	2377	2821	0.5	-1203	-1594	-0.4	7301	7494	0.1
		(21723)	(18327)	(1.4)	(15811)	(11796)	(1.8)	(27128)	(22425)	(1.5)
	(S, L)	1712	2565	0.8	-567	65	0.5	4494	4528	0.0
		(13992)	(11803)	(1.4)	(11016)	(8268)	(1.8)	(16528)	(13700)	(1.5)
	(L, S)	6450	8580	0.9	1569	2279	0.3	12432	13560	0.3
		(31513)	(31025)	(1.0)	(23109)	(26279)	(0.8)	(38644)	(33589)	(1.3)
	(L, L)	7595	10984	2.0	3858	5588	0.9	12923	16759	1.4
		(26364)	(31235)	(0.7)	(21787)	(25096)	(0.8)	(31014)	(35861)	(0.7)
Service	(S, S)	3408	4078	0.9	454	558	0.1	6929	7782	0.7
		(20488)	(18998)	(1.2)	(14951)	(13127)	(1.3)	(25125)	(23095)	(1.2)
	(S, L)	3384	4416	0.7	1814	2600	0.6	5145	6497	0.5
		(16394)	(18797)	(0.8)	(13211)	(12876)	(1.1)	(19218)	(23743)	(0.7)
	(L, S)	10890	11573	0.3	5351	5502	0.1	18610	17536	-0.3
		(38668)	(34641)	(1.2)	(30923)	(29914)	(1.1)	(46350)	(37877)	(1.5)
	(L, L)	9812	10678	0.9	6666	6964	0.3	14321	16451	1.1
		(26134)	(26185)	(1.0)	(22752)	(19732)	(1.3)	(29764)	(33072)	(0.8)

Note: In the columns "Pre" and "Post" indicating the pre- and post- SOX eras, standard deviation is listed in parentheses below the corresponding mean. The columns "t-/F-stat" report the statistics of a two-sided t-test on equal mean with critical value equal to 1.96 at the 5% confidence level, and the one-sided F-test on equal variance with critical value equal to 1. 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average. Assets (Compensation) is measured in millions (thousands) of 2006 U.S. dollars.

TABLE A3: NONPARAMETRIC TESTS (EXTENDED SAMPLE)

A: Test on PDF Abnormal of Returns

Sector	Primary		Consumer goods		Service	
(A, C)	Bad	Good	Bad	Good	Bad	Good
(S, S)	18.05	10.34	12.51	12.39	14.25	14.55
(S, L)	5.88	5.02	1.26	2.27	14.70	5.29
(L, S)	3.29	4.16	3.74	2.03	9.01	19.69
(L, L)	29.46	8.57	9.03	8.68	71.68	29.56

B: Test on Contract Shape

Sector	Primary		Consumer goods		Service	
Firm Type	Bad	Good	Bad	Good	Bad	Good
(S, S)	10.06	1.58	2.89	1.09	1.54	1.47
(S, L)	6.82	6.45	3.30	1.71	4.08	6.85
(L, S)	19.67	7.34	5.51	3.52	5.66	8.74
(L, L)	10.32	23.38	3.69	6.74	7.37	10.65

Note: 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average. Both tests are one-sided test and both statistics follow a standard normal distribution $N(0, 1)$.

Table D1: Estimates of the Probability Distribution of Accounting Returns

A. Pre-SOX (Main Sample)

Firm Type	Primary		Consumer goods		Service	
(A, C)	Bad/Good	Obs	Bad/Good	Obs	Bad/Good	Obs
(S, S)	1.2	1840	1.4	1500	1.3	2359
(S, L)	1.4	779	1.3	669	1.1	638
(L, S)	1.4	898	1.3	752	1.5	796
(L, L)	1.3	2134	1.4	1625	1.5	2880
Total	1.3	5651	1.4	4546	1.4	6673

B. Post-SOX (Main Sample)

Firm Type	Primary		Consumer goods		Service	
(A, C)	Bad/Good	Obs	Bad/Good	Obs	Bad/Good	Obs
(S, S)	1.6	343	1.1	322	1.1	637
(S, L)	1.5	130	0.7	96	1.3	149
(L, S)	1.2	169	0.8	148	1.1	223
(L, L)	1.4	381	1.0	277	1.7	588
Total	1.4	1023	1.0	843	1.3	1597

C. Pre-SOX (Extended Sample)

Firm Type	Primary		Consumer goods		Service	
(A, C)	Bad/Good	Obs	Bad/Good	Obs	Bad/Good	Obs
(S, S)	1.2	2039	1.4	1665	1.2	2738
(S, L)	1.3	852	1.2	724	1.1	719
(L, S)	1.3	989	1.2	893	1.3	924
(L, L)	1.2	2335	1.4	1773	1.4	3231
Total	1.2	6215	1.3	5001	1.3	7612

D. Post-SOX (Extended Sample)

Firm Type	Primary		Consumer goods		Service	
(A, C)	Bad/Good	Obs	Bad/Good	Obs	Bad/Good	Obs
(S, S)	1.3	534	1.1	494	1.1	944
(S, L)	1.3	197	0.8	150	1.1	221
(L, S)	1.2	256	0.8	222	1.0	331
(L, L)	1.1	576	1.1	412	1.6	912
Total	1.2	1563	1.0	1278	1.2	2408

Note: 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average.

Table D2: The 95% Confidence Regions of Risk-aversion and Corresponding Certainty Equivalent (in 2006 US$)

A: Main Sample

Period	Years	Risk Aversion	Certainty Equivalent
Pre-SOX	1993-2001	(0.0695, 0.6158)	(34722, 290206)
Post-SOX	2004-2005	(0.0695, 0.6158)	(34722, 290206)
Common		(0.0695, 0.6158)	(34722, 290206)

B: Extended Sample

Period	Years	Risk Aversion	Certainty Equivalent
Pre-SOX	1993-2002	(0.0784, 0.2335)	(39160, 115704)
Post-SOX	2003-2005	(0.0616, 0.2335)	(30781, 115704)
Common		(0.0784, 0.2335)	(39160, 115704)

Note: The subsampling procedure was performed using 100 replications of subsamples with 80% of full sample observations, each using 100 grid points on the searching interval $[0.0003, 54.598]$. The certainty equivalent corresponding to one particular value of the risk aversion in the estimated confidence region is the certainty equivalent of a equiprobable gamble of losing or winning 1 million dollars.

Table D3: MSE of the Density Approximation Under working

		Main Sample				Extended Sample			
		Pre-SOX		Post-SOX		Pre-SOX		Post-SOX	
Sector	Firm Type	Bad	Good	Bad	Good	Bad	Good	Bad	Good
Primary	(S, S)	0.006	0.004	0.004	0.013	0.006	0.003	0.006	0.017
	(S, L)	0.009	0.005	0.069	0.038	0.007	0.005	0.042	0.028
	(L, S)	0.006	0.002	0.044	0.013	0.006	0.002	0.018	0.004
	(L, L)	0.014	0.011	0.028	0.019	0.013	0.007	0.022	0.014
Consumer goods	(S, S)	0.004	0.003	0.014	0.018	0.003	0.002	0.008	0.015
	(S, L)	0.008	0.005	0.021	0.020	0.008	0.006	0.038	0.016
	(L, S)	0.003	0.004	0.022	0.003	0.002	0.002	0.013	0.003
	(L, L)	0.005	0.004	0.017	0.003	0.004	0.003	0.018	0.010
Service	(S, S)	0.005	0.003	0.012	0.002	0.008	0.004	0.007	0.003
	(S, L)	0.009	0.005	0.038	0.042	0.007	0.005	0.018	0.042
	(L, S)	0.003	0.006	0.010	0.002	0.006	0.010	0.013	0.008
	(L, L)	0.004	0.005	0.016	0.003	0.004	0.003	0.018	0.006

Note: Approximation used 200 equally spaced points. 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average.

Table D4: MSE of the Density Approximation Under Shirking

Main Sample		Pre-SOX		Post-SOX	
Sector	Firm Type	Bad	Good	Bad	Good
Primary	(S, S)	(0.007, 0.008)	(0.002, 0.004)	(0.002, 0.004)	(0.023, 0.025)
	(S, L)	(0.005, 0.007)	(0.008, 0.010)	(0.071, 0.074)	(0.053, 0.056)
	(L, S)	(0.004, 0.005)	(0.002, 0.003)	(0.012, 0.013)	(0.056, 0.109)
	(L, L)	(0.015, 0.017)	(0.010, 0.012)	(0.017, 0.019)	(0.005, 0.007)
Consumer goods	(S, S)	(0.001, 0.002)	(0.002, 0.009)	(0.001, 0.006)	(0.009, 0.022)
	(S, L)	(0.007, 0.008)	(0.005, 0.006)	(0.027, 0.062)	(0.010, 0.012)
	(L, S)	(0.003, 0.006)	(0.004, 0.011)	(0.004, 0.016)	(0.012, 0.023)
	(L, L)	(0.005, 0.006)	(0.004, 0.005)	(0.007, 0.008)	(0.005, 0.009)
Service	(S, S)	(0.005, 0.008)	(0.002, 0.003)	(0.007, 0.013)	(0.002, 0.007)
	(S, L)	(0.015, 0.020)	(0.006, 0.008)	(0.061, 0.071)	(0.036, 0.047)
	(L, S)	(0.004, 0.007)	(0.004, 0.018)	(0.013, 0.015)	(0.002, 0.002)
	(L, L)	(0.003, 0.007)	(0.005, 0.016)	(0.012, 0.016)	(0.003, 0.004)

Extended Sample		Pre-SOX		Post-SOX	
Sector	Firm Type	Bad	Good	Bad	Good
Primary	(S, S)	(0.008, 0.009)	(0.003, 0.003)	(0.004, 0.005)	(0.037, 0.039)
	(S, L)	(0.007, 0.007)	(0.008, 0.009)	(0.045, 0.046)	(0.050, 0.052)
	(L, S)	(0.005, 0.005)	(0.002, 0.003)	(0.015, 0.015)	(0.006, 0.008)
	(L, L)	(0.017, 0.018)	(0.012, 0.013)	(0.012, 0.012)	(0.005, 0.006)
Consumer goods	(S, S)	(0.001, 0.002)	(0.001, 0.002)	(0.002, 0.005)	(0.001, 0.002)
	(S, L)	(0.008, 0.008)	(0.006, 0.007)	(0.134, 0.135)	(0.025, 0.027)
	(L, S)	(0.002, 0.003)	(0.003, 0.005)	(0.019, 0.021)	(0.008, 0.012)
	(L, L)	(0.004, 0.005)	(0.004, 0.004)	(0.010, 0.011)	(0.008, 0.008)
Service	(S, S)	(0.009, 0.010)	(0.003, 0.003)	(0.009, 0.011)	(0.002, 0.003)
	(S, L)	(0.013, 0.014)	(0.007, 0.007)	(0.027, 0.027)	(0.036, 0.040)
	(L, S)	(0.007, 0.007)	(0.008, 0.012)	(0.027, 0.027)	(0.004, 0.004)
	(L, L)	(0.004, 0.005)	(0.004, 0.005)	(0.018, 0.019)	(0.002, 0.003)

Note: This table reports the confidence region of MSE or the approximation of $g_{st}(x)f_{st}(x)$ by a truncated normal distribution. The confidence region is bounded by the minimum and maximum value of the MSE for the identified set of γ that requires $\alpha_{j=1,2}$ to be invariant with bond price and the moral hazard costs to be nonnegative. 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 return is classified as "Good (Bad)" if it is greater (less) than the industry average.

Table D5: Administrative Cost (Extended Sample)

Sector	(A,C)	τ_1	$\Delta\tau_1$	
Primary	(S,S)	(2191, 2330)	(1398, 1406)	+
	(S,L)	(1341, 1404)	(2570, 2582)	+
	(L,S)	(4466, 4599)	(2780, 2910)	+
	(L,L)	(4311, 4398)	(2921, 3012)	+
Consumer Goods	(S,S)	(1764, 2258)	(-679, -555)	-
	(S,L)	(1235, 1363)	(949, 972)	+
	(L,S)	(4393, 5072)	(1724, 2113)	+
	(L,L)	(6708, 7353)	(-214, -130)	-
Service	(S,S)	(2639, 2990)	(1036, 1065)	+
	(S,L)	(2641, 2875)	(-76, -29)	-
	(L,S)	(9803, 10689)	(-1756, -1624)	-
	(L,L)	(9441, 9949)	(-1240, -1218)	-

Note: Cost are measured in thousands of 2006 US$. $\tau_1 \equiv \gamma^{-1}\frac{b_{t+1}}{b_t-1}\ln\alpha_{2,pre}$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table D6: Aggregate Agency Costs (Extended Sample)

Sector	(A, C)	τ_2	$\Delta\tau_2$	
Primary	(S, S)	(72, 210)	(3, 11)	+
	(S, L)	(33, 96)	(-18, -6)	-
	(L, S)	(68, 201)	(68, 198)	+
	(L, L)	(45, 132)	(47, 137)	+
Consumer goods	(S, S)	(259, 753)	(-189, -65)	-
	(S, L)	(67, 195)	(13, 37)	+
	(L, S)	(357, 1036)	(192, 582)	+
	(L, L)	(340, 985)	(55, 140)	+
Service	(S, S)	(183, 534)	(-44, -15)	-
	(S, L)	(122, 356)	(24, 71)	+
	(L, S)	(460, 1345)	(95, 227)	+
	(L, L)	(265, 772)	(10, 32)	+

Note: Cost are measured in thousands of 2006 US$.$\tau_2 \equiv \sum_{s=1}^{2}\varphi_{s,pre}E_{s,pre}\left[w_{s,pre}(x)\right] - \tau_1$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table D7: Welfare Costs of Moral Hazard (Extended Sample)

Sector	(A,C)	τ_3	$\Delta\tau_3$	
Primary	(S,S)	(172, 502)	(97, 188)	+
	(S,L)	(60, 144)	(-1, 38)	=
	(L,S)	(6, 184)	(154, 532)	+
	(L,L)	(301, 689)	(-118, -65)	-
Consumer Goods	(S,S)	(1060, 3294)	(423, 533)	+
	(S,L)	(149, 479)	(-274, -121)	-
	(L,S)	(949, 2575)	(624, 1320)	+
	(L,L)	(1478, 3606)	(-2868, -1161)	-
Service	(S,S)	(358, 1017)	(125, 229)	+
	(S,L)	(285, 655)	(61, 150)	+
	(L,S)	(1356, 3381)	(-2417, -603)	-
	(L,L)	(769, 2327)	(-1445, -381)	-

Note: Cost are measured in thousands of 2006 US\$. $\tau_3 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre}[y_{s,pre}(x)] - \tau_1$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table D8: Welfare Costs of Hidden Information (Extended Sample)

(measured in thousands of 2006 US\$)

Sector	(A, C)	τ_4	$\Delta\tau_4$	
Primary	(S, S)	(-292, -101)	(-177, -94)	-
	(S, L)	(-48, -27)	(-56, -5)	-
	(L, S)	(17, 62)	(-334, -86)	-
	(L, L)	(-558, -257)	(112, 255)	+
Consumer goods	(S, S)	(-2541, -802)	(-674, -488)	-
	(S, L)	(-284, -82)	(135, 312)	+
	(L, S)	(-1539, -592)	(-738, -432)	-
	(L, L)	(-2621, -1139)	(1217, 3008)	+
Service	(S, S)	(-483, -175)	(-273, -140)	-
	(S, L)	(-299, -163)	(-126, 10)	=
	(L, S)	(-2036, -896)	(698, 2644)	+
	(L, L)	(-1554, -504)	(391, 1477)	+

Note: Cost are measured in thousands of 2006 US\$. $\tau_4 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre}\left[w_{s,pre}(x) - y_{s,pre}(x)\right]$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

Table D9: Gross Losses to the Shareholders Firms the CEO from Shirking (Extended Sample)

Sector	(A, C)	ρ_1	$\Delta\rho_1$	
Primary	(S, S)	(11.94, 12.07)	(-0.59, -0.51)	-
	(S, L)	(12.53, 12.90)	(-5.40, -5.08)	-
	(L, S)	(10.21, 10.68)	(-2.57, -2.49)	-
	(L, L)	(5.90, 6.09)	(-1.02, -0.94)	-
Consumer Goods	(S, S)	(18.30, 18.51)	(-9.33, -9.25)	-
	(S, L)	(10.60, 10.61)	(11.81, 12.70)	+
	(L, S)	(9.15, 10.03)	(-1.43, -0.95)	-
	(L, L)	(7.52, 8.13)	(-2.84, -2.02)	-
Service	(S, S)	(17.44, 17.57)	(-3.80, -3.43)	-
	(S, L)	(12.99, 13.74)	(-6.58, -5.92)	-
	(L, S)	(18.61, 18.91)	(-12.46, -11.48)	-
	(L, L)	(9.96, 10.73)	(-7.06, -6.80)	-

Note: Gross losses are measured in percentage. $\rho_1 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre} \{x [1 - g_{s,pre}(x)]\}$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price (16.4).

Table D10: Compensating Differential from CEO Shirking versus Working (Extended Sample)

(measured in thousands of 2006 US$)

Sector	(A, C)	ρ_2	$\Delta\rho_2$	
Primary	(S, S)	(2792, 2995)	(51, 125)	+
	(S, L)	(1587, 1668)	(31, 51)	+
	(L, S)	(2303, 2471)	(1085, 1163)	+
	(L, L)	(1983, 2073)	(364, 472)	+
Consumer Goods	(S, S)	(7671, 8438)	(-1992, -1766)	-
	(S, L)	(2224, 2400)	(317, 368)	+
	(L, S)	(5353, 6165)	(1623, 1869)	+
	(L, L)	(4687, 5465)	(-2264, -2218)	-
Service	(S, S)	(4608, 5074)	(50, 87)	+
	(S, L)	(2542, 2840)	(23, 97)	+
	(L, S)	(8758, 9929)	(-6960, -6722)	-
	(L, L)	(5916, 6610)	(-3985, -3885)	-

Note: Differentials are measured in thousands of 2006 US$. $\rho_2 \equiv b_{t+1} [(b_t - 1)\gamma]^{-1} \ln(\alpha_{2,pre}/\alpha_{1,pre})$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.

TABLE D11: CHANGE IN WELFARE COSTS OF MORAL HAZARD CAUSED BY SIGNAL QUALITY
(EXTENDED SAMPLE)

Sector	(A,C)	$\tau_3(\alpha_{j,post}, g_{s,pre})$	ρ_3	
Primary	(S,S)	(179, 551)	(90, 139)	+
	(S,L)	(62, 153)	(-3, 29)	=
	(L,S)	(75, 468)	(85, 248)	+
	(L,L)	(429, 973)	(-401, -194)	-
Consumer Goods	(S,S)	(548, 1723)	(936, 2025)	+
	(S,L)	(207, 643)	(-438, -179)	-
	(L,S)	(1628, 4784)	(-889, -55)	-
	(L,L)	(542, 839)	(-250, -101)	-
Service	(S,S)	(366, 1061)	(117, 185)	+
	(S,L)	(302, 667)	(50, 133)	+
	(L,S)	(143, 184)	(609, 782)	+
	(L,L)	(107, 250)	(282, 632)	+

Note: Costs are measured in thousands of 2006 US\$. $\rho_3 \equiv \tau_3(\alpha_{j,post}, g_{s,post}) - \tau_3(\alpha_{j,post}, g_{s,pre})$. Here "+" ("-") means that the change is positive (negative), and "=" means we cannot reject the null hypothesis of no change. The confidence region is estimated for the single common bond price, 16.4.