Notes on Moral Hazard

Jonathan Dworak

February 22, 1999

1 Introduction

In moral hazard models we are concerned about the ability of an agent to take unobserveable actions once contracts have been agreed upon. Since these actions may not be specified directly in the contract, the terms of the arrangement must give the agent private incentives to act in the interests of the principal. In most contexts, the result is that the principal must evaluate the tradeoff between insurance and incentives. Below I focus on the simplest model that still illustrates the general features of moral hazard, the two-action model. Reading this section requires familiarity with Lagrangians. Then I turn to the more tractable linear model which will be very important in the discussion of the multitask problem.

2 The two-action moral hazard model

A risk-neutral principal seeks to contract with a risk-averse agent. The agent operates a technology that produces output valuable to the principal. How much output is produced is random, and depends on the effort of the agent, who may choose a high level of effort or a low level of effort. For simplicity, let us assume there are only a finite number of outcomes (possible levels of output). We will denote these outcomes, in increasing order, by $y \in \{y_1, \ldots, y_N\}$. If the agent chooses a high level of effort, the probability of outcome y_n is $p_n > 0$, while if a low effort level is chosen, the probability of outcome y_n is $q_n > 0$. Think, for example, that q's and p's are such that higher outcomes are more likely under high effort than low effort. One way to formalize this assumption is the *monotone liklihood ratio property: (MLRP)*

$$n' > n \Rightarrow \frac{q_{n'}}{p_{n'}} < \frac{q_n}{p_n}$$

That is, the ratio q/p is decreasing in n.

We may think of the outcomes y as representing dollar values to the principal. The principal seeks only to maximize the expected amount of money he receives. The agent, on the other hand, is risk-averse with utility function u. Also, there is a cost of taking high effort, given by c. The principal's problem then is to choose payments for the agent that depend on the outcome acheived. Denote these payments by $\{s_n\}$. Let us for now assume that there is no information problem; the principal may force the agent to take the high action. The principal must choose these payments in such a way as to maximize his expected profit subject to the agent's participation constraint. This translates into minimizing the expected value of payments to the agent:

$$\min_{\{s_n\}} \sum_{n=1}^{N} p_n s_n \text{ subject to}$$

$$\sum_{n=1}^{N} p_n u(s_n) - c \geq 0 \qquad (IR)$$

The Lagrangian associated with this problem is

$$\sum_{n=1}^{N} p_n s_n + \lambda \left(\sum_{n=1}^{N} p_n u(s_n) - c \right)$$

with typical first-order condition

$$p_n + \lambda p_n u'(s_n) = 0 \Rightarrow u'(s_n) = -\frac{1}{\lambda}$$

If we assume that the agent is risk averse ($u'' < 0$), the above equation implies that s_n is constant across n[1]. When there is no moral hazard problem, the agent is given full insurance; the principal bears all the risk. This first-best solution is the outcome whenever a risk-averse agent interacts with a risk-neutral one; all pareto optima have full insurance for the risk-averse party.

However, suppose that the principal cannot observe the agent's effort choice. The agent must find it in his own interest to choose high effort. When s_n is constant (at, say, s^*) his expected utility from high effort is $u(s^*) - c$ while from low effort the expected utility is $u(s^*)$. The agent has no incentive to take the high effort on his own. This is the classic moral hazard problem; once an agent has full insurance, he will always choose to take the least costly action, rather than more costly actions that increase social utility. When the principal cannot observe the agent's effort, he must take these incentives into account when designing a contract:

$$\min_{\{s_n\}} \sum_{n=1}^{N} p_n s_n \text{ subject to}$$

$$\sum_{n=1}^{N} p_n u(s_n) - c \geq 0 \qquad (IR)$$

$$\sum_{n=1}^{N} p_n u(s_n) - c \geq \sum_{n=1}^{N} q_n u(s_n) \qquad (IC)$$

[1]Note that the way we have set up the minimization problem implies that the multiplier λ is negative.

The second constraint expresses the condition that the agent wish to choose the high cost action. The Lagrangian becomes:

$$\sum_{n=1}^{N} p_n s_n + \lambda \left(\sum_{n=1}^{N} p_n u(s_n) - c \right) + \mu \left(\sum_{n=1}^{N} p_n u(s_n) - c - \sum_{n=1}^{N} q_n u(s_n) \right)$$

with typical first-order condition

$$p_n + \lambda p_n u'(s_n) + \mu u'(s_n)(p_n - q_n) = 0$$

which can be rewritten (since $p_n > 0$) as

$$-\frac{1}{u'(s_n)} = \lambda + \mu \left(1 - \frac{q_n}{p_n} \right)$$

What does this say? First, note that if the IC constraint didn't bind, we would have $\mu = 0$, and we would be back in the full insurance case. Therefore the IC must bind; the agent is made just indifferent between the high and low action. Secondly, keep in mind that the multiplier μ is negative; thus, the right-hand side is increasing in the liklihood ratio $\frac{q_n}{p_n}$. This means that marginal utility must be increasing in the liklihood ratio as well. Since $u'' < 0$, this means that the payments s_n decrease in the liklihood ratio $\frac{q_n}{p_n}$. This should make sense; the agent is paid less ("punished") for outcomes that are relatively more likely under the low action than the high action. Similarly, the payment is higher for outcomes that are relatively more likely under the high action than the low action. Note the interesting fact that the payments don't depend on the y's at all, but only on the liklihood ratios. We may have thought that the principal would like to pay the agent more for higher outcomes; the above equation suggests that in the general case that may not occur. If we assume the MLRP, however, we will restore the monotonicity payment in output. The other place the y's come in is in the choice of effort. Here I have simply assumed that the principal wishes the agent to take the high action. The above program shows the minimal cost of inducing the agent to take the high action. This cost, of course, is higher than it would be in the absence of moral hazard because the agent bears some risk. Although he is still held to his outside option (the IR is satisfied with equality), the expected level of the payments is higher since they are now risky. The principal could induce the low action, however, by offering the full insurance contract. Secondly, what would happen if the agent were risk-neutral? Then, intuitively, there is no cost of making the agent bear a lot of risk. One possibility is to choose payments such that $s_n = y_n - p$ where p satisfies

$$p = \sum_{n=1}^{N} p_n y_n - c$$

In this scheme, the principal "sells" the firm to the agent for price p, making the agent the residual claimant (and so inducing the efficient action) but sucking back all the surplus.

These considerations identify the key lesson of all moral hazard models: *the trade-off between incentives and insurance.* If there is no moral hazard problem, the principal would always fully insure the risk-averse agent. If the agent is risk-neutral, there is no moral hazard problem; the principal can give the agent high-powered incentives by making him the residual claimant. When actions are unobserveable and the agent is risk-averse, the principal must balance the increased incentives associated with a more risky payment scheme with the increased level of payments he must make to compensate the agent for the risk.

3 The linear-normal moral hazard model

As we have seen, in the general case the payment scheme depends very much on the relative liklihood of outcomes under high and low actions. Unfortunately, it is hard even in the simple two-action case to guarantee simple properties like monotonicity. Extending the model to more than two actions is even more difficult. Many real-world incentive schemes are simple profit sharing systems, however, and we would like to study how these operate. Also, continuous effort choice seems to be a feature of reality. A simple model which allows us to examine these (and a host of other issues) is the linear-normal model. In this model, the principal is restricted to choosing payments that are a linear function of output. The agent chooses effort e continuously. Output is given by

$$y = e + \varepsilon$$

where ε is normally distributed with mean zero and variance σ^2. The linearity restriction amounts to the condition that the share s paid to the agent satisfy

$$s = a + by$$

The agent's risk aversion is summarized by his attitudes toward the mean and variance of any gamble. Let μ denote the mean of the payment scheme he receives and s^2 denote its variance. His expected utility from taking action e and receiving the payment scheme is

$$\mu - c(e) - \frac{1}{2}\rho s^2$$

where the parameter ρ indexes the agent's risk-aversion. If $s = a + by$, then by assumption the mean of payments is $a + be$ and the variance is $b^2\sigma^2$. Thus the agent taking action e receives expected utility

$$a + be - c(e) - \frac{1}{2}\rho b^2\sigma^2$$

If the agent is maximizing his expected utility he will choose effort to satisfy the first-order condition $b = c'(e)$. Thus, the principal's problem becomes

$$\begin{aligned}
&\max_{a,b,e} (1-b)\,e - a \text{ subject to} \\
&a + be - c(e) - \frac{1}{2}\rho b^2 \sigma^2 \geq 0 \qquad \text{(IR)} \\
&b = c'(e) \qquad \text{(IC)}
\end{aligned}$$

Making the (IR) into an equality and substituting for a in the objective function results in

$$\begin{aligned}
&\max_{b,e} e - c(e) - \frac{1}{2}\rho b^2 \sigma^2 \qquad \text{subject to} \\
&b = c'(e)
\end{aligned}$$

and plugging in the IC results in the unconstrained problem

$$\max_{e} e - c(e) - \frac{1}{2}\rho \left(c'(e)\right)^2 \sigma^2$$

which has first-order condition

$$1 - c'(e) - \rho c'(e) c''(e) \sigma^2 = 0$$

which can be rewritten (using the fact that $b = c'(e)$) as

$$b = \frac{1}{1 + \rho \sigma^2 c''(e)}$$

Here we have illustrated the trade-off between insurance and incentives in a very simple formula. Assuming convex costs ($c'' > 0$) it is clear that b is between zero and one. If b is zero, the agent's payment does not vary with output and he receives full insurance. If $b = 1$, the agent is the residual claimant. Our intuitions about when these cases should occur is confirmed in the above expression. When either $\rho = 0$ (the agent is risk-neutral), or $\sigma^2 = 0$ (there is no uncertainty), we choose $b = 1$ and give the agent full incentives. The power of these incentives is declining in the risk-aversion of the agent and in the riskiness of the underlying technology. The interesting term is c''; the incentives depend on the degree of cost convexity. In this case, this term doesn't add much, but we will find that the multi-task moral hazard model is very much about the interaction of effort variables in an agent's cost function, and how that affects incentives.