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

January 14, 2022

Abstract

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

A Data construction details

A.1 Firm type

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. 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.

Binary variables based on firm size and capital structure (debt-to-equity ratio) also categorize the firm types. Firm size is measured by the total assets on a firm's balance sheet (A). The capital structure is reflected by the debt-to-equity ratio (C). The numerator of the ratio is the total liabilities and the denominator is the total common equity. 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 within a sector 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 if $(A, C) = (S, L)$ then the firm had lower total assets and a higher debt-to-equity ratio than the median. The calculations are made in the first year the sample or more generally the first year a firm joins the sample; this classification system implies firms do not change categories.

A.2 Accounting return

Accruals and deferrals are examples of accounting features often used to convey information to shareholders about the state of the firm. Managers exercise some discretion over how to report these items. More generally, accounting numbers reveal information to shareholders that managers are either obligated or choose to disclose. In our model, CEOs collect and convey their private information on the firm's prospects after accepting a contractual arrangement with the firm. We construct an empirical measure of the report by accounting returns to assets evaluated at book value, consistent with the concept of comprehensive income in accounting practice. Let $\widetilde{s}_{it}$ the accounting return for firm i at year t, calculated as

$$\widetilde{s}_{it} \equiv \left(Asset_{it} - Debt_{it} + Dividend_{it}\right) / \left(Asset_{i,t-1} - Debt_{i,t-1}\right)$$

where $Asset_{it}$ is the total assets at the end of year t, $Debt_{it}$ is the total liability minus minority interest, $Dividend_{it}$ is the dividend to common stock plus the dividend to preferred stock. All variables are deflated to base year 2006 before calculating the accounting return. Specifically, we define the binary private state, denoted as $s_{it} \in \{1, 2\}$, where $s_{it} = 1$ means that in year t the accounting return $\widetilde{s}_{it}$ is lower for firm i than the average for all firms within the same sector, size, and capital structure categories, while $s_{it} = 2$ means the reverse.

A.3 Excess return

Firm-year observations (i, t) are dropped if the firm changed its fiscal year end. Consequently all the observations in the empirical analysis are based on a 12 month period, and therefore comparable. Raw stock prices and adjustment factors are from the Compustat PDE dataset. For each firm-year in the sample, we calculate π_{it}, the monthly compounded stock returns adjusted for splitting and repurchasing for the fiscal year. Then we subtract π_t, the return to a value-weighted market portfolio (NYSE/NASDAQ/AMEX), from π_{it} to obtain $\widetilde{x}_{it} \equiv \pi_{it} - \pi_t$, the net excess returns to firm i at year t. The excess return is the sum of $\widetilde{x}_{it}$ plus total CEO compensation for the corresponding year, scaled by firm's value. Denoting by $\widetilde{w}_{it}$ denote total compensation to the CEO, and V_{it} the beginning firm value of firm i in the same fiscal year, we compute the excess return as $x_{it} \equiv \widetilde{x}_{it} + \widetilde{w}_{it}/V_{it}$.

A.4 Components of compensation

Following the concept of income-equivalent total compensation adopted by Antle and Smith [1985, 1986], Hall and Liebman [1998], and Margiotta and Miller [2000], we construct a measure of total compensation by adding change in wealth from options held and stocks held to the other components of compensation included in ExecuComp. Thus, in addition to the total compensation included in Compustat ExecuComp, we also calculate the holding value of firm-specific equities.

To evaluate the call options for each CEO in each year, we apply the dividend-adjusted Black-Scholes formula as follows. Let $CALL$ denote the call option value, $PRICE$ the exercise price, M the time to maturity in years, SEC the underlying security price, Y the dividend yield, R the risk-free return, and V the implied volatility. Let $\mathcal{N}(\cdot)$ denote the standard normal cumulative distribution function. Then the call option value is given by the formula

$$CALL = e^{-YM}\mathcal{N}(\mathfrak{d}_1)SEC - \exp(-RM)\,\mathcal{N}(\mathfrak{d}_1 - V\sqrt{M})PRICE,$$

where:

$$\mathfrak{d}_1 \equiv \frac{\ln(SEC) - \ln(PRICE) + (R - Y + V^2/2)M}{V\sqrt{M}}.$$

We do not observe all the inputs of the Black-Scholes formula for grants made before 1993, the first year of our sample. Compustat ExecuComp only provides the valuation information for those options newly granted after 1993, including the number of underlying stock shares, exercise prices, expiration dates, and issue dates. To accurately value the wealth change of a CEO, however, we must estimate the value of unexercised options and update it each year, including those options granted before 1993. To facilitate the calculation, we assume that (i) no option is exercised until expiration dates; (ii) stock options granted before 1993 are exercised in a FIFO fashion; and (iii), each CEO holds stock options granted before 1993 for a fixed period equal to the average length of the holding period across all years when he is in the sample. These assumptions suffice to interpolate the issue dates and exercised prices for options granted before 1993. The same procedure is used to impute firm specific options granted before the CEO enters our sample.

A.5 Bond prices

In our theoretical model the bond price equals the present value of a security paying one consumption unit annually for a given number of years (including in perpetuity). By assumption future interest rates and bond prices are known. In our empirical work we set the horizon at 30 years. Denoting by ι_t the one period real interest rate at t, the bond price b_t at t is

$$b_t = \sum_{s=1}^{30} {}_{k=0}^{s} \frac{1}{1 + \iota_{t+k}}$$

The Treasury Bill yield curve is interpolated to obtain nominal interest rates. We deflate all nominal prices $\widetilde{p}_t$, including bond prices, to a base year of 2006, defining $p_t = \widetilde{p}_t / deflator2006$. Table S-2 Panel A shows real bond prices b_t and b_{t+1} imputed for each year $t \in \{1993, ..., 2005\}$.

When estimating the confidence region for the structural parameters of the model and the welfare measures, we focus on bond prices in the pre- and post-SOX eras that apply to both eras, constructing a grid of five bond prices

$$b' \in \{15.55, 16.1, 16.4, 16.8, 17.8\}$$

that approximate bond prices in the pre-SOX era, and two $b' \in \{16.4, 16.8\}$ that roughly characterize bond prices in the post-SOX era, as indicated in Table S-2 Panel B. For example $b_{t+1} = 15.40$ in fiscal year 2000, which roughly corresponds to setting $b' = 15.55$.

The sets are obtained as follows. The average of the two extreme values on each end determine the grid boundary. For the pre-SOX era, 15.55 is the average of 15.40 and 15.69, and 17.8 is the average of 17.19 and 18.38. For the post-SOX era, 16.4 is the average of 16.33 and 16.45, and 16.80 is the average of 16.45 and 17.10. Note that 16.4 and 16.8 fall into the range of the pre-SOX bond prices too, so those prices are kept as grid points for the pre-SOX era as well. The remaining bond price in the five point grid of the pre-SOX era, 16.1, is the average of the other (mid-value) bond prices, namely 15.91, 16.00, 16.19, 16.21, and 16.21; it roughly bisects the distance between 15.55 and 16.4, its adjacent values. Panel A of Table S2 refers.

We predict the grid values of b as a quadratic function of b' running a least squares regression of b_t on b_{t+1} without an intercept, but including linear and quadratic terms. The coefficients (t-statistics) are 1.9271 (6.58) for the linear term and -0.056154 (-3.17) for the quadratic term. The predicted values of b from the estimated regression along with the corresponding value of b' are reported in the the right two columns of Panel B in Table S-2. Panel B of Table S-2 shows the years 1997-98 and 2000-01 are treated as having the same bond price pair, namely $(15.55, 16.39)$, even though the bond price pair in 1997-98 is $(15.69, 16.00)$ is not identical to $(15.40, 16.19)$, the pair in 2000-01.

A.6 Trimming and labelling

Recalling the notation in the text, $\mathcal{Z}$ denotes the Cartesian product of categorical variables used to partition firms, namely the three sectors, two firm sizes and two levels of financial leverage. Let $p \in \{1, 2\}$ designate the era, where $p = 1$ indicates the pre-SOX era and $p = 2$ indicates the post-SOX era. With regards bond prices, we note that (i) since $b = \delta(b')$ by assumption, the compensation plan at a given point in time only depends on one bond price, b', not two, and (ii) the set of bond prices considered are interpolated as described in Appendix A.5 and laid out in Panel B of Table S.2 with their corresponding years. In the description of our empirical work, we label firm-year observations (i, t) by $n \in \{1, \dots, N\}$.

The data on observation n is $(x_n, \widetilde{w}_n, b_n, s_n, z_n, p_n)$ where x_n denotes excess returns, $\widetilde{w}_n$ denotes total compensation measured with independent and identically distributed error, b_n is the current bond price imputed from Appendix A.5 corresponding to the year compensation was paid, s_n is the accounting state of the firm, z_n is the firm category (given by sector size and debt/equity ratio), and p_n is an indicator for the era (pre- or post-SOX).

Several components of our test statistics are nonparametrically estimated density or regression functions that condition the population and the sample on (s, z, p). These components are computed off 24 subsamples of length $N_{s,z,p} \equiv \sum_{i=1}^{N} 1\{(s, z, p) = (s_n, z_n, p_n)\}$ that partition the sample by (s, z, p). To elucidate our estimation and testing procedures, we annotate several of the variables by (s, z, p) to indicate the subsample they belong to. Thus $x_n^{(s,z,p)}$ is uniquely associated with the excess return of a particular m^{th} observation in the data that has excess return $x_m \equiv x_n^{(s,z,p)}$, accounting return s_m, firm category z_m and era p_m. We label measured compensation in a similar way by $\widetilde{w}_n^{(s,z,p)}$, and define a coarser partition based on (s, z) with analogous annotation: for example $x_n^{(s,z)}$ is uniquely associated with the excess return of a particular observation m' in the data that has excess return $x_{m'} \equiv x_n^{(s,z,p)}$, accounting return s_m and firm category z_m. Note that $x_n^{(s,z,p)} = x_n^{(s,z)}$ with probability one, the excess returns referring to the same observation, because the probability that $x_m = x_{m'}$ is zero if $m \neq m'$.

To eliminate the influence of outliers we trim the sample for the purposes of structural estimation. Let $\min\{x\}$ denote the value of excess returns corresponding to the 2.5 percentile in the distribution of excess returns, and $\max\{x\}$ denote the excess return value corresponding to the 97.5 percentile. For the purposes of structural estimation we included only those observations for which $\min\{x\} \leq x_n \leq \max\{x\}$ for $i \in \{1, \ldots, N\}$.

B Testing for structural change

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

B.1 Change in the density of excess returns

Let $f_s^p(x|z)$ denote the probability density function of excess returns in the pre-SOX era conditional on firm category $z \in \mathcal{Z}$, state $s \in \{1, 2\}$ and era $p \in \{1, 2\}$. The null hypothesis of no change is that $f_s^1(x|z) = f_s^2(x|z)$ for all $(x, s, z) \in \mathcal{R} \times \{1, 2\} \times \mathcal{Z}$. The test of equality (for each z) is based on the summing estimates of $\left[f_s^1(x|z) - f_s^2(x|z)\right]^2$ over observations in subsamples designated by $(s,z,1)$ and $(s,z,2)$. Following Li and Racine [2007, page 363] we calculate, for each (s,z), the statistic $T_{s,z}^{pdf}$ by

$$T_{s,z}^{pdf} = \widehat{\sigma}_{x,s,z}^{-1}(N_{s,z,1}N_{s,z,2}h_{s,z})^{1/2}(\widehat{\zeta}_{s,z} - \widehat{c}_{s,z}),$$

where

$$\begin{aligned}
\widehat{\zeta}_{s,z} \quad = \quad & \frac{1}{\sqrt{2\pi}h_{s,z}N_{s,z,1}^2}\sum_{m=1}^{N_{s,z,1}}\sum_{n=1}^{N_{s,z,1}}\exp\left[-\frac{1}{2}\left(\frac{x_m^{(s,z,1)}-x_n^{(s,z,1)}}{h_{s,z}}\right)^2\right] \\
& +\frac{1}{\sqrt{2\pi}h_{s,z}N_{s,z,2}^2}\sum_{m=1}^{N_{s,z,2}}\sum_{n=1}^{N_{s,z,2}}\exp\left[-\frac{1}{2}\left(\frac{x_m^{(s,z,2)}-x_n^{(s,z,2)}}{h_{s,z}}\right)^2\right] \\
& -\frac{2}{\sqrt{2\pi}h_{s,z}N_{s,z,1}N_{s,z,2}}\sum_{m=1}^{N_{s,z,1}}\sum_{n=1}^{N_{s,z,2}}\exp\left[-\frac{1}{2}\left(\frac{x_m^{(s,z,1)}-x_n^{(s,z,2)}}{h_{s,z}}\right)^2\right],
\end{aligned}$$

$$\widehat{c}_{s,z} = \left(h_{s,z}\sqrt{2\pi}\right)^{-1}\left(N_{s,z,1}^{-1} + N_{s,z,2}^{-1}\right),$$

and

$$\begin{aligned}
\widehat{\sigma}_{x,s,z}^2 \quad = \quad & \frac{1}{2\pi h_{s,z}N_{s,z,1}^2}\sum_{m=1}^{N_{s,z,1}}\sum_{n=1}^{N_{s,z,1}}\exp\left[-\frac{1}{4}\left(\frac{x_m^{(s,z,1)}-x_n^{(s,z,1)}}{h_{s,z}}\right)^4\right] \\
& +\frac{1}{2\pi h_{s,z}N_{s,z,2}^2}\sum_{m=1}^{N_{s,z,2}}\sum_{n=1}^{N_{s,z,2}}\exp\left[-\frac{1}{4}\left(\frac{x_m^{(s,z,2)}-x_n^{(s,z,2)}}{h_{s,z}}\right)^4\right] \\
& +\frac{1}{\pi h_{s,z}N_{s,z,1}N_{s,z,2}}\sum_{m=1}^{N_{s,z,1}}\sum_{n=1}^{N_{s,z,2}}\exp\left\{-\frac{1}{4}\left(\frac{x_m^{(s,z,1)}-x_n^{(s,z,2)}}{h_{s,z}}\right)^4\right\}
\end{aligned}$$

The bandwidth $h_{s,z}$ is calculated following Silverman's rule of thumb, conditional on $z \in \{1,2\} \times \mathcal{Z}$ but margining over the two eras. Similarly, $\widehat{\sigma}_{s,z}$ is the estimated standard deviation of excess returns margining over both eras. The test statistic $T_{s,z}^{pdf}$, asymptotically distributed as standard normal distribution under the null hypothesis that $f_s^1(x|z) = f_s^2(x|z)$, is one-sided. Hence the null is rejected at the 1% significance level if $T_{s,z}^{pdf} > 2.33$, at the 5% significance level if $T_{s,z}^{pdf} > 1.64$, and at the 10% significance level if $T_{s,z}^{pdf} > 1.28$.

B.2 Change in the compensation schedule

Denote by $w^{(s,z,p)}(x)$ the optimal CEO compensation schedule as a function of (x,s,z,p). Note that

$$w_n^{(s,z,p)} \equiv w^{(s,z,p)}\left(x_n^{(s,z,p)}\right) = E\left[\widetilde{w}_n^{(s,z,p)}\right]$$

because the difference $\widetilde{w}_n^{(s,z,p)} - w_n^{(s,z,p)}$ is identically and independently distributed measurement error with mean zero. Define the nonparametric (standard normal) kernel estimator with bandwidth $h_{s,z,2}$ for $w_n^{(s,z,p)}$ by

$$\widehat{w}_n^{(s,z,2)} \equiv \sum_{m=1}^{N_{s,z,2}} \widetilde{w}_m^{(s,z,2)} \exp\left[-\tfrac{1}{2}\left(\frac{x_m^{(s,z,2)} - x_n^{(s,z,2)}}{h_{s,z,2}}\right)^2\right] \Bigg/ \sum_{m=1}^{N_{s,z,2}} \exp\left[-\tfrac{1}{2}\left(\frac{x_m^{(s,z,2)} - x_n^{(s,z,2)}}{h_{s,z,2}}\right)^2\right]$$

Similarly $\widehat{w}_n^{(s,z)}$, a nonparametric (standard normal) kernel estimator for $w_n^{(s,z)} \equiv E\left[\widetilde{w}\,|x_n, s_n, z_n\right]$ with bandwidth $h_{s,z}$, is defined as

$$\widehat{w}_n^{(s,z)} = \sum_{m=1}^{N_{s,z}} \widetilde{w}_m^{(s,z)} \exp\left[-\tfrac{1}{2}\left(\frac{x_m^{(s,z)} - x_n^{(s,z)}}{h_{s,z}}\right)^2\right] \Bigg/ \sum_{m=1}^{N_{s,z,2}} \exp\left[-\tfrac{1}{2}\left(\frac{x_m^{(s,z)} - x_n^{(s,z)}}{h_{s,z}}\right)^2\right].$$

Testing the null hypothesis that $w^{(s,z,1)}(x) = w^{(s,z,2)}(x)$ amounts to a test of whether $w^{(s,z)}(x) = w^{(s,z,2)}(x)$, that the compensation schedule is not affected by conditioning on the era. It is based on the sum of squared differences of nonparametric estimates of $w_n^{(s,z,2)}$ and $w_n^{(s,z)}$, adjusted for trimming the sample. Define the trimming indicator variable $A_n^{(s,z)}$, by culling the observations for each subsample defined by (s,z) with outlier excess returns: $A_n^{(s,z)} = 1$ if $x_n^{(s,z)}$ falls within the 2.5% to 97.5% range of excess returns for those observations belonging to the (s,z) subsample, and $A_n^{(s,z)} = 0$ otherwise. Following Aït-Sahalia, Bickel, and Stoker [2001] the test statistic T_d^W is defined as

$$T_{s,z}^W = \widehat{\sigma}_{w,s,z}^{-1}\left[\sum_{n=1}^{N_{s,z,2}} \left(\widehat{w}_n^{(s,z,2)} - \widehat{w}_n^{(s,z)}\right)^2 \frac{A_n^{(s,z)}}{N_{s,z}} - \sum_{n=1}^{N_{s,z,2}} \frac{\widehat{\sigma}_{n,w,s,z,2}^2}{\widehat{f}_n^{(s,z,2)}} \frac{A_n^{(s,z)}}{4\pi h_{s,z,2} N} - \sum_{n=1}^{N_{s,z}} \frac{\widehat{\sigma}_{n,w,s,z}^2}{\widehat{f}_n^{(s,z)}} \frac{h_{s,z,2}\widehat{A}_n^{(s,z)}}{2\sqrt{\pi} h_{s,z} N}\right],$$

where $\widehat{A}_n^{(s,z)}$ is a nonparametric estimator for $E\left[A_n^{(s,z)}\,|x_n, s_n, z_n\right]$ with a standard normal kernel and bandwidth $h_{s,z}$, defined as

$$\widehat{A}_n^{(s,z)} = \sum_{m=1}^{N_{s,z}} A_m^{(s,z)} \exp\left[-\frac{1}{2}\left(\frac{x_m^{(s,z)} - x_n^{(s,z)}}{h_{s,z}}\right)^2\right] \Bigg/ \sum_{m=1}^{N_{s,z}} \exp\left[-\frac{1}{2}\left(\frac{x_m^{(s,z)} - x_n^{(s,z)}}{h_{s,z}}\right)^2\right].$$

the densities of excess returns, $f_n^{(s,z,2)}$ and $f_n^{(s,z)}$, are estimated nonparametrically by

$$\widehat{f}_n^{(s,z,2)} = \frac{1}{\sqrt{2\pi} N_{s,z,2} h_{s,z,2}} \sum_{m=1}^{N_{s,z,2}} \exp\left[-\frac{1}{2}\left(\frac{x_m - x_n}{h_{s,z,2}}\right)^2\right]$$

$$\widehat{f}_n^{(s,z)} = \frac{1}{\sqrt{2\pi} N_{s,z} h_{s,z}} \sum_{m=1}^{N_{s,z}} \exp\left[-\frac{1}{2}\left(\frac{x_m - x_n}{h_{s,z}}\right)^2\right].$$

and estimates of the conditional variance terms are given by

$$\widehat{\sigma}^2_{n,w,s,z,2} = \sum_{m=1}^{N_{s,z,2}} \left[\left(\widetilde{w}_m^{(s,z,2)}\right)^2 - \left(\widehat{w}_m^{(s,z,2)}\right)^2\right] \exp\left\{-\frac{1}{2}\left(\frac{x_m - x_n}{h_{s,z,2}}\right)^2\right\} \Big/ \sum_{m=1}^{N_{s,z,2}} \exp\left\{-\frac{1}{2}\left(\frac{x_m - x_n}{h_{s,z,2}}\right)^2\right\}$$

$$\widehat{\sigma}^2_{n,w,s,z} = \sum_{m=1}^{N_{s,z}} \left[\left(\widetilde{w}_m^{(s,z)}\right)^2 - \left(\widehat{w}_m^{(s,z)}\right)^2\right] \exp\left\{-\frac{1}{2}\left(\frac{x_m - x_n}{h_{s,z}}\right)^2\right\} \Big/ \sum_{m=1}^{N_{s,z}} \exp\left\{-\frac{1}{2}\left(\frac{x_m - x_n}{h_{s,z}}\right)^2\right\}$$

$$\widehat{\sigma}^2_{w,s,z} = \frac{1}{4\pi N_{s,z}} \sum\nolimits_{n=1}^{N_{s,z}} \frac{\left(\sigma^2_{n,w,s,z}\right)^2 A^2_{n,d}}{\widehat{f}_n^{(s,z)}}.$$

This one-sided test statistic is asymptotically distributed standard normal with mean 0 and variance 1. Thus the null hypothesis is rejected at the (i) 1% significance level if $T_d^W > 2.33$, (ii) 5% significance level if $T_d^W > 1.64$, and (iii) 10% significance level if $T_d^W > 1.28$.

C Identification and estimation details

This appendix presents the details of identification, structural estimation, and the counterfactual analyses.

C.1 Identification

This subsection set identifies the risk aversion parameter, γ. Given a pair of bond prices pertaining to the current period and the the next one, denoted by (b, b'), the following set of restrictions places limits on the observationally equivalent values of γ. To reduce the notational clutter we follow the text in suppressing the dependence of these restrictions on (b, b'). After explaining how these restrictions utilized in identification, we show how variation in the bond prices generates additional restrictions.

The model requires that at least one of the truth-telling constraint and the sincerity constraint should be binding. This implies that

$$\Psi_3(\gamma) \equiv E_2\left[v_1(x,\gamma) - v_2(x,\gamma)\right] \geq 0,$$

$$\Psi_4(\gamma) \equiv E_2\left[\alpha_1(\gamma)^{1/(b-1)} v_1(x,\gamma) g_2(x,\gamma) - \alpha_2(\gamma)^{1/(b-1)} v_2(x,\gamma)\right] \geq 0,$$

and

$$\Psi_5(\gamma) \equiv \Psi_3(\gamma)\Psi_4(\gamma) = 0.$$

The Kuhn-Tucker multipliers, $\eta_3(\gamma)$ and $\eta_4(\gamma)$, used in the definition of $g_1(x,\gamma)$, are non-negative. The complementary slackness conditions for the truth-telling and sincerity constraints must also be satisfied, implying $\Psi_6(\gamma) \equiv \Psi_3(\gamma)\eta_3(\gamma) = 0$ and $\Psi_7(\gamma) \equiv \Psi_4(\gamma)\eta_4(\gamma) = 0$. We impose another exclusion restriction, that α_1 does not depend on the private state, yielding

$$\begin{aligned}\Psi_1(\gamma) \quad \equiv \quad & E\left[v_s(x,\gamma)\right]^{-1} - E_1[v_1(x,\gamma)]^{-1} - \eta_3(\gamma)E_1[h(x)v_1(x,\gamma)]E_1[v_1(x,\gamma)]^{-1} \\ & -\eta_4(\gamma)\left[\tfrac{\alpha_1(\gamma)}{\alpha_2(\gamma)}\right]^{1/(b_t-1)} E_1[g_2(x,\gamma)h(x)v_1(x,\gamma)]E_1[v_1(x,\gamma)]^{-1} = 0.\end{aligned}$$

In addition, the likelihood $g_1(x,\gamma)$ is positive, implying

$$\Psi_2(\gamma) \equiv E_1\left[1\left\{g_1(x,\gamma) > 0\right\} - 1\right] = 0,$$

and by definition the expected value of $g_1(x,\gamma)$ is one, so

$$\Psi_8(\gamma) \equiv E_1\left[g_1(x,\gamma)\right] - 1 = 0$$

Since the data show the CEO is not paid a fixed wage in either state, we infer it must be optimal for shareholders to induce the CEO works in both private states rather than shirk in either of them. The corresponding inequality restriction is:

$$\Lambda(\gamma) \equiv \sum\nolimits_{s=1}^{2} \varphi_s\left\{E_s[V_s x - w_s(x)] - E_s\left[V_s x g_s(x,\gamma) - \tfrac{b'}{b-1}\gamma^{-1}\ln[\alpha_1(\gamma)]\right]\right\} \geq 0.$$

It is evident from the inequalities and equalities laid out above, that the collection of the restrictions defines a Borel set of the risk aversion parameters that depend on (b,b'). We now acknowledge the dependence explicitly by writing

$$\Gamma(b,b') \equiv \left\{\gamma > 0 : \begin{array}{l} \Lambda(\gamma;b,b') \geq 0 \\ \Psi_j(\gamma;b,b') = 0 \text{ for } j \in \{1,2,5,\ldots,8\} \\ \Psi_k(\gamma;b,b') \geq 0 \text{ for } k \in \{3,4\} \\ \eta_k(\gamma;b,b') \geq 0 \text{ for } k \in \{3,4\} \end{array}\right\}.$$

Variation in bond prices over time yield additional sets of restrictions. The intersection of these sets is defined as

$$\Gamma \equiv \bigcap_{(b,b')\in B^2} \Gamma(b,b') = \{\gamma > 0 : Q(\gamma) = 0\}$$

where B^2 denotes the set of distinct pairs of bond prices generated by the data, the criterion function $Q(\gamma)$ is defined as

$$Q(\gamma) \equiv \sum_{(b,b')\in B^2} Q(\gamma; b, b') \tag{1}$$

and

$$\begin{aligned} Q(\gamma; b, b') \quad \equiv \quad & \min\left[0, \Lambda(\gamma; b, b')\right]^2 + \sum\nolimits_{j\in\{1,2,5,\ldots,8\}} \Psi_j^2(\gamma; b, b') \\ & + \sum\nolimits_{k=3}^{4} \left\{ \min\left[0, \Psi_k(\gamma; b, b')\right]^2 + \min\left[0, \eta_k(\gamma; b, b')\right]^2 \right\}. \end{aligned}$$

C.2 Estimation

The structural estimation can be divided into three stages. We estimate the optimal compensation plan for each sector, pre- and post-SOX, and bond prices, then the density of excess returns when the CEO works conditional on the same variables, both of which are inputs into the last stage of obtaining a consistent estimate of the identified set for the risk aversion parameters.

Compensation Conditioning on the accounting state of the firm state $s \in \{1, 2\}$, firm category $z \in \mathcal{Z}$, the era (pre- and post-SOX) denoted by $p \in$ {pre-SOX, post-SOX} and bond prices (b, b'), we estimate the optimal compensation plan in three steps:

1. Select 200 evenly distributed points of excess returns over the range $\min\{x\}$ to $\max\{x\}$, and denote them by x_k for $k \in \{1, .., 200\}$.

2. Estimate $E\left[\widetilde{w}\,|, x_k, b, s, z, p\right]$, total compensation conditional on each value of (x_k, b, s, z, p), with the nonparametric regression estimator $\widehat{w}_k^{(b,s,z,p)}$, obtained with standard normal kernel density function

$$\widehat{w}_k^{(b,s,z,p)} = \frac{\sum_{n=1}^{N} \widetilde{w}_n^{(s,z,p)} K\left(\frac{x_n - x_k}{h_x}\right) K\left(\frac{b_n - b}{h_b}\right)}{\sum_{n=1}^{N} K\left(\frac{x_n - x_k}{h_x}\right) K\left(\frac{b_n - b}{h_b}\right)}$$

3. For each (b, s, z, p) interpolate $\widehat{w}_k^{(b,s,z,p)}$ with a spline formed from $\left\{\widehat{w}_k^{(b,s,z,p)}\right\}_{k=1}^{200}$ to impute total compensation for every observation $n \in \{1, \ldots, N\}$ in the sample, denoted by $\widehat{w}_n$.

Firm performance The probability density function for excess returns from working, conditional on (s, z, p), denoted by $f_s^p(x|z)$, is nonparametrically estimated with a standard normal kernel density function by

$$\widehat{f}_s^p(x|z) = \frac{1}{\sqrt{2\pi} N_{s,z,p} h_{s,z,p}} \sum_{n=1}^{N_{s,z,p}} \exp\left[-\frac{1}{2}\left(\frac{x_n - x}{h_{s,z,p}}\right)^2\right] \tag{2}$$

where $h_{s,z,p}$ is the bandwidth.

Confidence region for γ: Let $Q^{(N)}(\gamma)$ denote a consistent estimate of $Q(\gamma)$, formed from the data of sample size N by substituting sample analogues for their population counterparts that are used in defining $Q(\gamma)$ given in (1). Note that $Q^{(N)}(\gamma)$ is strictly positive only because expectations and limits in the population differ from their respective sample analogues. When x is unbounded, $Q^{(N)}(\gamma)$ converges to zero at rate $N^{2/3}$ (Gayle and Miller, [2015]). We construct a confidence region for the observationally equivalent risk aversion parameters for each specification we investigate, defining

$$\widehat{\Gamma}_\delta \equiv \left\{\gamma > 0 : N^{2/3} Q^{(N)}(\gamma) \leq \widehat{c}_\delta\right\}, \tag{3}$$

where $\widehat{c}_\delta$ is a consistent estimator of c_δ, the critical value of the confidence region for a test of size δ. Asymptotically, the probability that the set of observationally equivalent γ's constructed this way are not contained in Γ is δ.

We modify a subsampling procedure proposed by Chernozhukov, Hong and Tamer [2007] to estimate c_δ. 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 $c_0^{(N)}$ and $\widehat{\Gamma}_0$ as

$$\begin{aligned} c_0^{(N)} &\equiv \inf_{\widetilde{\gamma} > \gamma_N} \left[N^{2/3} Q^{(N)}(\widetilde{\gamma})\right] + \kappa_N \\ \widehat{\Gamma}_0 &\equiv \{\gamma \geq \gamma_N : N^{2/3} Q^{(N)}(\gamma) \leq c_0^{(N)}\}, \end{aligned}$$

where $\kappa_N \propto \ln N$ and γ_N, a strictly positive sequence, converges to zero at a rate faster than

$N^{2/3}$. For each subset $i \in \{1, ..., B_N\}$ of size N_b define

$$C^{(i,N_b)} \equiv \sup_{\gamma \in \widehat{\Gamma}_0} \left[(N_b)^{2/3} Q^{(i,N_b)}(\gamma) \right],$$

and denote by $\widehat{c}_\delta$ the $(1-\delta)$ quantile of the sample $\left\{C^{(1,N_b)}, \ldots, C^{(B_N,N_b)}\right\}$. Substituting $\widehat{c}_\delta$ into (3) we form $\widehat{\Gamma}_\delta$ from $Q^{(N)}(\gamma)$.

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 S-7, 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 S-7 gives economic meaning to the estimates of risk aversion in Table S-7, where the amount a CEO would pay to avoid an equiprobable gamble with losing or winning \$1,000,000.

C.3 Counterfactual analysis

This subsection outlines procedures used to approximate $f_s(x)$ and $g_s(x)$ for conducting the counterfactual analysis. Approximating $f_s(x)$ with a parametric function essentially smooths a nonparametric function. Estimating the likelihood ratio $g_s(x)$, or equivalently $g_s(x)f_s(x)$, is more involved because both functions are only set identified from the curvature of the optimal compensation schedule, which in turn depends on bond prices and the risk aversion parameter.

With regards $f_s(x)$, for each $(s, z, p) \in \{1, 2\} \times \mathcal{Z} \times \{1, 2\}$ we approximate the density for excess returns, when the CEO works, by a truncated normal density, writing

$$\widehat{f}_s(x, x_L, \mu_s, \sigma_s) = \left[\Phi\left(\frac{\mu_s - x_L}{\sigma_s} \right) \sigma_s \sqrt{2\pi} \right]^{-1} \exp\left[-\frac{1}{2} \left(\frac{x - \mu_s}{\sigma_s} \right)^2 \right],$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function, x_L is the lower bound of the support, (μ_s, σ_s) denotes the mean and standard deviation of the parent normal distribution, and the dependence of $\widehat{f}_s(\cdot)$ on (z, p) is suppressed here to reduce notational clutter. The cutoff point for the support, x_L, estimated by the lowest excess return in the sample, denoted by $\widehat{x}_L$, which is a superconsistent estimator for x_L. Then we evaluate the nonparametric estimator for $f_s^{(N)}(x_k)$ using (2) at 200 evenly spaced points $\{x_k\}_{k=1}^{200}$, and minimize the mean squared deviation between $f^{(N)}(x_k)$ and $f_s(x_k, x_L, \mu_s, \sigma_s)$ summed over k with respect to μ_s and σ_s to obtain

$$(\widehat{\mu}_s, \widehat{\sigma}_s) = \underset{(\mu_s,\sigma_s)}{\arg\min} \sum\nolimits_{k=1}^{200} \left[f_s^{(N)}(x_k) - \widehat{f}_s(x_k, \widehat{x}_L, \mu_s, \sigma_s) \right]^2 .$$

Table S-8 presents the Mean Square Errors (MSE) for these approximations. It shows the truncated normal distribution closely approximates the distribution of excess returns.

We approximate $g_s(x)$ by the ratio of two truncated normal distributions with a common lower truncated point x_L. This implies the shirking density, denoted by $\underline{f}_s(x) \equiv g_s(x) f_s(x)$, is approximated by

$$\widehat{\underline{f}}_s(x, x_L, \underline{\mu}_s, \underline{\sigma}_s) = \left[\Phi\left(\frac{\underline{\mu}_s - x_L}{\underline{\sigma}_s} \right) \underline{\sigma}_s \sqrt{2\pi} \right]^{-1} \exp\left[-\frac{1}{2} \left(\frac{x - \underline{\mu}_s}{\underline{\sigma}_s} \right)^2 \right].$$

For each γ in the estimated identified set, we follow a similar procedure described above for the work density, to approximate the confidence intervals of the identified region for $\widehat{\underline{f}}_s(x, x_L, \underline{\mu}_s, \underline{\sigma}_s)$. Specifically, for each $\gamma \in \widehat{\Gamma}_\delta$ defined in (3) we first compute an estimate of $g_s(x_k, \gamma)$, denoted by $\widehat{g}_s(x_k, \gamma)$. Then we minimize the squared distance between $\widehat{\underline{f}}_s(x_k, x_L, \underline{\mu}_s, \underline{\sigma}_s)$ and the product of $f_s^{(N)}(x_k)\widehat{g}_s(x_k, \gamma)$ summed over $k \in \{1, \ldots, 200\}$ with respect to $\left(\underline{\mu}_s, \underline{\sigma}_s\right)$, to obtain

$$(\widehat{\underline{\mu}}_s(\gamma), \widehat{\underline{\sigma}}_s(\gamma)) = \underset{\left(\underline{\mu}_s,\underline{\sigma}_s\right)}{\arg\min} \sum_{k=1}^{200} \left[f_s^{(N)}(x_k|Z)\widehat{g}_s(x_k, \gamma) - \widehat{\underline{f}}_s(x_k, \widehat{x}_L, \underline{\mu}_s, \underline{\sigma}_s)) \right]^2 .$$

for all $\gamma \in \widehat{\Gamma}_\delta$, and form the union of the minimizers. Table S-9 displays the confidence interval of the MSE for these approximations. It shows the truncated normal distribution closely approximates the distribution of excess return under shirking.

D Solving the pure moral hazard model

To derive optimal compensation as a function of x in the analogous two-state pure moral hazard model for a given set of primitives (that is fixing values for $f_s(x)$, $g_s(x)$, α_1, α_2 and γ), 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

$$y_s(x) = \gamma^{-1}\frac{b'}{b-1}\ln\alpha_2 + \gamma^{-1}b'\ln[1+\eta_s^p\left(\frac{\alpha_2}{\alpha_1}\right)^{\frac{1}{b-1}} - \eta_s^p g_s(x)],$$

where $y_s(x)$ is the optimal compensation schedule and η_s^p solves

$$\int_{\overline{x}}^{\infty}\frac{g_s(x)-(\alpha_2\,/\alpha_1\,)^{\frac{1}{b-1}}}{1+\eta_s^p\,(\alpha_2\,/\alpha_1\,)^{\frac{1}{b-1}}-\eta_s^p g_s(x)}f_s(x)dx = 0. \tag{4}$$

We numerically compute the integral (4) as follows. For each candidate solution value of η_s^p, we conduct a grid search to detect the points in the range of x where the left side of (4) is either infinite or zero. These points divide the domain of x into a number of intervals. The integral (4) is then numerically computed for the candidate η_s^p on each interval and then summed over the integrated values. Searching over different positive values of η_{st}^p to minimize the squared value of the left side of (4) yields the unique root, whose existence is guaranteed by the Kuhn-Tucker theorem.

E Extended Sample

Panel C and Panel D in Table S-3 presents the distribution of accounting returns in the extended sample. Table S-4 presents the time-series summary of the firm characteristics from 1993 to 2005. Table S-5 presents the firm characteristics and compensation before and after SOX was enacted.

REFERENCES

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

ANTLE, R., AND A. SMITH. 'An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives.' Journal of Accounting Research 24 (1986): 1-39.

AÏT-SAHALIA, Y.; P. J. BICKEL; AND T. M. STOKER. 'Goodness-of-fit Tests for Kernel Regression with an Application to Option Implied Volatilities.' Journal of Econometrics 105 (2001): 363–412.

GAYLE, G.-L., AND R. A. MILLER. ‘Identifying and Testing Models of Managerial Compensationn.’ Review of Economic Studies 82 (2015): 1074–1118.

HALL, B. J., AND J. B. LIEBMAN. ‘Are CEOs Really Paid Like Bureaucrats?’ The Quarterly Journal of Economics 103 (1998): 653–691.

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

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

Table S-1: Nonparametric Tests (Balanced Sample)

A: Test on PDF of Excess Returns

Sector	Primary		Consumer		Service	
(Size, D/E)	Bad	Good	Bad	Good	Bad	Good
(S,S)	7.36	1.31	30.57	18.74	10.84	12.82
(S,L)	4.67	1.98	0.97	10.02	9.3	5.91
(L,S)	4.30	4.67	3.54	3.38	10.52	6.6
(L,L)	29.3	9.44	11.03	23.28	47.31	31.01

B: Test on Contract Shape

Sector	Primary		Consumer		Service	
(Size, D/E)	Bad	Good	Bad	Good	Bad	Good
(S,S)	6.82	2.57	3.86	1.67	4.46	3.02
(S,L)	12.66	5.18	2.38	2.32	3.36	8.09
(L,S)	12.61	5.03	2.09	2.73	4.95	3.84
(L,L)	9.82	3.56	4.21	3.02	9.50	2.00

Note: The critical value for these one-sided tests at the 5% confidence level is 1.64.

Table S-2: Bond Prices

A. Raw values			B. Grid selection			
Fiscal year	b_t	b_{t+1}	Fiscal year	b_{t+1}	Grid b'	Grid b
1993	16.86	18.38	2000	15.40	15.55	16.39
1994	18.38	15.91	1997	15.69		
1995	15.91	16.21	1994	15.91	16.1	16.47
1996	16.21	16.00	1996	16.00		
1997	16.00	15.69	1999	16.19		
1998	15.69	17.19	2001	16.21		
1999	17.19	16.19	1995	16.21		
2000	16.19	15.40	2002	16.57		
2001	15.40	16.21	1998	17.19	17.8	16.51
2002	16.21	16.57	1993	18.38		
2003	16.57	17.10	2005	16.33	16.4	16.50
2004	17.10	16.45	2004	16.45	16.8	16.53
2005	16.45	16.33	2003	17.10		

Note: Panel A lists the raw values of the bond prices (b_t, b_{t+1}) that are calculated using the method explained in the section A.f. Panel B ranks the bond price b_{t+1} ascendingly within each era, respectively, and reports the grid points used in the structural estimation.

Table S-3: Estimates of the Probability Distribution of Accounting Returns

A. Pre-SOX (Main Sample)						
Firm Type	Primary		Consumer goods		Service	
(A, C)	φ_1/φ_2	Obs	φ_1/φ_2	Obs	φ_1/φ_2	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)	φ_1/φ_2	Obs	φ_1/φ_2	Obs	φ_1/φ_2	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)	φ_1/φ_2	Obs	φ_1/φ_2	Obs	φ_1/φ_2	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)	φ_1/φ_2	Obs	φ_1/φ_2	Obs	φ_1/φ_2	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 its value is greater (less) than the industry average and its probability is $\varphi_2(\varphi_1)$.

Table S-4: 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 S-5: 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 returns		1.100	1.126	4.0	1.115	1.085	-3.6	1.164	1.092	-11.1
		(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 S-6: 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 S-7: The 95% Confidence Regions of Risk-aversion and Corresponding Certainty Equivalent (in 2006 US$)

A: Main Sample			
Eras	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			
Eras	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 S-8: 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 S-9: 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 S-10: Gross Losses to the Shareholders Firms the CEO from Shirking (in %, Extended Sample)

$\rho_1 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre} \{x [1 - g_{s,pre}(x)]\}$

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: 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 S-11: Compensating Differential from CEO Shirking versus Working (in $ thousands, Extended Sample)

$\rho_2 \equiv b_{t+1}\left[(b_t - 1)\,\gamma\right]^{-1} \ln(\alpha_{2,pre}/\alpha_{1,pre})$

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: 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 S-12: Administrative Cost (in $ thousands, Extended Sample)

$$\tau_1 \equiv \gamma^{-1} \frac{b_{t+1}}{b_t - 1} \ln \alpha_{2,pre}$$

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: 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.

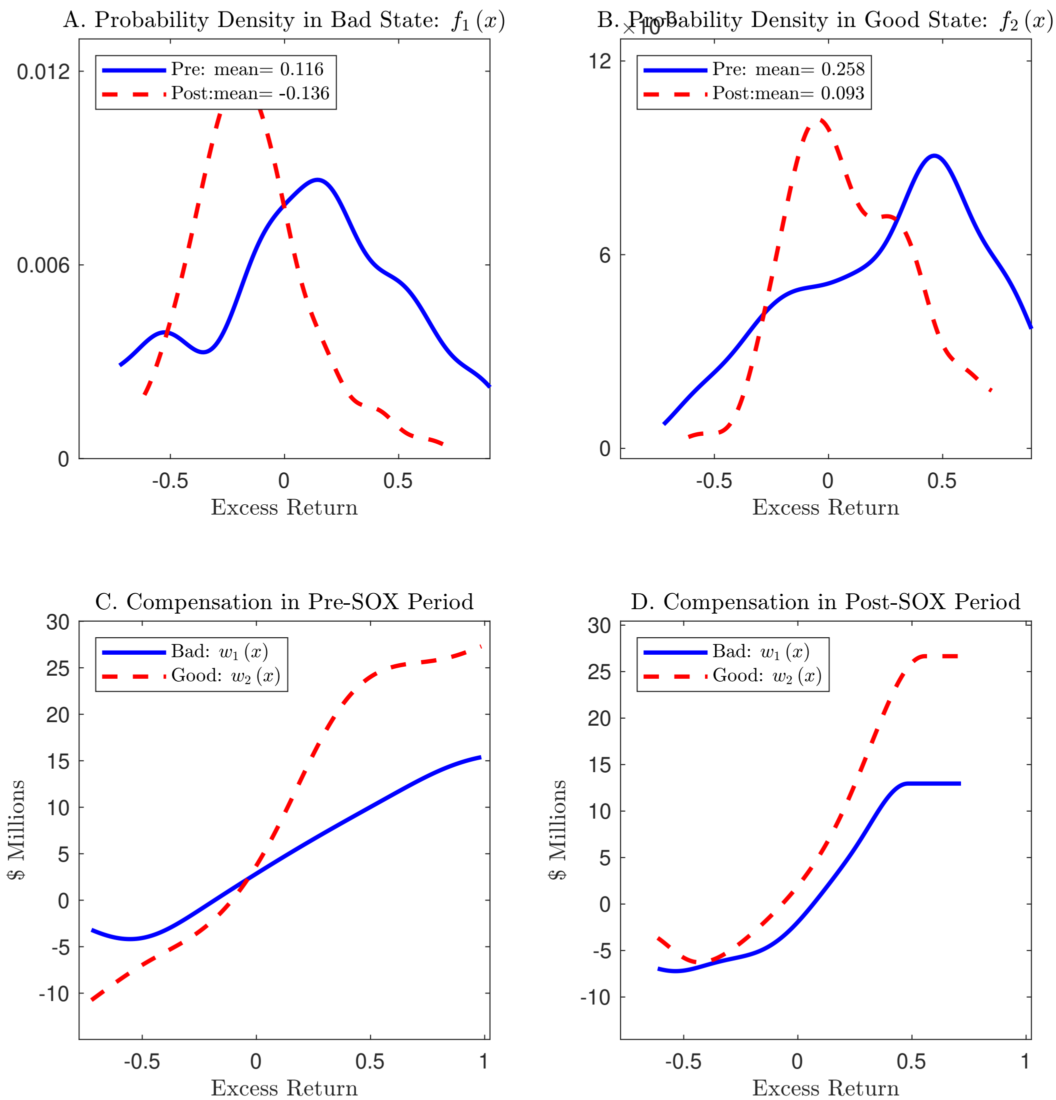
Table S-13: Aggregate Agency Costs (in $ thousands, Extended Sample)

$$\tau_2 \equiv \sum_{s=1}^{2} \varphi_{s,pre} E_{s,pre}\left[w_{s,pre}(x)\right] - \tau_1$$

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: 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 S-1: Empirical Compensation Schedule and Excess Return Density (Balanced Sample)



Note: The plots present the non-parametrically estimated density of excess returns and the optimal compensation of firms with small size and low leverage in the Consumer Goods sector. "Pre" and "Post" indicating the pre-SOX (2000-2001) and post-SOX (2004-2005) eras. The compensation of both eras is anchored at bond prices equal to 16.5 (b_t) and 16.4 (b_{t+1}).