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

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

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

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

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 $\Gamma_{\mathrm{H}t}$, 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

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

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 $\Gamma_{\mathrm{H}\alpha}^{(N)}(T)$ as

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

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

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

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

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

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 $\gamma \in \Gamma_{\text{H}}^{(N)}(T)$ 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 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 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.