Continuous Action Moral Hazard Notes

Jacob Kohlhepp

October 29, 2021

1 Moral hazard with Continuous Actions

To help reinforce the technical aspects of solving the moral hazard problem with continuous actions I will do a problem where both principal and the agent are risk averse. So there will be "risk sharing."

1.1 Setup

Output is given by a random variable q which depends on effort (a), chosen by an agent. Both the principal and the agent are risk averse, with increasing concave and twice differentiable utility functions $v(x), u(x)$. v for the principal and u for the agent. Effort is costly, so the agent's utility is given by: $u(q, a) = u(w(q)) - c(a)$. The agent has outside option $\bar{u}$ The principal views wages and output as interchangeable, so its utility from output q and wages w is: $v(q - w(q))$.

We assume that the principal makes a TIOLI offer to the agent.

1.2 First-best

First assume effort is observed and contractible. Solve for the first-best wages and effort.

Principal's problem:

$$\max_{a, w(q)} \mathbb{E}[v(q - w(q))|a]$$

s.t. $\mathbb{E}[u(w(q)) - c(a)|a] \geq \bar{u}$

Langrangian:

$$\mathcal{L} = \int \left\{ v(q - w(q)) + \lambda \Big(u(w(q)) - c(a) - \bar{u} \Big) \right\} f(q|a) dq$$

We argue that $\lambda > 0$: Suppose not. Then we could offer another contract with slightly lower wages at every point. This raises principal utility, contradicting optimality.

Point-wise maximize:

$$\max_{w} v(q - w) + \lambda u(w)$$

$$\frac{v'(q - w)}{u'(w)} = \lambda$$

$$\implies u'^{-1} \left\{ \frac{v'(q - w)}{\lambda} \right\} = w(q)$$

Optimal wages are not fixed - there is risk sharing. Suppose we impose exponential utility, so that $v(x) = -e^{-r_1 x}$ and $u(x) = -exp(-r_2 x)$. Then we have that:

$$\frac{r_1 exp(-r_1(q-w))}{r_2 exp(-r_2 w)} = \frac{r_1}{r_2} exp\Big\{ w(r_1 + r_2) - r_1 q \Big\} = \lambda$$

Solve for wages:

$$w(q) = \frac{1}{r_1 + r_2}\left(log\left\{\frac{r_2}{r_1}\lambda\right\} + r_1 q\right)$$

So we have that wages are linear, and the risk the principal imposes on the agent (the coefficient on q) depends on $\frac{r_1}{r_1+r_2}$, which is the risk aversion of the principal relative to a measure of overall risk aversion. This demonstrates optimal risk sharing. If risk-aversion is the same for both players, then the coefficient on q is 1/2, demonstrating that risk is split equally.

Deriving this general form for optimal wages is about as far as I go here. Optimal effort maximizes principal utility subject to the optimal wage equation derived above, with IR binding. We cannot do the simplifications we did before with constant wages (when principal was risk neutral).

1.3 Moral Hazard

Now suppose effort a is not observed. What is the optimal wage schedule?

For this problem, focus on the steps I use rather than on this particular problem.

1. **Setup the Problem:**

$$\max_{a, w(q)} \mathbb{E}[v(q - w(q))|a]$$

s.t.

$$\mathbb{E}[u(w(q)) - c(a)|a] \geq \bar{u} \tag{IR}$$

$$a \in \arg\max_{\tilde{a}} \mathbb{E}[u(w(q)) - c(a)|\tilde{a}] \tag{ICs}$$

2. **Derive the IC-FOC.**

Consider the agent's problem of choosing an effort given a wage:

$$\max_a \mathbb{E}[u(w(q)) - c(a)|\tilde{a}] = \int \Big(u(w(q)) \Big) f(q|a) dq - c(a)$$

FOC:

$$\int u(w(q)) f_a(q|a) dq - c'(a) = 0$$

where we are bringing the derivative inside the integral.[1]

3. **Replace the ICs with the IC-FOC from the last step.**

$$\max_{a, w(q)} \mathbb{E}[v(q - w(q))|a]$$

[1] For those interested in the math, this is probably justified using dominated convergence theorem or monotone convergence theorem or something similar.

s.t.

$$\mathbb{E}[u(w(q)) - c(a)|a] \geq \bar{u} \tag{IR}$$

$$\int u(w(q))f_a(q|a)dq - c'(a) = 0 \tag{IC-FOC}$$

Note: If the agent's utility conditional on effort is concave for the optimal wage, this replacement is valid. It is hard to show this concavity, so in practice people sometimes show it numerically. For this class, you will probably not be asked to show it, so you can just do this step. This approach is called the "first-order approach." In lecture the professor will review when it is indeed valid.

4. **Form the Lagrangian.**

$$\mathcal{L} = \int \Big\{ v(q - w(q)) \Big\} f(q|a)dq + \lambda \int \Big(u(w(q)) - c(a) - \bar{u} \Big) f(q|a)dq + \mu \Big(\int u(w(q))f_a(q|a)dq - c'(a) \Big)$$

Combine everything into one integral:

$$\mathcal{L} = \int \Big\{ v(q - w(q)) + \lambda \Big(u(w(q)) - c(a) - \bar{u} \Big) + \mu \Big(u(w(q)) \frac{f_a(q|a)}{f(q|a)} - c'(a) \Big) \Big\} f(q|a)dq$$

5. **Point-wise Maximize the Wage.**

Fix q, w:

$$\max_w v(q - w) + \lambda u(w) + \mu u(w) \frac{f_a(q|a)}{f(q|a)}$$

FOC:

$$-v'(q - w) + \lambda u'(w) + \mu u'(w) \frac{f_a(q|a)}{f(q|a)} = 0$$

$$u'(w) \Big(\lambda + \mu \frac{f_a(q|a)}{f(q|a)} \Big) = v'(q - w)$$

$$\Big(\lambda + \mu \frac{f_a(q|a)}{f(q|a)} \Big) = \frac{v'(q - w)}{u'(w)}$$

This is true for all q, so:

$$\Big(\lambda + \mu \frac{f_a(q|a)}{f(q|a)} \Big) = \frac{v'(q - w(q))}{u'(w(q))}$$

This implies some sort of risk sharing. Take the case of exponential utility. Then the RHS becomes:

$$\frac{r_1 exp(-r_1(q - w))}{r_0 exp(-r_0 w)} = \frac{r_1}{r_0} exp\Big\{ w(r_1 + r_0) - r_1 q \Big\}$$

Let's say they are equally risk-averse, that is $r_0 = r_1 = r$. Then the above becomes:

$$exp\{2rw(q) - rq\} = \Big(\lambda + \mu \frac{f_a(q|a)}{f(q|a)} \Big)$$

$$\leftrightarrow w(q) = \frac{1}{2r} log \Big(\lambda + \mu \frac{f_a(q|a)}{f(q|a)} \Big) + \frac{1}{2} q$$

We can see that this optimal wage is similar to the first-best but with an extra component that depends on q (the likelihood ratio). This extra component demonstrates a trade-off between incentives and risk: to get more effort we need to impose risk. But imposing risk pushes us farther away from optimal wages, which have equal sharing of risk.

6. **Argue that $\mu, \lambda > 0$, that is both constraints bind.**

   The argument for $\lambda > 0$ is the same as in two-actions: if it was 0, then this implies the IR is slack, meaning that we can construct a new contract $\tilde{w}$ from the old contract w such that $u(\tilde{w}(q)) = u(w(q)) - \epsilon)$ but IR still is satisfied. This of course raises principal utility, contradicting optimality.

   Arguing the sign of μ is harder. First note that $\mu \neq 0$ because the IC-FOC is an equality constraint. So it is either positive or negative. The proof that it is positive follows from MLRP and the wage function we derived in the last step. The proof is in the lecture notes and will be covered Monday. It is good to know the intuitive idea of the proof, which is that negative μ will make marginal utility at optimal a negative, which cannot be (since then we would want to slack off by providing a little bit less effort). This violates the original IC conditions.