HAD6750 Lecture 2: Moral Hazard

Alex Hoagland, PhD

Last updated: 2026-01-06

1 Lecture 2: Moral Hazard / Principal–Agent

Reminders

Start with a discussion of referee report guidelines:

- Think: contribution, flaws/assumptions, presentation, usefulness, theoretical/empirical extensions.
- No decision for the journal, obviously, but everything else!

2 Foundations: Risk Aversion and Expected Utility

Risk aversion is captured by concave utility:

$$u''(\cdot) < 0,$$

which implies (Jensen's inequality) that

$$u(\mathbb{E}[S]) \;>\; \mathbb{E}[u(S)].$$

Can you draw the **certainty equivalent** and corresponding *risk premium* on this figure?

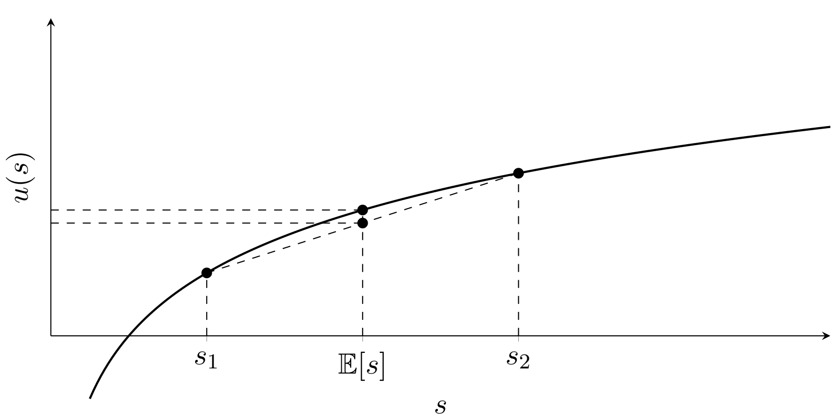

Figure 1: Concave utility and the certainty equivalent

Risk premium and certainty equivalent

For a lottery over wealth/outcomes S with mean $\mathbb{E}[S]$, define the *certainty equivalent* c_e by

$$u(c_e) = \mathbb{E}[u(S)].$$

The *risk premium* is

$$\mathrm{RP} = \mathbb{E}[S] - c_e,$$

interpreted as "the amount someone would pay to avoid a bet."

Expected utility formula

For a two-state (or an N-state) lottery:

$$\mathbb{E}U = p_1 u(s_1) + p_2 u(s_2) \quad \Rightarrow \quad \sum_{i=1}^{N} p_i u(x_i),$$

and in general (continuous outcomes):

$$\mathbb{E}U = \int u(x)\, dF(x),$$

where $F(x)$ denotes the probability distribution for a continuous outcome.

3 Effort Model: Principal–Agent (PA)

- **Principal:** like the manager. Wants to maximize profit (assumed to be risk neutral), but doesn't observe effort. Hence, they choose a wage schedule $w(\pi)$.
- Agent wants to maximize expected utility: $V = \mathbb{E}\big[v(w(\pi)) \mid e\big] - g(e)$. Assumed to be risk averse here.
- The main problem from private information is **hidden effort** here

Principal's problem:

Choose a wage schedule $w(\cdot)$ to minimize expected wage under high effort, subject to participation and incentive compatibility:

$$\min_{w(\cdot)} \int w(\pi)\, dF(\pi \mid e_H)$$

subject to

$$\int v(w(\pi))\, dF(\pi \mid e_H) - g(e_H) \geq 0 \qquad \text{(Participation)},$$

$$\int v(w(\pi))\, dF(\pi \mid e_H) - g(e_H) \geq \int v(w(\pi))\, dF(\pi \mid e_L) - g(e_L) \qquad \text{(Incentive Compatibility)}.$$

4 Example parameterization:

Setup

Consider

$$v(w) = \ln w \quad \text{(risk averse)}, \qquad g(e) = e^2 \quad \text{(costly effort)}.$$

- Reservation utility: $U_A = 10$ ("reservation wage").
- Output (or surplus) takes two values:

$$Q \in \{400,\, 100\}.$$

- Let $p = \Pr(Q = 400)$, where:

$$p = \begin{cases} \frac{2}{3} & \text{if } e = 1, \\ \frac{1}{3} & \text{if } e = 0. \end{cases}$$

Effort Observability vs. Unobservability

Case 1: Effort is observed

If effort is observed, the principal can condition on e and choose a wage w for each level of e.

If $e = 0$:

$$\mathbb{E}U_A = \frac{1}{3}\ln w - 0 + \frac{2}{3}\ln w - 0 \geq 10 \quad \Rightarrow \quad \ln w \geq 10 \quad \Rightarrow \quad w^\star = 100.$$

If $e = 1$:

$$\mathbb{E}U_A = \frac{1}{3}\ln w - 1 + \frac{2}{3}\ln w - 1 \geq 10 \quad \Rightarrow \quad \ln w \geq 11 \quad \Rightarrow \quad w^\star = 121.$$

What would the principal choose to enforce? How would they enforce it?

$$\mathbb{E}U_P(e = 0) = \frac{1}{3}(400 - 100) + \frac{2}{3}(100 - 100) = 100,$$

$$\mathbb{E}U_P(e = 1) = \frac{1}{3}(100 - 121) + \frac{2}{3}(400 - 121) = 179.$$

Case 2: Effort is unobserved

When e is unobserved, the contract must be tied to output, so wages are (w_L, w_H).

Participation if $e = 0$.

$$\mathbb{E}U_A(0) = \frac{1}{3}\ln w_H + \frac{2}{3}\ln w_L \geq 10.$$

The principal then solves the least-cost (expected wage) problem:

$$\min_{w_L, w_H} \frac{1}{3}w_H + \frac{2}{3}w_L \quad \text{s.t.} \quad \frac{1}{3}w_H^{1/2} + \frac{2}{3}w_L^{1/2} = 10,$$

which gives:

$$w_H = 4\Big(w_L - 30\ln w_L + 225\Big).$$

Participation if $e = 1$.

$$\mathbb{E}U_A(1) = \frac{1}{3}\big(\ln w_L - 1\big) + \frac{2}{3}\big(\ln w_H - 1\big) \geq 10,$$

which is written equivalently as

$$\frac{1}{3}w_L^{1/2} + \frac{2}{3}w_H^{1/2} = 11.$$

Incentive Compatibility. To induce high effort, there must be a constraint so that w_L ensures low effort does *not* satisfy the participation threshold under the offered contract:

$$\frac{1}{3}\ln w_H + \frac{2}{3}\ln w_L < 10.$$

Solving graphically by setting the relevant constraints equal yields:

$$w_H^\star = 144, \qquad w_L^\star = 81.$$

5 Zeckhauser (1970): Insurance, Treatment, and Moral Hazard

In this model, we have patients choosing whether or not to seek care based on:

- Their own health
- The costs of being sick
- The (OOP) costs of health production

Baseline risk without treatment

Consumers are either sick or healthy:

$$p = \Pr(\text{sick}), \qquad 1 - p = \Pr(\text{healthy}).$$

Without treatment, expected utility is

$$EU_c = p\,U(w - s) + (1 - p)\,U(w),$$

with s representing the sickness-related loss. **How do we interpret the loss in this way? Is there an alternative approach?**

Treatment choice (ex post)

Suppose a treatment effort/quantity h exists, producing benefit $f(h)$ with

$$f'(h) > 0, \qquad f''(h) < 0.$$

Now the expected utility is

$$EU_c|\text{treatment} = p\,\underbrace{U(w - s + f(h) - h)}_{\text{ex-post utility}} + (1 - p)\,U(w),$$

Then (conditioning on being sick) the consumer chooses h to maximize

$$w - s + f(h) - h.$$

The ex-post optimal treatment level satisfies

$$f'(h^\star) = 1.$$

Insurance contract (ex ante), Illness is still verifiable

What about before knowing you are sick?

- Without insurance, solve the same problem but taking p into account
- What about under an insurance contract?

With premium π and transfer τ when sick, the notes write expected utility as

$$EU_c = (1-p)\,U(w-\pi) + p\,U\big(w - \pi - s + \tau + f(h^\star) - h^\star\big).$$

The contract pays for "full recuperation" when you are sick (note: this is a **zero-profit constraint** that we typically impose on insurers):

$$\tau = s - f(h^\star) + h^\star,$$

with an **actuarially fair** premium

$$\pi = p\,\tau.$$

What do consumers do in the optimum here? Would you choose to get treatment if $s \leq 0$? Why or why not?

6 Unverifiable Illness and Coinsurance

Coinsurance and ex-post moral hazard

Digression: why do we call this moral hazard?

- Is (unverifiable) illness like hidden effort?
- Grossman seems to think so
- There are also ex-post treatment decisions: now h^* will change as the price of h^* changes (this is more like the **law of demand** than hidden effort)

If sickness is unverifiable, insurance often uses coinsurance: let the insurer pay $(1-\theta)h$ and the consumer pay θh. If sick, the consumer solves

$$\max_h \; w - s + f(h) - \theta h,$$

giving the first-order condition

$$f'\big(h^\star(\theta)\big) = \theta, \qquad \text{so} \qquad \frac{dh^\star(\theta)}{d\theta} \leq 0.$$

Equivalently, increasing coverage (lower θ) increases treatment utilization: *ex-post moral hazard.*

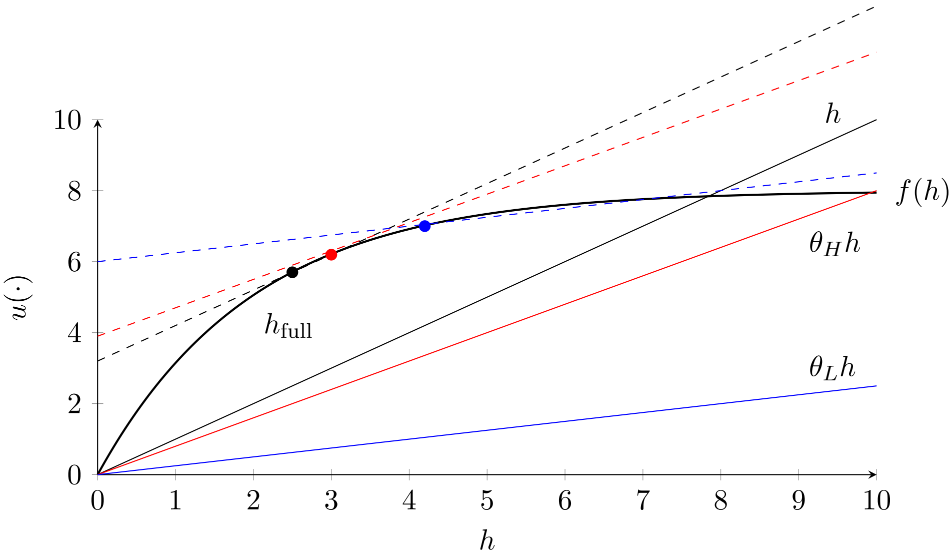

Figure 2: Coinsurance and treatment choice: lower θ implies higher $h^\star(\theta)$. **How is this the law of demand?**

Optimal Insurance Design Problem

How can we find the equilibrium?

$$\max_{\pi,\theta} \,(1-p)\,U(w-\pi) + p\,U\big(\underbrace{w-s-\pi+f(h(\theta))}_{\tilde{x}} - \theta h(\theta)\big)$$

subject to fair insurance:

$$\pi = p(1-\theta)\,h(\theta).$$

What is the Lagrangian?

Consider the first-order conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \pi} &= -(1-p)U'(w-\pi) - pU'\big(\,\cdot\,\big) - \lambda = 0,\\
\frac{\partial \mathcal{L}}{\partial \theta} &= pU'\big(\tilde{x}\big)\Big(f'(h(\theta))h'(\theta) - h(\theta) - \theta h'(\theta)\Big) + \lambda p\Big(-h(\theta) + (1-\theta)h'(\theta)\Big) = 0,\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= \pi = p(1-\theta)\,h(\theta) = 0.
\end{aligned}$$

Some things to note here:

1. Equation 1 shows the loss in marginal utility from the insurance premium needs to be equal to λ (common interpretation of the Lagrangian parameter as a "shadow price")

2. In equation 2, we can drop the p in both terms and rearrange to get

$$\underbrace{-U'(sick, treatment)[(f'(h(\theta)) - \theta)h'(\theta) - h(\theta)]}_{\text{marginal costs of paying for treatment}} = \underbrace{\lambda[(1-\theta)h'(\theta) - h(\theta)]}_{\text{utility gains from insurance (risk pooling)}}$$

7 Overconsumption and (In)Achievability of First Best

Can we achieve the first-best equilibrium?

Note: what do we mean here by the first-best equilibrium? What is the deviation from that?

When an individual is fully insured, the marginal cost of additional medical care to the patient at the time of service becomes zero, even though the actual societal cost is positive. This leads individuals to consume more healthcare than they would if they had to pay the full price, resulting in "excess consumption".

Mathetmatically: with $\theta < 1$ and using the ex-post condition $f'(h(\theta)) = \theta$, the FOC simplifies to:

$$u'(\text{sick treatment}) \cdot \big[h(\theta)\big] = \lambda\Big((1-\theta)h'(\theta) - h(\theta)\Big).$$

Since $(1-\theta)h'(\theta) < 0$ (because $1-\theta > 0$ and $h'(\theta) < 0$), whenever we have **risk aversion** (meaning $u'' < 0$), then reducing the cost of care (θ) will lead to *overconsumption* relative to first best under full insurance. That is, as the price of care drops, the FOC compensates by having lower equilibrium marginal utility of treatment, which corresponds to an increase in h^*.