Internet Appendix for
"Why Are CEOs Rarely Fired?
Evidence from Structural Estimation"

October 24, 2009

Contents:

Appendix A: The Board's Learning Problem

This Appendix solves the board's learning problem, which is a Kalman filtering problem. I use the notation $\kappa_\epsilon \equiv \sigma_\epsilon^2 / \left(\phi^2 \sigma_0^2\right)$, and $\kappa_z \equiv \sigma_z^2/\sigma_0^2$. The surprises in the additional signal and persistence-adjusted profitability equal

$$\begin{aligned}
\delta_{z,t} &\equiv z_t - \mu_t & \text{(A1)}\\
\delta_{y,t} &\equiv \frac{1}{\phi}\left(y_t - y_{t-1}\right) + y_{t-1} - \mu_t = \alpha + \frac{1}{\phi}\epsilon_t - \mu_t. & \text{(A2)}
\end{aligned}$$

Standard results on Bayesian learning (e.g. Zellner 1971) imply that $\sigma^2(\tau)$, the board's variance of ability α after τ periods of learning has occurred, decays monotonically and deterministically with tenure according to

$$\sigma^2(\tau) = \sigma_0^2 \left[1 + \tau\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right]^{-1}.$$

Then we have

$$
\begin{aligned}
\mu_{t+1} &= \mu_t + \delta_{y,t}\theta_y\left(\tau_t\right) + \delta_{z,t}\theta_z\left(\tau_t\right) && \text{(A3)} \\
\theta_y\left(\tau\right) &\equiv \frac{\sigma^2\left(\tau\right)\phi^2}{\sigma_\epsilon^2}\left(1 + \sigma^2\left(\tau\right)\phi^2/\sigma_\epsilon^2 + \sigma^2\left(\tau\right)/\sigma_z^2\right)^{-1} && \text{(A4)} \\
&= \kappa_\epsilon^{-1}\left(1 + \left(\tau+1\right)\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right)^{-1} && \text{(A5)} \\
\theta_z\left(\tau\right) &= \kappa_z^{-1}\left(1 + \left(\tau+1\right)\left(\kappa_\epsilon^{-1} + \kappa_z^{-1}\right)\right).^{-1} && \text{(A6)}
\end{aligned}
$$

The posterior mean follows a random walk with no drift. The board rationally ignores the industry component of profitability, v_t, which contains no information about the CEO's skill. Also, the board adjusts for persistence in profitability (Equation (A2)).

Next I compare the influence of the profitability signal and additional z signal on the board's beliefs about CEO skill. Specifically, I compare the change in posterior beliefs resulting from a one standard deviation z shock and a one standard deviation profitability signal shock. The model predicts that the response to the z shock is $P \equiv \sigma_\epsilon/(\phi\sigma_z)$ times larger than the response to the profitability signal shock. This result follows from equations (A3)-(A6). A one standard deviation z shock corresponds to $\delta_z = \sigma_z$, which moves beliefs by $\theta_z(\tau)\sigma_z$. A one standard deviation X shock corresponds to $\delta_X = \sigma_\epsilon/\phi$, which moves beliefs by $\theta_X(\tau)\sigma_\epsilon/\phi$. Taking ratios,

$$
\frac{\theta_z(\tau)\sigma_z}{\theta_X(\tau)\sigma_\epsilon/\phi} = \frac{\kappa_z^{-1}\sigma_z}{\kappa_X^{-1}\sigma_\epsilon/\phi} = \frac{\sigma_\epsilon}{\sigma_z\phi} \equiv P.
$$

Appendix B: Bellman Equation for the Board's Optimization Problem

This Appendix provides the Bellman equation for the board's optimization problem. I introduce notation to distinguish between μ_t^{inc}, the posterior mean of the incumbent CEO's skill α going into period t, and μ_t, the prior mean of the CEO chosen to serve in period t. If the firm decides not to fire the incumbent, then $\mu_t = \mu_t^{inc}$, otherwise $\mu_t = \mu_0$.

Proposition 1 *(Bellman equation): The board's objective function can be simplified as*

$$
\begin{aligned}
\frac{U_t}{\kappa B_t} &= E_t\left[\sum_{s=0}^{\infty}\beta^s v_{t+s}\right] + \left(\frac{1-\phi}{1-\beta\left(1-\phi\right)}\right) y_{t-1} + && \text{(A7)} \\
& \left(\frac{\phi}{1-\beta\left(1-\phi\right)}\right)\left(\frac{1}{1-\beta}\right)\mu_0 + V\left(\eta_t^{inc}, \tau_t, b_t\right) && \text{(A8)}
\end{aligned}
$$

where $\eta_t^{inc} = \mu_t^{inc} - \mu_0$, *and the value function* $V(\eta, \tau, 0)$ *solves the Bellman equation*

$$
\begin{aligned}
V(\eta, \tau, 0) &= \max\{V_{fire}, V_{keep}(\eta, \tau)\}, && \text{(A9)}\\
V_{fire} &= V(0,0,0) - c && \text{(A10)}\\
c &\equiv c^{(firm)} + c^{(pers)}/\kappa && \text{(A11)}\\
V_{keep}(\eta, \tau) &= \left(\frac{\phi}{1-\beta(1-\phi)}\right)\eta + \beta f(\tau) V(\eta, \tau, 1) + && \text{(A12)}\\
&\quad \beta(1 - f(\tau)) E\left[V(\eta + \theta_X(\tau)\delta_X + \theta_z(\tau)\delta_z, \tau + 1, 0)\right] && \text{(A13)}\\
\begin{pmatrix} \delta_y \\ \delta_z \end{pmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2/\phi^2 + \sigma^2(\tau) & 0 \\ 0 & \sigma_z^2 + \sigma^2(\tau) \end{bmatrix}\right),
\end{aligned}
$$

subject to a boundary condition if the CEO has just retired:

$$V(\eta, \tau, 1) = V(0,0,0) - c. \tag{A14}$$

Proof: I distinguish between total turnover costs from forced turnover (c_{fire}) and total turnover costs from voluntary turnover (c_{retire}). In my main model results and estimation, I set $c_{fire} = c_{retire} = c$. In the robustness section, I allow $c_{fire} \neq c_{retire}$, so separating the two here is useful. Substituting equation (A11) into (4), and then substituting the result into (3), the board's optimization problem is

$$\max_{\{d_{t+s}\}_{s=0}^{\infty}} U_t = \max_{\{d_{t+s}\}_{s=0}^{\infty}} \kappa E_t\left[\sum_{s=0}^{\infty} \beta^s B_{t+s}\left(v_{t+s} + y_{t+s} - d_{t+s}c_{fire} - b_{t+s}c_{retire}\right)\right],$$

where d_t and b_t are indicator variables equal to 1 if the CEO is fired or retired, respectively, in period t. Since the firm pays out profits immediately as dividends, the firm's book value is constant over time, so $B_{t+s} = B_t$ and

$$
\begin{aligned}
\max_{\{d_{t+s}\}_{s=0}^{\infty}} \frac{U_t}{\kappa B_t} &= \max_{\{d_{t+s}\}_{s=0}^{\infty}} E_t\left[\sum_{s=0}^{\infty} \beta^s\left(v_{t+s} + y_{t+s} - d_{t+s}c_{fire} - b_{t+s}c_{retire}\right)\right]\\
&= E_t\left[\sum_{s=0}^{\infty} \beta^s v_{t+s}\right] + VF_t,
\end{aligned}
$$

$$VF_t = \max_{\{d_{t+s}\}_{s=0}^{\infty}} E_t\left[\sum_{s=0}^{\infty} \beta^s\left(y_{t+s} - d_{t+s}c_{fire} - b_{t+s}c_{retire}\right)\right].$$

Next I write y_{t+s} as a function of y_{t-1}, shocks, and future posterior means:

$$
\begin{aligned}
y_t &= y_{t-1}(1-\phi) + \phi\mu_t + \phi\delta_{y,t}\\
y_{t+1} &= \left[y_{t-1}(1-\phi) + \phi\mu_t + \phi\delta_{y,t}\right](1-\phi) + \phi\mu_{t+1} + \phi\delta_{y,t+1}\\
&\vdots\\
y_{t+s} &= y_{t-1}(1-\phi)^{s+1} + \phi\sum_{\tau=0}^{s}\mu_{t+\tau}(1-\phi)^{s-\tau} + \phi\sum_{\tau=0}^{s}\delta_{y,t+\tau}(1-\phi)^{s-\tau}\\
E_t[y_{t+s}] &= y_{t-1}(1-\phi)^{s+1} + E_t\left[\phi\sum_{\tau=0}^{s}\mu_{t+\tau}(1-\phi)^{s-\tau}\right],
\end{aligned}
$$

since $E_t[\delta_{y,t+\tau}] = E_t[E_{t+\tau}[\delta_{y,t+\tau}]]$ and $E_{t+\tau}[\delta_{y,t+\tau}] = 0$. Next, we have

$$\begin{aligned}
E_t\left[\sum_{s=0}^{\infty}\beta^s y_{t+s}\right] &= \sum_{s=0}^{\infty}\beta^s E_t[y_{t+s}] \\
&= \sum_{s=0}^{\infty}\beta^s\left[y_{t-1}(1-\phi)^{s+1} + E_t\left[\phi\sum_{\tau=0}^{s}\mu_{t+\tau}(1-\phi)^{s-\tau}\right]\right] \\
&= y_{t-1}(1-\phi)\sum_{s=0}^{\infty}\beta^s(1-\phi)^s + \phi\sum_{s=0}^{\infty}\sum_{\tau=0}^{s}\beta^s(1-\phi)^{s-\tau}E_t[\mu_{t+\tau}] \\
&= \left(\frac{1-\phi}{1-\beta(1-\phi)}\right)y_{t-1} + \left(\frac{\phi}{1-\beta(1-\phi)}\right)\sum_{s=0}^{\infty}\beta^s E_t[\mu_{t+s}] \\
&= \left(\frac{1-\phi}{1-\beta(1-\phi)}\right)y_{t-1} + \left(\frac{\phi}{1-\beta(1-\phi)}\right)\sum_{s=0}^{\infty}\beta^s(\mu_0 + E_t[\eta_{t+s}]),
\end{aligned}$$

where I have used the relation:

$$\mu_t = \mu_0 + \eta_t.$$

In sum, we have

$$\begin{aligned}
E_t\left[\sum_{s=0}^{\infty}\beta^s y_{t+s}\right] &= \left(\frac{1-\phi}{1-\beta(1-\phi)}\right)y_{t-1} + \left(\frac{\phi}{1-\beta(1-\phi)}\right)\left(\frac{1}{1-\beta}\right)\mu_0 \\
&\quad + \left(\frac{\phi}{1-\beta(1-\phi)}\right)\sum_{s=0}^{\infty}\beta^s E_t[\eta_{t+s}]
\end{aligned}$$

Plugging this into the expression for VF,

$$VF_t \equiv \left(\frac{1-\phi}{1-\beta(1-\phi)}\right)y_{t-1} + \left(\frac{\phi}{1-\beta(1-\phi)}\right)\left(\frac{1}{1-\beta}\right)\mu_0 + V_t^*,$$

$$V_t^* = \max_{d_t}\left\{\left(\frac{\phi}{1-\beta(1-\phi)}\right)\eta_t - d_t c_{fire} - b_t c_{retire} + \beta E_t\left[V_{t+1}^*\right]\right\},$$

so

$$V\left(\eta_t^{inc}, \tau_t, b_t\right) = \max_{d_t}\left\{\left(\frac{\phi}{1-\beta(1-\phi)}\right)\eta_t - d_t c_{fire} - b_t c_{retire} + \beta E_t\left[V\left(\eta_{t+1}^{inc}, \tau_{t+1}, b_{t+1}\right)\right]\right\}.$$

If the incumbent CEO has just retired, the firm hires a new CEO ($\eta = 0$) and pays the retirement cost:

$$V_{retire} = V\left(\eta_t^{inc}, \tau_t, 1\right) = V(0,0,0) - c_{retire}.$$

Otherwise, if $b_t = 0$ and $d_t = 1$ (the firm fires its CEO), then the firm hires a new CEO and pays the firing cost:

$$V_{fire}\left(\eta_t^{inc}, \tau_t, 0\right) = V(0,0,0) - c_{fire}.$$

If $b_t = 0$ and $d_t = 0$ (the firm keeps its CEO), then

$$\begin{aligned} V_{keep}\left(\eta_t^{inc}, \tau_t, 0\right) &= \left(\frac{\phi}{1-\beta(1-\phi)}\right)\eta_t^{inc} + \beta E_t\left[V\left(\eta_{t+1}^{inc}, \tau_{t+1}, b_{t+1}\right)\right] \\ &= \left(\frac{\phi}{1-\beta(1-\phi)}\right)\eta_t^{inc} + \beta f(\tau_t) V^{retire} + \beta(1-f(\tau_t)) E_t\left[V\left(\eta_{t+1}^{inc}, \tau_{t+1}, 0\right)\right]. \end{aligned}$$

The firm chooses d_t (fire or keep CEO) according to

$$V\left(\eta_t^{inc}, \tau_t, 0\right) = \max\left\{V^{fire}\left(\eta_t^{inc}, \tau_t, 0\right), V^{keep}\left(\eta_t^{inc}, \tau_t, 0\right)\right\}.$$

Recalling from equation (A3) that

$$\begin{aligned} \mu_{t+1}^{inc} &= \mu_t^{inc} + \theta_y(\tau_t)\delta_{y,t} + \theta_z(\tau_t)\delta_{z,t} \\ \mu_0 + \eta_{t+1}^{inc} &= \mu_0 + \eta_t^{inc} + \theta_y(\tau_t)\delta_{y,t} + \theta_z(\tau_t)\delta_{z,t} \\ \eta_{t+1}^{inc} &= \eta_t^{inc} + \theta_y(\tau_t)\delta_{y,t} + \theta_z(\tau_t)\delta_{z,t} \end{aligned}$$

I write the Bellman in its final form by dropping time and incumbent subscripts and substituting in for V^{retire}. End of proof.

Equation (A7) shows that the board's objective function is the sum of an industry-specific component, a component due to persistence in profitability, and a component V which depends on the CEO's posterior mean skill and tenure in office. Each period the board makes a firing decision by comparing its utility from firing the CEO (V_{fire}) and not firing him (V_{keep}) (Equation (A9)). Expression (A10) shows that after firing the CEO, the board hires a new one and incurs the firing cost; the firing utility V_{fire} is constant over time. The board's decision depends on the total κ-adjusted turnover cost, defined in equation (A11), not on the firm and personal costs separately. In equation (A12), the utility V_{keep} from keeping the CEO depends on his expected contribution this period (the μ term) and the expected utility V next period, which in turn depends on whether the CEO quits (with probability $f(\tau)$) at the end of the period. If the CEO does not quit, he enters next period with posterior mean (minus the prior) equal to $\eta' = \eta + \theta_X(\tau)\delta_X + \theta_z(\tau)\delta_z$ (from the learning rule), and one more year of tenure (hence $\tau + 1$). The boundary condition in equation (A14) shows that following a voluntary succession, the board hires a new CEO and pays cost c. The prior mean μ_0 drops out of the Bellman equation, which still depends on η, distance between the posterior mean and the prior mean.

Appendix C: Numerical Solution of Bellman Equation

This Appendix describes how I numerically solve the Bellman equation to find the board's optimal CEO firing rule. I obtain an approximate solution for $V(\mu, \tau, 0)$ by discretizing the state space and iterating on the Bellman equation.

I approximate the value function using the Jacobi Iteration method. I start by discretizing the state space. State variable τ_t takes values in set $\varsigma = \{0, 1, ..., \overline{\tau} - 1\}$, where $\overline{\tau} = \sup \tau$ is the maximum possible number of terms in office. I let μ takes values in finite set M, which contains 1,001 equally spaced points in the interval $[\mu_0 - c_{fire} - 2\sigma_0, \mu_0 + c_{fire} + 2\sigma_0]$; the length of the interval does not need to be extremely large, because the extrapolation used below ends up being quite accurate. To speed up the iteration, I start with a guess of V^0 over the grid $\varsigma \times M$:

$$V^0(\mu, \tau, 0) = \left(\frac{\phi}{1 - \beta(1 - \phi)}\right)\left[\frac{\mu_0}{(1 - \beta)} + \max(\mu - \mu_0, 0)\frac{1 - \beta^{\overline{\tau} - \tau}}{1 - \beta}\right].$$

Then I update the value function according to

$$\begin{aligned} V^{t+1}(\mu, \tau, 0) &= \max\{V^t(\mu_0, 0, 0) - c_{fire}, \left(\frac{\phi}{1 - \beta(1 - \phi)}\right)\mu + \beta f(\tau)\left[V^t(\mu_0, 0, 0) - c_{retire}\right] + \\ &\quad \beta(1 - f(\tau)) E\left[V^t(\mu + \theta_X(\tau)\delta_X + \theta_z(\tau)\delta_z, \tau + 1, 0)\right]\}. \\ &\quad \begin{pmatrix} \delta_X \\ \delta_z \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2/\phi^2 + \sigma^2(\tau) & 0 \\ 0 & \sigma_z^2 + \sigma^2(\tau) \end{bmatrix}\right) \end{aligned}$$

I approximate the expectation above using Gauss-Hermite quadrature, as follows. Recall $V^t(\mu, \tau)$ is defined only for μ in the finite set M. First, I create a function $\widehat{V}^t(\mu, \tau)$ which is defined for all $\mu \in \mathbb{R}$ by performing piecewise cubic spline interpolation and extrapolation of the function $V^t(\mu, \tau)$. Second, I apply two-dimensional Gauss-Hermite quadrature with 7 nodes as follows: For each $\mu \in M$ and $\tau = 0, 1, ..., \overline{\tau} - 1$,

$$\begin{aligned} & E\left[V^t(\mu + \theta_X(\tau)\delta_X + \theta_z(\tau)\delta_z, \tau + 1, 0)\right] \\ \approx\ & \pi^{-1}\sum_{i=1}^{7}\sum_{j=1}^{7}\omega_i\omega_j\widehat{V}^t\left(\mu + \theta_X(\tau)\left[\sqrt{2(\sigma_\epsilon^2/\phi^2 + \sigma^2(\tau))}x_i\right] + \theta_z(\tau)\left[\sqrt{2(\sigma_z^2 + \sigma^2(\tau))}x_j\right], \tau + 1, 0\right) \end{aligned}$$

where $\{x_i\}$ and $\{\omega_i\}$ are the Gauss-Hermite quadrature nodes and weights, respectively. I stop iterating as soon as

$$\max_{(\tau, \mu) \in \varsigma \times M}\left|V^{t+s} - V^t\right| < 10^{-5}.$$

Appendix D: Simulation Method

I define a CEO spell as all the periods a CEO serves in office. To simulate a single spell, I draw the CEO's true skill α from the prior distribution, I generate firm-specific profitability y_t and additional signals z_t using the CEO's true skill α, and I update the board's beliefs according to the learning rule in equation (AA3). Simulated CEOs are fired according to the optimal rule from the Bellman equation, and they leave office voluntarily with probability $f(\tau)$.

Appendix E: Additional Details on SMM Estimation

I use the optimal weighting matrix

$$W = \left[N \mathrm{var}\left(\widehat{M_N}\right)\right]^{-1}.$$

I compute the 14x14 covariance matrix $\widehat{M_N}$ using the seemingly unrelated regressions approach. The moments can be expressed as the coefficients from the following system of regression equations:

$$\begin{aligned}
y_{it}^* &= \lambda_0 + \lambda_1 y_{it-1}^* + \Delta^{(-2)} + \Delta^{(-1)} + \Delta^{(0)} + \Delta^{(1)} + \Delta^{(2)} + \delta_{it} \\
\delta_{it}^2 &= Var\left(\delta\right) + w_{it} \\
d_{it} &= h^{(1-2)} + h^{(2-3)} + h^{(4-6)} + h^{(7+)} + \eta_{it} \\
Var_i\left(X_{it}\right) &= E\left[Var\left(X\right)\right] + e_i \\
\left(E_i\left[X_{it}\right] - E\left[E_i\left[X_{it}\right]\right]\right)^2 &= Var\left(E\left[X\right]\right) + \iota_i
\end{aligned}$$

The coefficients $h^{(j)}$ are fixed effects for tenure (j). Var_i denotes variance within CEO spell i, and E_i denotes average within CEO spell i. I estimate each regression separately using ordinary least squares, which provides consistent estimates for each moment as well as regression disturbances. Each regression above has the form

$$Y_i = X_i \beta_i + \varepsilon_i,$$

where Y_i is $N_i \times 1$ and β_i is $k_i \times 1$. The covariance between moments estimators β_i and β_j is the $k_i \times k_j$ matrix

$$Cov\left(\widehat{\beta}_i, \widehat{\beta}_j\right) = \left(X_i' X_i\right)^{-1} X_i' \Omega_{ij} X_j \left(X_j' X_j\right)^{-1},$$

where $\Omega_{ij} = Cov\left(\varepsilon_i, \varepsilon_j\right)$ is the $N_i \times N_j$ matrix whose element t, s is $Cov\left(\varepsilon_{it}, \varepsilon_{js}\right)$. I estimated the covariance matrix Ω_{ij} for each pair of moments ij, allowing for time series autocorrelation and also correlation across regressions.

I define

$$G_N = M_N - \frac{1}{S}\sum_{s=1}^{S} m_n^s(\theta).$$

Applying the result of Pakes and Pollard (1989) with the efficient weighting matrix, we obtain

$$\begin{aligned}
\sqrt{N}\left(\widehat{\theta} - \theta_0\right) &\rightarrow {}^d\mathcal{N}\left(0, \Omega\right) \\
\Omega &= \left(1 + \frac{1}{S}\right)\left(\Gamma' \Lambda^{-1} \Gamma\right)^{-1},
\end{aligned}$$

where S is the number of simulated data sets (I choose $S = 10$), $\Gamma = \text{plim}_{N\to\infty} \partial\widehat{G}(\theta_0)/\partial\theta'$ and $\Lambda = N\text{avar}(\widehat{M}(\theta_0)) = N\text{avar}(\widehat{m}(\theta_0))$. I estimate Γ by numerically differentiating $\widehat{G}\left(\widehat{\theta}\right)$ with respect to θ, and using $\widehat{\Lambda} = N\widehat{\text{var}}\left(\widehat{M}\right)$, as described above.

We have

$$\sqrt{N}\widehat{G}(\theta_0) \to^d \mathcal{N}\left(0, \left(1 + \frac{1}{S}\right)\Lambda\right),$$

so SMM provides the following test of the model's over-identifying restrictions:

$$\frac{NS}{1+S}\widehat{G}(\theta_0)'\Lambda^{-1}\widehat{G}(\theta_0)' \to^d \chi^2\left(\#\text{moments - }\#\text{parameters}\right).$$