Pure Moral Hazard

Robert A. Miller

Structural Econometrics

November 2017

- Auctions and limit order markets are indicative of how market microstructure works.
- Procurement auctions that exploit signals of the contracting bidders can explain why prices also depend on screening signal when there are only a small number of bidders and the auctioneer or procurement officer cannot extract all the surplus through competition.
- But signals about are often used in competitive situations as well:
  - insurance premium ratings (when actuarial odds are private information)
  - sales commissions and managerial compensation (where there is hidden effort).

A Pure Moral Hazard Model

Managerial compensation in aerospace, chemicals and electronics 1948 - 77

		Aerospace	Chemicals	Electronics	All
After-tax compensation	All	126,822	86,094	76,958	93,395
		(410,590)	(493,072)	(428,977)	(468,892)
	CEO	144,731	120,618	96,688	121,840
		(475,720)	(602,724)	(610,090)	(579,963)
	Non-CEO	118,211	69,283	64,522	79,082
		(375,687)	(428,969)	(257,306)	(400,985)
Pretax salary and bonus	All	136,408	121,786	82,223	119,594
		(61,319)	(63,111)	(34,787)	(67,757)
	CEO	175,965	154,324	106,522	151,388
		(58,025)	(70,338)	(38,316)	(67,263)
	Non-CEO	117,386	105,943	66,907	13,596
		(53,305)	(52,440)	(21,073)	(51,904)
After-tax value of options granted	All	16,821	11,759	16,947	13,505
		(58,726)	(46,206)	(65,051)	(51,874)
	CEO	19,463	14,525	19,721	16,322
		(61,331)	(55,171)	(77,674)	(60,354)
	Non-CEO	15,551	10,412	15,200	12,088
		(57,474)	(41,100)	(55,780)	(46,987)
Return on stock held	All	8,797	−5,601	(2,872)	−2,229
		(294,955)	(477,451)	(384,681)	(411,763)
	CEO	6,790	5,835	(5,681)	4,264
		(306,196)	(531,743)	(564,701)	(499,470)
	Non-CEO	9,763	−11,169	(1,102)	−5,497
		(289,817)	(400,172)	203,096	(359,753)
Pretax value of stock bonus	All	0	273	518	249
		0	(1,852)	(4,821)	(2,341)
	CEO	0	239	0	154
		0	(1,496)	0	(1,205)
	Non-CEO	0	290	845	296
		0	(2,003)	(6,139)	(2,705)
Return on options held	All	17,530	1,444	3,507	5,080
		(163,653)	(114,107)	(92,296)	(123,808)
	CEO	18,002	5,523	1,386	7,419
		(218,544)	(156,450)	(110,359)	(165,162)
	Non-CEO	17,303	−548	4,845	3,903
		(129,605)	(86,293)	79,080	(96,559)
Age	All	55.5	57.7	56.7	57.1
		(7.1)	(8.1)	(8.4)	(7.9)
	CEO	57.6	58.0	57.5	57.9
		(6.9)	(7.7)	(7.0)	(7.5)
	Non-CEO	54.5	57.6	56.2	56.8
		(6.9)	(8.5)	(9.1)	(8.4)

A Pure Moral Hazard Model (Margiotta and Miller, 2000)

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.

A Pure Moral Hazard Model

Choices of the agent

- 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 and:

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

- We assume there is an upper range of revenue that might be achieved with diligence, but is extremely unlikely to occur if the agent shirks.
- Formally:

$$\lim_{x \to \infty} [g(x)] = 0$$

- Intuitively this assumption states that a truly extraordinary performance can only be attained if the agent works.
- We assume that $g(x)$ is bounded, an assumption that rules out the possibility of setting a contract that is arbitrarily close to the first best resource allocation, first noted by Mirrlees (1975), by severely punishing the agent when $g(x)$ takes an extremely high value.

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:

$$-I_0 - I_1 \alpha_1 E\left[e^{-\gamma w(x)} g(x)\right] - I_2 \alpha_2 E\left[e^{-\gamma w(x)}\right]$$

  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:

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

Solving the Pure Moral Hazard Model

Cost minimization inducing work

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

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

  the participation constraint can be expressed as:

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

  and the incentive compatibility constraint becomes:

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

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

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

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.

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 $xg(x) - \gamma^{-1}\log(\alpha_1)$.
- Thus work is preferred by the principal if and only if:

$$\max\{0, \gamma E[xg(x)] - \log(\alpha_1)\} \leq \gamma E[x - w^{o}(x)]$$

while a shirking contract is offered if and only if:

$$\max\{0, \gamma E[x - w^{o}(x)]\} \leq \gamma E[xg(x)] - \log(\alpha_1)$$

- Otherwise no contract is offered.

Identification in the Pure Moral Hazard Model

Parameters

- The parameters of the model are characterized by $f(x)$ and $g(x)$, which together define the probability density functions of gross profits, (α_1, α_2), the preference parameters for shirking and diligent work (relative to the normalized utility from taking the outside option), as well as the risk aversion parameter γ.
- For the purposes of this introductory example, we assume the data comprise independent draws of profits and compensation, (x_n, w_n) for a sample of N observations generated in equilibrium.
- When the principal induces shirking, the density $f(x)g(x)$ can be estimated from observations on profits, the wage is constant at $w_n \equiv \gamma^{-1} \log(\alpha_1)$ for all n, but nothing more can be gleaned from the data about the structure of the model.

Identification in the Pure Moral Hazard Model

An implication of the regularity condition

- Our analysis focuses on cases when diligence is induced, and compensation w_n, depends on revenue x_n.
- Hence $f(x)$ is identified, along with N points on the compensation schedule $w_n \equiv w^o(x_n)$.
- Under the assumptions of the model $f(x)$ can be estimated with a nonparametric density estimator.
- From the compensation equation, the regularity condition on $g(x)$ and the fact that $g(x)$ is nonnegative, the maximum compensation the agent can receive is:

$$\lim_{x \to \infty} w^o(x) = \gamma^{-1} \ln \alpha_2 + \gamma^{-1} \ln \left[1 + \eta \left(\frac{\alpha_2}{\alpha_1} \right) \right] \equiv \overline{w} \qquad (1)$$

- Thus $\overline{w}$ is identified, and consistently estimated by the maximum compensation observed in the data.
- This essentially leaves γ, α_1, α_2, and $g(x)$ to identify from $f(x)$, $w^o(x)$, and $\overline{w}$.

- Our analysis proceeds in three steps.
- First we show that if γ is known, then α_1, α_2, and $g(x)$ are identified from the cost minimization problem.
- This means that the set of observationally equivalent parameters can be indexed by the positive real number γ, the risk aversion parameter.
- Second, we show that the firm's preference for working over shirking provides an additional inequality that helps delineate the values of observationally equivalent γ.
- Third, we prove that the set of restrictions we have derived in the first two steps fully characterize the identified set.

Identification in the Pure Moral Hazard Model

Defining a likelihood ratio

- Suppose γ is known, and define the mappings $g(x,\gamma)$ as:

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

- Taking the expectation with respect to $f(x)$ proves $E[g(x,\gamma)] = 1$ for all γ.
- Also $\overline{w} \geq w^{o}(x)$, so $e^{\gamma \overline{w}} \geq E\left[e^{\gamma w^{o}(x)}\right]$ and $e^{\gamma \overline{w}} \geq e^{\gamma w^{o}(x)}$ for all $\gamma > 0$.
- Therefore $g(x,\gamma) \geq 0$ for all $\gamma > 0$.
- Furthermore $w(x) \to \overline{w}$ as $x \to \infty$, and hence $g(x,\gamma) \to 0$, as the regularity condition stipulates.
- This proves $g(x,\gamma)$ can be interpreted as a likelihood ratio satisfying the regularity condition for all $\gamma > 0$.

Identification in the Pure Moral Hazard Model

Defining some taste parameters

- Next define $\alpha_1(\gamma)$ and $\alpha_2(\gamma)$ as:

$$\alpha_1(\gamma) \equiv \frac{1 - E\left[e^{\gamma w^o(x) - \gamma \overline{w}}\right]}{E\left[e^{-\gamma w^o(x)}\right] - e^{-\gamma \overline{w}}}, \quad \alpha_2(\gamma) \equiv \left\{E\left[e^{-\gamma w^o(x)}\right]\right\}^{-1}$$

- Clearly $\alpha_2(\gamma) > 0$ because $e^{-\gamma w^o(x)} > 0$. Similarly the numerator and denominator of the equation for $\alpha_1(\gamma)$ have the same sign for all γ, so $\alpha_1(\gamma)$ is also positive.
- Rearranging the expression for the ratio of the two taste parameters:

$$\frac{\alpha_1(\gamma)}{\alpha_2(\gamma)} = \frac{e^{\gamma \overline{w}} - E\left[e^{\gamma w^o(x)}\right]}{e^{\gamma \overline{w}} - \left\{E\left[e^{-\gamma w^o(x)}\right]\right\}^{-1}}$$

- Since the inverse function is convex, Jensen's inequality implies $E\left[e^{-\gamma w^o(x)}\right] > E\left[e^{\gamma w^o(x)}\right]^{-1}$ or $\left\{E\left[e^{-\gamma w^o(x)}\right]\right\}^{-1} < E\left[e^{\gamma w^o(x)}\right]$.
- Therefore $\alpha_1(\gamma) < \alpha_2(\gamma)$ for all $\gamma > 0$.

Identification in the Pure Moral Hazard Model

Using the cost minimization problem in identification

- Summarizing, given a density $f(x)$ for x and a compensation schedule $w^o(x)$ satisfying $w^o(x) \rightarrow \overline{w}$ as $x \rightarrow \infty$, identified from observations (x_n, w_n), for any positive γ we can construct, as primitives for the principal agent model, a $g(x, \gamma)$, a $\alpha_1(\gamma)$, and a $\alpha_2(\gamma)$.
- But we can also prove a stronger result:

Theorem (Gayle and Miller, 2015)

Suppose the data on x_n and w_n is generated by a parameterization of the model denoted by α_1^*, α_2^*, γ^*, $g^*(x)$ and $f^*(x)$ in which shareholder induce diligent work by solving the cost minimization problem. Then:

$$\begin{aligned} \alpha_1^* &= \alpha_1(\gamma^*) \\ \alpha_2^* &= \alpha_2(\gamma^*) \\ g^*(x) &= g(x, \gamma^*) \end{aligned}$$

- Making $g(x)$ the subject of the compensation equation and differentiating with respect to x yields:

$$g'(x) = -\eta^{-1} e^{-\gamma w(x)} \partial w(x) / \partial x$$

- From this equation it is evident that the slope is defined up to one normalization; a second normalization determines the level of $g(x)$.
- In our setup the regularity condition provides one normalization; the fact that $E[g(x)] = 1$ provides another.
- The formula for $\alpha_2(\gamma^*)$ is due to the participation constraint being met with equality.
- Since the incentive compatibility constraint is also met with equality:

$$\alpha_1 E\left[e^{-\gamma w^o(x)} g(x)\right] = \alpha_2 E\left[e^{-\gamma w^o(x)}\right] = 1$$

and substituting in the formula for $g(x)$ and rearranging to make α_1 the subject of the equation produces the formula evaluated at γ^*.

Identification in the Pure Moral Hazard Model

Restrictions from profit maximization

- The restrictions from cost minimization place no restrictions on γ.
- Imposing profit maximization limits the set of admissible γ.
- If paying $w^o(x)$ is more profitable than paying $\gamma^{-1}\log(\alpha_1)$ then:

$$E[x] - E[w^o(x)] - E[xg(x)] + \gamma^{-1}\ln(\alpha_1) \geq 0$$

- Substituting for $g(x) = g(x,\gamma)$ and $\alpha_1 = \alpha_1(\gamma)$ define $Q_0(\gamma)$ as:

$$\begin{aligned} & E[x] - E[w^o(x)] \\ & -E\left[x\frac{e^{\gamma\overline{w}} - e^{\gamma w^o(x)}}{e^{\gamma\overline{w}} - E\left[e^{\gamma w^o(x)}\right]}\right] + \gamma^{-1}\log\left(\frac{1 - E\left[e^{\gamma w^o(x) - \gamma\overline{w}}\right]}{E\left[e^{-\gamma w^o(x)}\right] - e^{-\gamma\overline{w}}}\right) \end{aligned}$$

- From the theorem $Q_0(\gamma^*) \geq 0$.

Identification in the Pure Moral Hazard Model

The identified set

- This inequality $Q_0(\gamma^*) \geq 0$ restricts the set of admissible γ.
- Are there any other restrictions? The short answer is no.
- Define Γ, a Borel set of risk aversion parameters, as:

$$\Gamma \equiv \{\gamma > 0 : Q_0(\gamma) \geq 0\}$$

Theorem (Gayle and Miller, 2015)

Consider any data generating process for (x_n, w_n). Then Γ is and sharp and tight. Moreover if Γ is empty the process was not generated by a principal agent model in the class described above.

- The identified set of risk parameters defined Γ has a simple empirical analogue.
- Suppose we have N cross sectional observations on (x_n, w_n) on identical firms and their managers.
- To estimate $Q_0(\gamma)$, we replace $\overline{w}$ with $\overline{w}^{(N)} \equiv \max\{w_1, \ldots, w_N\}$ and substitute sample moments for their population corresponding expectations, to obtain upon rearrangement:

$$\begin{aligned}
Q_0^{(N)}(\gamma) \equiv & \sum_{n=1}^{N} (x_n - w_n)/N \\
& - \sum_{n=1}^{N} x_n \left(e^{\gamma \overline{w}^{(N)}} - e^{\gamma w_n}\right) \Big/ \sum_{n=1}^{N} \left(e^{\gamma \overline{w}^{(N)}} - e^{\gamma w_n}\right) \\
& + \gamma^{-1} \log\left[\sum_{n=1}^{N} \left(e^{\gamma \overline{w}^{(N)}} - e^{\gamma w_n}\right)\right] \\
& + \gamma \log\left[\sum_{n=1}^{N} e^{\gamma\left(\overline{w}^{(N)} - w_n\right)} - N\right]
\end{aligned}$$

- Our tests are based on the fact that if $\gamma \in \Gamma$ then sampling error is the only explanation for why $Q_0^{(N)}(\gamma)$ might be negative.
- Clearly $Q_0^{(N)}(\gamma)$ converges at the rate of its slowest converging component.
- For simplicity suppose there exists some $\overline{x} < \infty$ such that $g(x) = 0$ for all $x > \overline{x}$.
- In words, there is a revenue threshold that shirking cannot achieve.
- Thus compensation is flat at $\overline{w}$ for all profits levels above $\overline{x}$, and $\overline{w}^{(N)}$ converges to $\overline{w}$ at a faster rate than $\sqrt{N}$.
- Since all the other components of $Q_0^{(N)}(\gamma)$ are sample moments, we conclude $Q_0^{(N)}(\gamma)$ converges at rate $\sqrt{N}$.

Empirical Implementation

A Test

- Denote by $\Gamma_{\delta}^{(N)}$ the set of risk aversion parameters that asymptotically covers the observationally equivalent set of $\gamma > 0$ with probability $1 - \delta$.
- For the critical value c_{δ} associated with test size δ, this set is defined:

$$\Gamma_{\delta}^{(N)} \equiv \left\{ \gamma > 0 : \min \left\{ 0, \sqrt{N} Q_{0}^{(N)} \left(\gamma \right) \right\}^{2} \leq c_{\delta} \right\}$$

- A consistent estimate of c_{δ} for given δ can be determined numerically by following subsampling procedures in Chernozhukov, Hong and Tamer (2007).
- Intuitively, if $\sqrt{N} Q_{0}^{(N)} \left(\gamma \right)$ is negative and large in absolute value for all $\gamma > 0$ we reject the null hypothesis that the pure moral hazard model generated the data.
- On the other hand if $\sqrt{N} Q_{0}^{(N)} \left(\gamma^{**} \right)$ is small in absolute value, or positive, we do not reject the null hypothesis that γ^{**} belongs to the identified set.