CEMMAP Lectures on Contracts & Market Microstructure

6. Moral Hazard and Managerial Compensation

Robert A. Miller

Carnegie Mellon University

September 2022

- In an Arrow Debreu world with a Walrasian equilibrium, it doesn't matter whether an employee is paid, net of amenity value, the expected value of his marginal product (a certain wage) or the ex-post value (a piece rate).
- Both the employer and the employee can adjust their portfolio of financial assets at the competitive equilibrium rate to achieve the same resource allocation.
- For example if the uncertainty is idiosyncratic, both the employee or the employer could full insure at actuarially fair rates.
- The lectures today analyze compensation and labor supply when contract form matters.
- This arises naturally in environments with asymmetric information.

A Pure Moral Hazard Model

- Gross revenue to a *risk neutral* principal is a random variable x.
- The probability distribution for x depends on choices by a *risk averse* agent.
- The principal proposes a *compensation plan* to the agent, denoted by $w(x)$.
- The agent an *employment* choice:
    - *rejecting* the principal's offer in favor of an outside option ($l_0 = 1$)
    - *accepting* the principal's offer ($l_0 = 0$).
- If $l_0 = 0$ the agent makes an *effort* choice:
    - *working*, pursuing value maximization ($l = 1$).
    - *shirking*, his optimal action if paid a fixed wage ($l = 0$).
- The principal observes l_0 whether the offer is accepted, but not the agent's work effort l.
- After revenue is realized, the principal pockets $x - w(x)$ as profit.

A Pure Moral Hazard Model

Marginal product of the agent

- Denote by $f(x)$ the pdf for x conditional on the agent working.
- Let $f(x)g(x)$ denote the pdf for x when the agent shirks.
- The expected value of x from work exceeds the expected value from shirking:

$$E[xg(x)] \equiv \int xf(x)g(x)\,dx < \int xf(x)\,dx \equiv E[x]$$

- The inequality indicates the *principal prefers working to shirking*.
- Since $f(x)$ and $f(x)g(x)$ are densities, $g(x)$, the ratio of the two densities, is a likelihood ratio.
- That is the likelihood ratio $g(x)$ is nonnegative for all x and:

$$E[g(x)] \equiv \int g(x)f(x)\,dx = 1$$

- We assume:
  - $g(x)$ is bounded, thus ruling out contracts that are arbitrarily close to the first-best resource allocation, where *agents are severely punished when $g(x)$ is extremely high (Mirrlees, 1975)*.

A Pure Moral Hazard Model

Preferences of the agent

- The agent is an expected utility maximizer with utility:

$$-l_0 - \alpha E\left[e^{-\rho w(x)}\right] l - \beta E\left[e^{-\rho w(x)} g(x)\right](1 - l)$$

  where:

  - utility is normalized to the outside option (when $l_0 = 1$).
  - ρ is the coefficient of absolute risk aversion.
  - α is a utility parameter with consumption equivalent $-\rho^{-1} \ln(\alpha)$ measuring the distaste for working.
  - β is a utility parameter with consumption equivalent $-\rho^{-1} \ln(\beta)$ measuring the distaste for shirking.

- A conflict of interest arises between the principal and the agent because:
  - *the agent prefers shirking*, meaning $\beta < \alpha$.
  - yet the principal prefers working since $E[xg(x)] < E[x]$.

Solving the Pure Moral Hazard Model

Participation and incentive compatibility constraints

- To induce the agent to accept employment and shirk, it suffices to propose a contract giving the agent an expected utility of at least minus one:

$$\beta E\left[e^{-\rho w(x)} g(x)\right] \equiv \beta E\left[v(x) g(x)\right] \leq 1$$

- To elicit work from the agent as well, the principal must offer a contract that gives the agent a higher expected utility than:
  1. the outside option (the *participation constraint*):

$$\alpha E\left[e^{-\rho w(x)}\right] \equiv \alpha E\left[v(x)\right] \leq 1$$

  2. the utility from shirking (the *incentive compatibility constraint*):

$$\alpha E\left[e^{-\rho w(x)}\right] \equiv \alpha E\left[v(x)\right] \leq \beta E\left[v(x) g(x)\right] \equiv \beta E\left[e^{-\rho w(x)} g(x)\right]$$

Solving the Pure Moral Hazard Model

Cost minimization inducing work

- The principal minimizes expected compensation subject to the participation and incentive compatibility constraints.
- Equivalently, the principal maximizes the strictly concave objective function $v(x) \equiv \exp\left[-\rho w\left(x\right)\right]$ with linear constraints:

$$E\left\{\ln\left[v(x)\right]\right\} + \eta_0 E\left[1 - \alpha_2 v(x)\right] + \eta_1 E\left[\beta g\left(x\right) v(x) - \alpha v(x)\right]$$

to obtain:

$$w^o\left(x\right) \equiv \rho^{-1}\ln\alpha + \rho^{-1}\ln\left[1 + \eta\left(\frac{\alpha}{\beta}\right) - \eta g(x)\right]$$

where η is the unique positive solution to the equation:

$$E\left[\frac{g\left(x\right)}{\alpha + \eta[(\alpha/\beta) - g\left(x\right)]}\right] = E\left[\frac{(\alpha/\beta)}{\alpha + \eta[(\alpha/\beta) - g\left(x\right)]}\right]$$

Solving the Pure Moral Hazard Model

Intuition for cost minimizing contract

- Why expose the agent to uncertainty in a shirking contract by tying compensation to revenue?
- Hence a agent paid to shirk is offered a fixed wage that just offsets his nonpecuniary benefits, $\rho^{-1} \ln \beta$.
- The certainty equivalent of the cost minimizing contract that induces diligent work is $\rho^{-1} \ln \alpha_2$, higher than the optimal shirking contract to compensate for the lower nonpecuniary benefits because $\alpha > \beta$.
- Moreover the agent is paid a positive risk premium of $E\left[w^o\left(x\right)\right] - \rho^{-1} \ln \alpha$.
- In this model of pure moral hazard these two factors, that working is less enjoyable than shirking, and more certainty in compensation is preferable, explains why compensating an agent to align his interests with the principal is more expensive than merely paying them enough to accept employment.

Solving the Pure Moral Hazard Model

Profit maximization

- Profit maximization by the principal determines which cost minimizing contract the principal should offer the agent.
- The profits from inducing the agent to work are $x - w^o(x)$, while the profits from employing the agent to shirk are $xg(x) - \rho^{-1}\ln(\beta)$.
- Thus work is preferred by the principal if and only if:

$$\max\{0, \rho E[xg(x)] - \ln(\beta)\} \leq \rho E[x - w^o(x)]$$

while a shirking contract is offered if and only if:

$$\max\{0, \rho E[x - w^o(x)]\} \leq \rho E[xg(x)] - \ln(\beta)$$

- Otherwise no contract is offered.

1. With *perfect monitoring* the principal would pay the agent a fixed wage of $\rho^{-1}\ln\alpha_2$ so the maximal amount prinicpal would pay for a perfect monitor is:

$$\tau_1 \equiv E_t\left[w^o(x) - \rho^{-1}\ln\alpha\right] = \rho^{-1}E\left\{\ln\left[1+\eta\left(\frac{\alpha}{\beta}\right) - \eta g(x)\right]\right\}$$

2. The agent's *nonpecuniary benefits from shirking* is:

$$\tau_2 \equiv \rho^{-1}\ln\left(\alpha/\beta\right)$$

3. The *gross loss* the principal incurs from the agent shirking instead of working is:

$$\tau_3 \equiv E\left[x - xg(x)\right]$$

- The model is defined by:
  - $f(x)$ the probability density function of x from working
  - $g(x)$ the likelihood ratio for shirking versus working
  - α distaste for working relative to outside option
  - β distaste for shirking relative to outside option
  - ρ risk-aversion parameter.
- The panel data set is $\{x_n, w_n\}_{n=1}^{N}$ where $w(x) = E[w_n | x_n]$.
- Thus $f(x)$ and $w(x)$ are identified.
- This leaves only $g(x)$ plus (α, β, ρ) to identify.

- The FOC for the Lagrangian can be expressed as:

$$v(x)^{-1} = \alpha\left[1+\eta\left(\alpha/\beta\right)-\eta g(x)\right] = \overline{v}^{-1} - \alpha\eta g(x)$$

where:

$$\lim_{x\to\infty}\left[g(x)\right] = 0 \Rightarrow \lim_{x\to\infty}\left[v(x)^{-1}\right] = \alpha\left[1+\eta\left(\alpha/\beta\right)\right] \equiv \overline{v}^{-1}$$

- These equalities imply:

$$g(x) = \frac{\overline{v}^{-1} - v(x)^{-1}}{\alpha\eta} = \frac{\overline{v}^{-1} - v(x)^{-1}}{\overline{v}^{-1} - E\left[v(x)^{-1}\right]} \tag{1}$$

- Also since both pariticpation and incentive compatibility constraints bind:

$$\alpha = E\left[v(x)\right]^{-1} \tag{2}$$

$$\beta = E\left[v(x)g\left(x\right)\right]^{-1} \tag{3}$$

Identification

The identified set (Gayle and Miller, 2015)

- Noting $v(x) = e^{-\rho w(x)}$ and $\overline{v} \equiv e^{-\rho \overline{w}}$ equations (1), (2) and (3) imply:

$$\alpha(\rho) = E\left[e^{-\rho w^o(x)})\right]^{-1}$$

$$\beta(\rho) = \frac{1 - E\left[e^{\rho w^o(x) - \rho \overline{w}}\right]}{E\left[e^{-\rho w^o(x)}\right] - e^{-\rho \overline{w}}}$$

$$g(x, \rho) = \frac{e^{\rho \overline{w}} - e^{\rho w^o(x)}}{e^{\rho \overline{w}} - E\left[e^{\rho w^o(x)}\right]}$$

- Finally since paying $w^o(x)$ is more profitable than paying $\rho^{-1} \ln(\beta)$:

$$\begin{aligned} 0 &\leq E[x] - E\left[w^o(x)\right] - E[xg(x)] + \rho^{-1} \ln(\beta) \\ &= \frac{cov\left(x, e^{\rho w^o(x)}\right)}{e^{\rho \overline{w}} - E\left[e^{\rho w^o(x)}\right]} - E\left[w^o(x)\right] + \rho^{-1} \ln\left(\frac{1 - E\left[e^{\rho w^o(x) - \rho \overline{w}}\right]}{E\left[e^{-\rho w^o(x)}\right] - e^{-\rho \overline{w}}}\right. \end{aligned}$$

A Dynamic Extension to the Static Model

.

- Adding *simple dynamics* to this model further restricts the set of *observationally equivalent* parameterizations.
- In a multiperiod model where the agent can borrow and save:
  - *interest rate adjustments* affect the value of (smoothing) an extra dollar
  - shiftng the incentive compatibility and participation constraints (Gayle and Miller, 2009)
- Accordingly suppose that each period t:
  - the agent chooses his consumption c_t.
  - the principal announces a compensation function $w_t(x_{t+1})$.
  - the agent chooses $l_{0t} \in \{0, 1\}$ (pariticipation) and $l_t \in \{0, 1\}$ (effort).
  - Output x_{t+1} occurs and he is paid
- For some discount factor $\delta \in (0, 1)$ his lifetime utility is:

$$-\sum_{t=0}^{\infty} \delta^t \exp\left(-\rho c_t\right)\left[l_{0t} + l_t \alpha + (1 - l_t)\beta\right]$$

where the preference parameters (α, β, ρ) and the production parameters $(f(x), g(x))$ have the same interpretation as above.

A Dynamic Extension to the Static Model

Modifying the participation and incentive compatibility constraints

- Similar to the static model define:

$$v_t(x) \equiv \exp\left(-\rho w_t(x) / b_{t+1}\right)$$

where b_t denote the bond price, and assume b_{t+1} is known at period t

- One can show the participation and incentive-compatibility constraints also follow their static model analogues:

$$\alpha^{-1/(b_t-1)} \geq E\left[v_t(x)\right] \tag{4}$$

$$0 \geq E\left[\left(g(x) - (\alpha/\beta)^{1/(b_t-1)}\right) v_t(x)\right]. \tag{5}$$

- The principal chooses v_t for each x to maximize:

$$\int \ln\left[v_t(x)\right] f(x)\mathrm{d}x$$

subject to (4) and (5).

- Changes in b_t tilt the constraints through the effect on $v_t(x)$.
- Since b_t is exogenous, it is a instrument facilitating identification.

A Dynamic Extension to the Static Model

Short term contracts are optimal (Proposition 5, Margiotta and Miller, 2000)

Lemma

The optimal long-term contract is implemented by replicating optimal short-term contracts, where the agent retires for sure in period t or $t+1$, choosing (l_{t0}, l_{t1}, l_{t2}) to maximize:

$$-l_{0t} - E_t \left\{ \left[l_t \alpha^{1/(b_t - 1)} (1 - l_t) + \beta^{1/(b_t - 1)} g_t(x) \right] \exp\left(-\frac{\rho w_t(x)}{b_{t+1}} \right) \right\}$$

- Comparing the principal's problem in this dynamic setting to its static analogue, the only differences are that:
  - $\rho \,/\, b_{t+1}$ replaces ρ, the risk aversion parameter. *Idiosyncratic wealth shocks are optimally smoothed over the agent's lifetime.*
  - $\alpha^{1/(b_t-1)}$ replaces α and $\alpha^{1/(b_t-1)}$ replaces α. *The consumption equivalent of $\alpha^{1/(b_t-1)}$ is $[(b_t - 1)\rho]^{-1} \ln \alpha$, augmenting (reducing) wealth when $\alpha \lesseqgtr 1$.*

- This result builds on Malcomson and Spinnewyn (1988), and Fudenberg, Holmstrom and Milgrom (1990).
- Intuitively no information about shirking in t will arrive after $t + 1$. Since the agent only faces a lifetime wealth constraint, postponing rewards or penalties beyond one period is pointless.
- Several of the assumptions are somewhat contentious:
  - Do managers take actions that only become evident years later?
    - The median (average) tenure of a CEO is about 5 (7) years.
    - In practice, stocks and options are granted and then later vested.
    - CEOs are occasionally fired, and not vested with all previous grants.
  - Managers manipulate returns:
    - for fraudulent purposes (Bertomeu, Marinovic, Miller and Varas, 2018)
    - to signal the state of the firm (Gayle and Miller, 2015).
  - The assumption of complete markets is often questioned, but:
    - the evidence against them is spotty (Altug and Miller,1990).
    - managers save, not borrow, and are financially savvy.

A Fully Parametric Specification

Truncated Normal distribution and Absolute Risk Aversion (CARA)

- Assume x is distributed truncated normal with lower truncation point ψ (representing bankruptcy or limited liability) with mean μ_w (μ_s) and variance σ^2 for parent normal if agent works (shirks):

$$
\begin{aligned}
f(x) &= \frac{1}{\sigma_w\sqrt{2\pi}}\Phi\left(\frac{\mu_w-\psi}{\sigma}\right)^{-1}\exp\left[\frac{-(x-\mu_w)^2}{2\sigma^2}\right] \\
\ln g(x) &= \ln\Phi\left[(\mu_s-\psi)/\sigma\right]-\ln\Phi\left[(\mu_w-\psi)/\sigma\right] \\
&\quad +\frac{\mu_w^2-\mu_s^2}{2\sigma^2}+\frac{(\mu_s-\mu_w)}{\sigma^2}x
\end{aligned}
$$

- Thus the model is parameterized by $(\psi, \mu_w, \sigma, \mu_s, \gamma, \alpha, \beta)$.
- Suppose there are N observations on $(\widetilde{w}_n, x_n)$ where:

$$\widetilde{w}_n \equiv w_n + \epsilon_n \text{ and } E\left[\epsilon_n \,|x_n\right] = 0.$$

A Fully Parametric Specification

Estimation

- Margiotta and Miller (2000) estimate:
  1. ψ with $\widehat{\psi} \equiv \min\{x_1, \ldots, x_N\}$. (Note $\widehat{\psi}$ converges to ψ at rate faster than $\sqrt{N}$ but is sensitive to measurement error.)
  2. (μ_w, σ) with LIML by forming likelihood for $f(x)$ with $\{x_1, \ldots, x_N\}$ under the assumption that $\widehat{\psi} = \psi$. (No first stage correction is necessary.)
  3. $(\mu_s, \rho, \alpha, \beta)$ with NLS based on

$$\widetilde{w}_n = \rho^{-1} \ln \alpha + \gamma^{-1} \ln \left[1 + \eta \left(\frac{\alpha}{\beta} \right) - \eta g(x) \right] + \epsilon_n$$

     using an inner loop at each iteration to solve for η as a mapping of (α, β, μ_s) given $(\widehat{\psi}, \widehat{\mu}_w, \widehat{\sigma})$.
  4. Correct the standard errors for $(\mu_s, \rho, \alpha, \beta)$ in the third step induced by $(\widehat{\mu}_w, \widehat{\sigma})$ obtained from the second step.

A Fully Parametric Specification

Estimating the importance of moral hazard (Table 8, Margiotta and Miller 2000)

- We used the Masson-Antle-Smith (MAS) data set (37 firms in aerospace, electronics, chemicals from 1944 - 1977).
- The annual cost of moral hazard pales in comparison to losses shareholders would make if managers were paid a fixed wage.

TABLE 8

COST OF MORAL HAZARD (MEAN AGED EXECUTIVE IN 1967 US$)

Measure	Industry	Executive	Cost
Δ_1	Aerospace	CEO	186,689
		Non-CEO	2,370
	Chemicals	CEO	232,966
		Non-CEO	2,680
	Electronics	CEO	173,643
		Non-CEO	2,327
Δ_2		CEO	259,181
		Non-CEO	3,272
Δ_3	Aerospace		263,283,500
	Chemicals		85,355,000
	Electronics		104,222,000

50 Years of Managerial Compensation

Changes in managerial compensation (Table 3, Gayle and Miller, 2009)

- We compare MAS data with data from:
  - S&P 500 COMPUSTAT CRSP (2,610 firms 1995 -2004, 2000 $US)
  - a subset formed from those firms in the three MAS sectors

TABLE 3—CROSS-SECTIONAL INFORMATION ON TOTAL COMPENSATION
(In thousands of US$ (2000); standard deviations in parentheses)

Rank	Sector	Old	New restricted	New all
All	All	528	4,121	2,319
		(1,243)	(19,283)	(12,121)
CEO	All	729	6,109	5,320
		(1,472)	(24,250)	(19,369)
Non-CEO	All	400	2,256	1,562
		(1,026)	(12,729)	(9,303)
All	Aerospace	744	6,407	
		(1,140)	(20,689)	
CEO	Aerospace	950	11,664	
		(1,292)	(19,416)	
Non-CEO	Aerospace	624	1,997	
		(695)	(18,563)	
All	Chemicals	543	2,802	
		(1,348)	(9,555)	
CEO	Chemicals	718	3,673	
		(1,527)	(7,072)	
Non-CEO	Chemicals	401	477	
		(241)	(23,390)	
All	Electronics	370	4,501	
		(1,057)	(22,118)	
CEO	Electronics	457	5,325	
		(1,407)	(24,576)	
Non-CEO	Electronics	108	1,635	
		(61)	(18,810)	

50 Years of Managerial Compensation

Changes in components of managerial compensation (Table 4, Gayle and Miller, 2010)

TABLE 4—CROSS-SECTIONAL INFORMATION ON COMPONENTS OF COMPENSATION
(In thousands of US$ (2000); standard deviations in parentheses)

Variable	Rank	Old	New restricted	New all
Salary and bonus	All	219 (114)	838 (1,066)	667 (905)
	CEO	261 (115)	1,037 (1,365)	1,127 (1,282)
	Non-CEO	179 (97)	640 (576)	552 (738)
Value of options granted	All	79 (338)	2,401 (13,225)	903 (3,753)
	CEO	111 (439)	3,402 (18,172)	1,782 (7,169)
	Non-CEO	51 (198)	1,401 (4,237)	681 (2,106)
Value of restricted stock granted	All	11 (95)	187 (1,633)	152 (936)
	CEO	8 (72)	242 (2,021)	298 (1,464)
	Non-CEO	13 (112)	133 (1,118)	115 (743)
Change in wealth from options held	All	5 (134)	785 (14,636)	281 (8,710)
	CEO	7 (167)	1,667 (17,078)	1,474 (13,567)
	Non-CEO	3 (94)	−76 (11,706)	−18 (6,939)
Change in wealth from stock held	All	−3 (439)	−40 (5,681)	125 (4,350)
	CEO	0.434 (479)	−14 (6,712)	264 (6,791)
	Non-CEO	−7 (398)	−64 (4,496)	90 (3,473)

50 Years of Managerial Compensation

Changes in sample composition of firms (Table 2, Gayle and Miller, 2010)

TABLE 2—CROSS-SECTIONAL INFORMATION ON SECTORS
(All currency in millions of US$ (2000); standard deviations in parentheses)

Variable	Sector	Old	New restricted	New all
Sales	All	1,243 (2,250)	3,028 (6,830)	4,168 (109,000)
	Aerospace	1,886 (3,236)	11,500 (14,900)	
	Chemicals	1,246 (2,018)	2,252 (2,091)	
	Electronics	319 (536)	2,469 (6,223)	
Value of equity	All	589 (1,034)	1,273 (2,863)	1,868 (4,648)
	Aerospace	391 (680)	3,132 (3,826)	
	Chemicals	677 (1,107)	800 (869)	
	Electronics	159 (365)	1,283 (3,096)	
Number of firms	All	37	151	1,517
	Aerospace	5	11	
	Chemicals	25	40	
	Electronics	7	100	
Number of employees	All	27,370 (28,850)	12,208 (26,676)	18,341 (46,960)
	Aerospace	49,920 (34,335)	58,139 (69,452)	
	Chemicals	23,537 (25,268)	8,351 (9,323)	
	Electronics	10,485 (7,664)	9,195 (18,266)	
Total assets	All	525 (924)	3,035 (6,550)	9,926 (40,300)
	Aerospace	726 (130)	10,600 (12,900)	
	Chemicals	548 (851)	2,385 (2,380)	
	Electronics	146 (233)	2,551 (6,311)	
Observations	All	1,797	3,260	82,578
	Aerospace	355	233	
	Chemicals	1,092	935	
	Electronics	252	2,092	

- If managers in COMPUSTAT ran firms the same size as managers in MAS, their certainty equivalent wage would have increased by a factor of 2.3, the increase in national income per capita.
- After adjusting for the general increase in living standards over these years, the model attributes:
  - hardly any of the increased managerial compensation to changes in $\rho^{-1} \ln \alpha$, or the certainty equivalent wage.
  - practically all the increase to changes in τ_1, the risk premium.
- The factors driving the change in τ_1 were:
  - not risk preferences: managers in the MAS (COMPUSTAT) population were willing to \$240,670 (\$248,620) to avoid a gamble of winning or losing \$1 million.
  - not $\Delta f(x)$: the biggest $\Delta \tau_1$ in aerospace where the abnormal returns became less dispersed, which reduces the risk premium.
  - the sharp increase in α / β mainly due to increased firm assets, which provides managers with more opportunities to shirk.