Optimal Contracting when there is Moral Hazard

Robert A. Miller

Tilburg University

November 2023

A Pure Moral Hazard Model

Motivation

- In an Arrow Debreu world with a Walrasian equilibrium, it doesn't matter whether an employee is paid the value of his marginal product less the amenity value with a certain wage or 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.
- This lecture analyzes compensation and labor supply when the contract form matters.
- It arises naturally in environments with asymmetric information.

A Pure Moral Hazard Model

Framework

- A risk neutral principal proposes a compensation plan to a risk averse agent, an explicit contract or an implicit agreement, which depends on the future realization of gross revenue to the principal.
- The agent accepts or rejects the principal's (implicit) offer.
- If he rejects the offer he receives a fixed utility from an outside option.
- If he accepts the offer, the agent chooses between pursuing the principal's objectives of value maximization (working), versus following objectives he would pursue if he was paid a fixed wage (shirking).
- The principal observes whether the offer is accepted, but not the agent's work routine.
- After revenue is realized, the agent receives compensation according to the explicit contract or implicit agreement, and the principal pockets the remainder as profit.

- Denote the workplace employment decision of the agent by an indicator $l_0 \in \{0, 1\}$, where $l_0 = 1$ means the agent rejects the principal's offer.
- Denote the effort level choices by $l_j \in \{0, 1\}$ for $j \in \{1, 2\}$, where diligence work is defined by setting $l_2 = 1$, and shirking is defined by setting $l_1 = 1$.
- Since taking the outside option, working diligently and shirking are mutually exclusive activities, $l_0 + l_1 + l_2 = 1$.

A Pure Moral Hazard Model

Revenue and profits of the principal

- Gross revenue to the principal is denoted by x, a random variable drawn from a probability distribution that is determined by the agent's work routine.
- After x is revealed the both the principal and the agent at the end of the period, the agent receives compensation according to the contract or implicit agreement.
- To reflect its potential dependence on (or measurability with respect to) x, we denote compensation by $w(x)$.
- The principal's profit is revenue less compensation, $x - w(x)$.

A Pure Moral Hazard Model

Marginal product of the agent

- Denote by $f(x)$ the probability density function for revenue conditional on the agent working, and let $f(x)g(x)$ denote the probability density function for revenue when the agent shirks.
- We assume:

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

- The inequality reflects the preference of principal for working over 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 $g(x)$ is nonnegative for all x, bounded, and:

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

A Pure Moral Hazard Model

Preferences of the agent

- We assume the agent is an expected utility maximizer and utility is exponential in compensation, taking the form:

$$-l_0 - l_1 \alpha_1 E\left[e^{-\gamma w(x)} g(x)\right] - l_2 \alpha_2 E\left[e^{-\gamma w(x)}\right]$$

  where without further loss of generality we normalize the utility of the outside option to negative one.
- Thus γ is the coefficient of absolute risk aversion, and α_j is a utility parameter with consumption equivalent $-\gamma^{-1}\log(\alpha_j)$ that measures the distaste from effort level $j \in \{1, 2\}$.
- We assume $\alpha_2 > \alpha_1$ meaning that shirking gives more utility to the agent, than working.
- A conflict of interest arises between the principal and the agent because he prefers shirking, meaning $\alpha_1 < \alpha_2$, yet the principal prefers working since $E[xg(x)] < E[x]$.

Solving the Pure Moral Hazard Model

Participation constraint

- To induce the agent to accept the principal's offer and engage in his preferred activity, shirking, it suffices to propose a contract that gives the agent an expected utility of at least minus one.
- In this case we require $w(x)$ to satisfy the inequality:

$$\alpha_1 E\left[e^{-\gamma w(x)} g(x)\right] \leq 1$$

Solving the Pure Moral Hazard Model

Participation and incentive compatibility constraints

- To elicit work from the agent, the principal must offer a contract that gives the agent a higher expected utility than the outside option, and a higher expected utility than shirking.
- In this case we require:

$$\alpha_2 E\left[e^{-\gamma w(x)}\right] \leq 1$$

and:

$$\alpha_2 E\left[e^{-\gamma w(x)}\right] \leq \alpha_1 E\left[e^{-\gamma w(x)} g\left(x\right)\right]$$

- Defining $v(x) \equiv \exp\left[-\gamma w\left(x\right)\right]$ note that:

$$-E\left[w\left(x\right)\right] = \gamma^{-1} E\left\{\log\left[v(x)\right]\right\}$$

the participation constraint can be expressed as:

$$\alpha_2 E\left[v(x)\right] \leq 1$$

and the incentive compatibility constraint becomes:

$$\alpha_2 E\left[v(x)\right] \leq \alpha_1 E\left[v(x) g\left(x\right)\right]$$

- In the transformed problem we maximize a strictly concave objective function with linear constraints.
- Applying the Kuhn Tucker theorem, choose v for each x to maximize:

$$E\left\{\log\left[v(x)\right]\right\} + \eta_0 E\left[1 - \alpha_2 v(x)\right] + \eta_1 E\left[\alpha_1 g\left(x\right) v(x) - \alpha_2 v(x)\right]$$

Lemma (Margiotta and Miller, 2000)

To minimize the cost of inducing the agent to accept employment and work diligently the board offers the contract:

$$w^o(x) \equiv \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[1 + \eta \left(\frac{\alpha_2}{\alpha_1} \right) - \eta g(x) \right]$$

where η is the unique positive solution to the equation:

$$E\left[\frac{g(x)}{\alpha_2 + \eta[(\alpha_2/\alpha_1) - g(x)]} \right] = E\left[\frac{(\alpha_2/\alpha_1)}{\alpha_2 + \eta[(\alpha_2/\alpha_1) - g(x)]} \right]$$

- Differentiate the Lagrangian with respect $v(x)$ to obtain:

$$v(x)^{-1} = \eta_0 \alpha_2 + \eta_1 \alpha_2 - \eta_1 \alpha_1 g(x)$$

- We can show both constraints are met with equality, establishing the formula for η, and showing $\eta_0 = 1$, to yield:

$$v(x)^{-1} = \alpha_2 + \eta_1 \alpha_2 - \eta_1 \alpha_1 g(x)$$

Solving the Pure Moral Hazard Model

Intuition for cost minimizing contract

- There is no point exposing the manager 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, $\gamma^{-1} \ln \alpha_1$.
- The certainty equivalent of the cost minimizing contract that induces diligent work is $\gamma^{-1} \ln \alpha_2$, higher than the optimal shirking contract to compensate for the lower nonpecuniary benefits because $\alpha_2 > \alpha_1$.
- Moreover the agent is paid a positive risk premium of $E\left[w^o(x)\right] - \gamma^{-1} \ln \alpha_2$.
- 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.

Measuring the Importance of Moral Hazard

Three measures

- Recall the optimal compensation with moral hazard is $w^o(x)$ and to meet the participation constraint, shareholders must pay $\gamma^{-1} \ln \alpha_2$.
- Therefore the maximal amount shareholders would pay to rid the firm of the moral hazard problem is:

$$\Delta_1 \equiv E_t\left[w^o(x) - \gamma^{-1}\ln\alpha_2\right] = \gamma^{-1} E\left\{\ln\left[1 + \eta\left(\frac{\alpha_2}{\alpha_1}\right) - \eta g(x)\right]\right\}$$

- A second measure of moral hazard is the nonpecuniary benefits the manager obtains from shirking.
- This is the monetized utility loss from working versus shirking:

$$\Delta_2 \equiv \gamma^{-1}\ln\alpha_1 - \gamma^{-1}\ln\alpha_2 = -\gamma^{-1}\ln\left(\alpha_2/\alpha_1\right)$$

- Third is the gross loss a firm incurs from the manager shirking instead of working:

$$\Delta_3 \equiv E\left[x - xg(x)\right]$$

Identification

Model primitives and the data generating process

- 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.

Identification

What if the risk parameter is known?

- 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:

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

- Also since both pariticpation and incentive compatibility constraints bind:

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

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

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 $\rho^{-1}\ln(\beta)$:

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

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:

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

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:
  - $\rho \big/ b_{t+1}$ 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 $\alpha \lesseqgtr 1$.*

A Dynamic Extension to the Static Model

Discussion

- 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 $(\psi, \mu_w, \sigma, \mu_s, \gamma, \alpha_1, \alpha_2)$.
- Suppose there are N observations on $(\widetilde{w}_n, x_n)$ where:

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

- 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. $(\mu_s, \gamma, \alpha_1, \alpha_2)$ with NLS based on

     $$\widetilde{w}_n = \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[1 + \eta \left(\frac{\alpha_2}{\alpha_1}\right) - \eta g(x)\right] + \epsilon_n$$

     using an inner loop at each iteration to solve for η as a mapping of $(\alpha_2, \alpha_2, \mu_s)$ given $(\widehat{\psi}, \widehat{\mu}_w, \widehat{\sigma})$.
  4. Correct the standard errors for $(\mu_s, \gamma, \alpha_1, \alpha_2)$ 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.

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.

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

50 Years of Managerial Compensation

Changes in components of managerial compensation (Table 4, Gayle and Miller, 2009)

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

50 Years of Managerial Compensation

Changes in sample composition of firms (Table 2, Gayle and Miller, 2009)

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

50 Years of Managerial Compensation

What were the driving forces behind these changes?

- If managers in the COMPUSTAT population ran firms the same size as managers in MAS, their compensation 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 $\gamma^{-1} \ln \alpha_2 / \alpha_0$, or the certainty equivalent wage
  - practically all the increase to changes the risk premium Δ_1
- The factors driving the change in Δ_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 changes in $f(x)$: the biggest change in Δ_1 in aerospace where the abnormal returns became less dispersed *(reducing the premium)*.
  - the sharp increase in α_2 / α_1 mainly due to increased firm assets *(increasing the utility from shirking)*.