Lecture 6

Moral Hazard: The Principal-Agent Model – Part I

Alexander K. Koch∗

Department of Economics, Royal Holloway, University of London

March 13, 2006

Required reading:


Supplementary reading:


Well, then, says I, what's the use you learning to do right when it's troublesome to do right and ain't no trouble to do wrong, and the wages is just the same?

Huckleberry Finn (Mark Twain, 1884, The Adventures of Huckleberry Finn)

1 Introduction

So far in this course we have seen the consequences of asymmetric information existing at the time of contracting. For example, if one party is privately informed about characteristics that are relevant to all the contracting parties a problem of adverse selection may arise. We now turn to settings where relevant private information is expected to emerge after contracts have been signed. The informational asymmetry provides the party that has access to better information with scope for opportunistic behavior. Such problems of post-contractual opportunism are typically analyzed by recasting them as a principal-agent relation: One individual (agent) acts on behalf of another individual (principal) and is supposed to advance the principal's goals. However, the agent and the principal have differing individual objectives. This leads to problems of post-contractual opportunism since the principal cannot easily determine whether the agent's actions or reports are truly serving the principal's goals rather than being self-interested behavior of the agent. Two kinds of informational asymmetries can arise:

- **hidden information**: the agent obtains better information about aspects relevant to the principal. For example, the manager in charge of sales in a particular region is likely to acquire a better knowledge of the market than his superiors.

- **hidden action (moral hazard)**: the agent undertakes an action that affects the principal's utility and this action is difficult to monitor for the principal. For example, the amount of effort or the quality of work put in by an employee is difficult to monitor directly and the principal has to rely on results of the employee's work (e.g., output produced, the number of breakdowns/repairs on manufacturing equipment etc.). Another example we will see in lecture 8: the bank handing out credit may not be able to assess the riskiness of the project chosen by the creditor.

The readings in ? provide many specific examples for problems arising from hidden information and moral hazard.

The second part of the course will deal with the analysis of the adverse effects of post-contractual private information and how contracts can be designed to mitigate these. In
particular, the information asymmetry creates a problem for the principal of providing incentives to induce appropriate behavior. The Principal-Agent framework that we will study helps deal with questions such as ‘How do you make managers work hard?’; ‘How do you ensure quality in production?’; ‘How do you induce people to study, invest in the right projects, and to save?’

2 The moral hazard model with 2 actions

Consider a principal (“she” – e.g., the owner of a firm) who wants to hire an agent (“he” – e.g., a worker for the firm) to perform some task. The task consists of the agent choosing an effort level \( e \in \{l, h\} \). The high effort level \( e = h \) comes at utility cost of \( \psi_h = \psi \) to the agent, whereas there is no disutility for the low effort level \( e = l \): \( \psi_l = 0 \). The task output \( q \in Q \) is stochastic. In the following we will treat different cases separately, where we modify the assumptions about the set of possible output values \( Q \):

- two possible outcomes: \( Q = \{q_L, q_H\} \), \( q_L < q_H \) (this lecture);
- \( n \) possible outcomes: \( Q = \{q_1, q_2, \ldots, q_n\} \), \( q_1 < q_2 < \cdots < q_n \) (Lecture 7);

We assume that the principal is risk neutral with utility equal to her profits

\[
\Pi(q, w) = q - w,
\]

where \( q \) is the realized output and \( w \) is the wage paid to the agent. The agent is assumed to be a von Neumann-Morgenstern utility maximizer with Bernoulli utility \( u(w) \), where \( u'(\cdot) > 0 \) and \( u''(\cdot) < 0 \), i.e, the agent is risk averse. Moreover, we assume that the agent’s utility is separable in income and effort costs:

\[
U(w, e) = u(w) - \psi_e.
\]

The principal makes a contract offer to one agent from the large pool of applicants for the position. Assume that an agent can realize expected utility \( \bar{U} \) if he pursues the best other option available to him. Thus, his reservation utility level for accepting the contract offered by the principal is \( \bar{U} \). We will normalize \( \bar{U} = 0 \) to simplify formulas (in the problem sets we will encounter a non-zero reservation utility). Moreover, assume that the principal makes zero profits when no agent is hired.

\[1\] This convenient because it allows us to ignore income effects on the cost of effort. These effects are of lesser interest in the analysis of the typical moral hazard problem.
The rules of the game and timing are as follows:

1. The principal makes a contract offer to an agent from the pool of job applicants.
2. The agent can refuse the offer (and then the principal will start at step 1 with another job applicant) or accept the offer.
3. If the agent accepts the contract, he chooses an effort level \( e \in \{l, h\} \).
4. Output \( q \in Q \) is realized and the contractually specified payments are made.

3 The basic model variant with 2 outcomes

In this basic model variant we assume that the task output can take on two possible values \( Q = \{q_L, q_H\} \), where we will assume for simplicity that \( q_L = 0 < q_H = 1 \). The probability of producing high output is increasing in the effort the agent exerts:

\[
p_h \equiv Pr(q = 1|e = h) > p_l \equiv Pr(q = 1|e = l) > 0.
\]

3.1 The symmetric information benchmark and the first best solution

It is instructive to consider first the symmetric information benchmark case, where the agent’s effort choice is observable and verifiable by a court that can enforce contracts. The contract between the principal and the agent can condition monetary transfers between parties on any variable that is observable and verifiable. In the symmetric information case, the contract stipulates wages as a function of actual effort \( e \) and realized output \( q \): \( w(e, q) \). If the agreed upon effort, say \( \hat{e} \), is exerted and high output realizes the principal pays wage \( w_1 = w(e = \hat{e}, q = 1) \) and if low output realizes she pays wage \( w_0 = w(e = \hat{e}, q = 0) \). We will summarize such a contract using the following notation: \( Y = [e, w_0, w_1] \), where \( e \) is the effort that the contract implements (i.e., which the agent will actually choose) and \( w_0/w_1 \) are the wages if low/high output realizes.

In sum, the principal and agent face the most favorable circumstances for avoiding contracting problems: they can enter a binding agreement at the contracting stage on the effort, say \( \hat{e} \), to be exerted since the actual effort level \( e \) is not hidden from any of the contracting parties. The idea is that a court can hand out sufficiently severe punishments for breach of contract (expressed mathematically by an infinite loss in utility) so that no party will ever want to violate the contract terms in a way that can be proven before court. That is, in the symmetric information benchmark case, the agent will deliver the agreed upon effort \( e = \hat{e} \) (or face a penalty \( w(e, q) = -\infty \) for \( e \neq \hat{e} \)) and the principal will pay the promised wage \( w(\hat{e}, q) \).

4
To determine the optimal contract $Y^{SI} = [e, w_0, w_1]$ the principal offers a contract that maximizes her profits subject to the agent’s *individual rationality (IR)* constraint. The IR constraint guarantees that the agent is willing to accept the contract: his expected utility derived from the contract must be at least as high as that offered by the outside option. That is, the principal solves the following program:

$$\max_{e, w_0, w_1} p_e (1 - w_1) + (1 - p_e) (-w_0) \quad \text{(P:SI)}$$

subject to

$$p_e u(w_1) + (1 - p_e) u(w_0) - \psi_e \geq \bar{U} = 0. \quad \text{(IR)}$$

We will solve this problem in two steps:

1. For each possible effort level we determine separately the wages which implement this effort level at the least cost to the principal.

2. Using the implementation costs derived in the first step we look for the effort level which maximizes the principal’s profits.

**Implementing effort $e = h$.** Since the contract can condition on the actual effort the agent exerted the principal only needs to make sure that the agent is willing to accept the contract requiring effort $e = h$. To minimize her wage bill, the principal will choose $w_0$ and $w_1$ so that the agent’s participation constraint is binding (i.e., holds with $=$) – why offer the agent a higher expected utility (i.e., a higher wage payment) than necessary to convince him to accept the contract?

Hence, we can write down the Lagrangian:

$$\mathcal{L} = p_h (1 - w_1) + (1 - p_h) (-w_0) - \lambda \left[ \psi - p_h u(w_1) - (1 - p_h) u(w_0) \right],$$

where $\lambda$ is the Lagrange multiplier for the IR constraint. The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial w_1} = -p_h + \lambda p_h u'(w_1) = 0 \quad \Leftrightarrow \quad \lambda = \frac{1}{u'(w_1)}, \quad (2)$$

$$\frac{\partial \mathcal{L}}{\partial w_0} = -(1 - p_h) + \lambda (1 - p_h) u'(w_0) = 0 \quad \Leftrightarrow \quad \lambda = \frac{1}{u'(w_0)}, \quad (3)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = p_h u(w_1) + (1 - p_h) u(w_0) = \psi. \quad (4)$$

It is easy to check that the second-order conditions satisfy the requirements for the solution to be a maximum.

---

2 The superscript $SI$ serves as labels for the symmetric information case.

3 Note that the effort level is fixed and we are maximizing the expected profit over $w_0$ and $w_1$. 
Conditions (2) and (3) yield the following optimality condition:

\[
\frac{1}{u'(w_0)} = \frac{1}{u'(w_1)} \Leftrightarrow w_0 = w_1, \tag{5}
\]

since \(u'' \neq 0\). The principal is risk neutral while the agent is risk averse, and therefore it is optimal to offer the agent full insurance: \(w_0 = w_1 = w\). Another way to see this, look at the contracting parties’ indifference curves in \((w_0, w_1)\) space. The slope of the principal’s indifference line (she is risk neutral!) is:

\[
\frac{dw_1}{dw_0} = -\frac{1 - ph}{ph}. \tag{6}
\]

The slope of the agent’s indifference curve is:

\[
\frac{dw_1}{dw_0} = -(1 - ph) \frac{u'(w_0)}{ph u'(w_1)}. \tag{7}
\]

Suppose that \(w_1 > w_0\) and the IR constraint is binding (point A in Figure 1). Since the agent’s utility function is strictly concave (\(u''(\cdot) < 0\)) we then have \(u'(w_1) < u'(w_0)\), so that the agent’s marginal rate of substitution between the wage in the high output state and that in the low output state is lower than the corresponding marginal rate of substitution for the principal:

\[
-\frac{(1 - ph)}{ph} \frac{u'(w_0)}{u'(w_1)} < -\frac{1 - ph}{ph}. \tag{8}
\]

Thus, the principal can decrease \(w_1\), lowering the expected wage payment by \(ph \frac{dw_1}{d w_0}\), while the increase in \(w_0\) necessary to keep the agent at expected utility level \(\bar{U} = 0\) pushes up the expected wage bill by only \((1 - ph) \frac{dw_0}{d w_1}\). As one can easily see in the figure, the principal can leave the agent on the same indifference curve, while moving to a higher indifference curve herself as long as \(w_0 \neq w_1\). Thus, full insurance \(w_0 = w_1 = w\) is optimal (point B in Figure 1).

Using the above insights, we can derive the optimal wages. The principal chooses \(w = w_0 = w_1\) so that the IR constraint binds:

\[
u(w) - \psi = 0 \quad \Rightarrow \quad w_0 = w_1 = w^{SI}(h) = u^{-1}(\psi).
\]

**Implementing effort** \(e = l\). The procedure is analogous to the one above. Again, the agent is fully insured against fluctuations in wages and the principal sets \(w = w_0 = w_1\) so that the IR constraint binds:

\[
u(w) = 0 \quad \Rightarrow \quad w_0 = w_1 = w^{SI}(l) = u^{-1}(0).
\]
Finding the profit maximizing effort level to implement. Comparing expected profits for the principal, we see that implementing effort $e = h$ is optimal for the principal if and only if the expected profit is positive,

$$p_h - w^{SI}(h) \geq 0,$$  \hspace{1cm} (9)

and the benefit from more effort exceeds its cost to the principal,

$$\underbrace{p_h - p_l} \geq \underbrace{w^{SI}(h) - w^{SI}(l) = u^{-1}(\psi) - u^{-1}(0)}$$  \hspace{1cm} (10)

benefit \hspace{1cm} cost of effort in units of output

of effort $e = h$

instead of $e = l$

Otherwise, it is optimal to implement effort $e = l$ (if $p_l - w^{SI}(l) \geq 0$) or offer no contract at all (if $p_l - w^{SI}(l) < 0$). Note that the first-best solution arising under symmetric information is Pareto optimal: the principal’s profit is maximized leaving the agent no worse off than under the outside option. Thus, the socially efficient amount of effort is actually implemented.
Proposition 1
The first-best allocation obtained under symmetric information is Pareto-optimal: the risk-neutral principal offers full insurance to the risk-averse agent (efficient risk allocation) and implements the socially efficient level of effort (efficient effort choice).

3.2 The asymmetric information case and the second best solution
We now turn to the situation where the principal cannot observe the effort that the agent exerted. Now the transfers stipulated in a contract \( Y = [e, w_0, w_1] \) can only condition on the observed outcome, i.e., wages are only a function of \( q \) and not of effort \( e \) (since the latter is not observable by the principal or a court). As a consequence, even if the agent promises to deliver \( e \) when the contract is signed, the principal must make sure that this effort level is then actually chosen by the agent when faced with the wage scheme offered. Thus, the principal faces an additional constraint in her optimization program: the incentive (IC) constraint which states that effort \( e \) must maximize the agent’s utility for the given wage scheme:

\[
\max_{e, w_0, w_1} p_e (1 - w_1) + (1 - p_e) (-w_0) \quad \text{(P:AI)}
\]

subject to

\[
p_e u(w_1) + (1 - p_e) u(w_0) - \psi_e \geq \bar{U} = 0. \quad \text{(IR)}
\]

\[
e \in \arg\max_{e'} p(e') u(w_1) + (1 - p(e')) u(w_0) - \psi_e'. \quad \text{(IC)}
\]

It is easiest to understand the IC constraint when going through the individual effort levels separately. In fact, again it is useful to solve the principal’s problem going through the two steps that we used in the symmetric information setting.

Implementing effort \( e = h \). Writing down the (IC) constraint for \( e = h \)

\[
e \in \arg\max_{e'} p(e') u(w_1) + [1 - p(e')] u(w_0) - \psi_e' \]

\[
\Leftrightarrow p_h u(w_1) + (1 - p_h) u(w_0) - \psi \geq p_l u(w_1) + (1 - p_l) u(w_0). \]

Clearly, a fixed wage \( w = w_1 = w_0 \) will not do the trick. Then the agent has utility from wages \( u(w) \) regardless of whether high or low output realizes and will therefore choose the effort level which minimizes his disutility from effort: \( e = l \). As a result of asymmetric information about the agent’s effort level, full insurance is not feasible when implementing effort level \( e = h \). The principal needs to rely on realized output as indicators for whether the agent actually exerted

\[4\] The superscript AI serves as label for the asymmetric information case.
effort \( e = h \) or not. Rewriting the above incentive constraint we see that the principal must introduce a wedge between the utility received from the low-output wage and that from the high-output wage so that the agent indeed chooses \( e = h \):

\[
p_h u(w_1) + (1 - p_h) u(w_0) - \psi \geq p_l u(w_1) + (1 - p_l) u(w_0)
\]

\[
\Leftrightarrow (p_h - p_l) [u(w_1) - u(w_0)] \geq \psi.
\]

Clearly, the constraint is binding in optimum (make sure you understand this).

**The substituting-in approach to solving the problem.**

We can solve for the optimal wages by substituting in for \( u(w_1) \) and \( u(w_0) \) using the fact that the IR and IC constraints bind in the optimum. The IC constraint tells us that to implement effort \( e = h \) the principal must set wages so that

\[
\psi = p_h u(w_1) + (1 - p_h) u(w_0) - \psi \geq p_l u(w_1) + (1 - p_l) u(w_0)\]

\[
\Leftrightarrow (p_h - p_l) [u(w_1) - u(w_0)] \geq \psi.
\]

Thus,

\[
\psi = \left( \frac{1 - p_l}{p_h - p_l} \right) u(w_0) + \left( \frac{1 - p_h}{p_l - p_h} \right) u(w_1).
\]

Note that the wage received in the high output state under asymmetric information exceeds the fixed wage under symmetric information,

\[
w_{1I}(h) = u^{-1} \left( \psi \frac{1 - p_l}{p_h - p_l} \right) > w^{SI}(h) = u^{-1} (\psi),
\]

while that in the low output state falls below it,

\[
w_{0I}(h) = u^{-1} \left( -\psi \frac{p_l}{p_h - p_l} \right) < w^{SI}(h) = u^{-1} (\psi).
\]

As a consequence of this income risk, the agent needs to be compensated with a risk premium in addition to the reimbursement of the disutility of effort. From the IR constraint we have

\[
\psi = p_h u(w_{1I}(h)) + (1 - p_h) u(w_{0I}(h)).
\]
Since $u'' < 0$ we know that:

$$p_h u(w_1^A(h)) + (1 - p_h) u(w_0^A(h)) < u\left(p_h w_1^A(h) + (1 - p_h) w_0^A(h)\right). \quad (19)$$

Combining (18) and (19), we obtain

$$\psi < u\left(p_h w_1^A(h) + (1 - p_h) w_0^A(h)\right) \iff \frac{u^{-1}(\psi)}{C^{SI}(h)} < \frac{p_h w_1^A(h) + (1 - p_h) w_0^A(h)}{C^A(h)} \quad (20)$$

since $u^{-1}$ is monotone increasing. That is, the expected cost of implementing $e = h$ under asymmetric information, $C^A(h) = p_h w_1^A(h) + (1 - p_h) w_0^A(h)$, exceeds that for implementing the same effort under symmetric information, $C^{SI}(h) = w^{SI}(h) = u^{-1}(\psi)$.

**Implementing effort $e = l$.** Implementing effort level $e = l$ is trivial. Offering the full insurance wages as in the symmetric information setting, $w_0 = w_1 = w^{SI}(l)$, the agent will clearly choose $e = l$ rather than the costly high effort $e = h$. Therefore, the incentive constraint is not binding when implementing low effort.

**Finding the profit maximizing effort level to implement.** Without specifying a functional form for the agent’s utility function we cannot perform this second step and derive the optimal contract (in the problem set we go through such an example).

However, even without performing this final step in detail we can make some general statements about the efficiency of the contractual outcome. The contracting outcome under asymmetric information is called second-best solution because it needs to satisfy the additional incentive constraint and therefore does not always correspond to the first-best solution. As a consequence of this additional constraint, the contracting outcome under asymmetric information may fail to be Pareto optimal, then leading to either inefficient risk sharing or to an inefficiently low effort level.

**Risk sharing.** Since the principal is risk neutral while the agent is risk averse efficient risk allocation calls for full insurance of the agent by the principal. Under asymmetric information full insurance is only feasible if the lowest effort level $e = l$ is to be implemented. Otherwise, the agent needs to be exposed to income risk in order for him to exert high effort $e = h$. This gives rise to a trade-off between risk and incentives: either there can be efficient risk sharing (full insurance of the agent) or there can be appropriate incentives for high effort.

---

5This follows directly from the definition of a strictly concave function.
**Implemented effort level.** As we have seen, implementing effort level \( e = l \) costs the same as under symmetric information. Therefore, whenever the optimal contract under symmetric information calls for \( e = l \) the principal will also choose to implement \( e = l \) under asymmetric information (see Figure 2). Thus, she will neither decide to shut down operations when this is not also the socially optimal thing to do nor will she implement a higher effort level than is socially efficient. However, in contrast to the symmetric information case, the principal needs to expose the agent to some risk to implement \( e = h \) under asymmetric information. As shown above the principal must compensate the agent for this with a risk premium on top of reimbursing the agent for the disutility of effort suffered. As a consequence, the effort level implemented may not correspond to the first-best effort level. Given parameter values for which it is optimal for the principal under symmetric information to implement \( e = h \) (i.e., also socially efficient) it may well be that she finds it optimal to implement effort \( e = l \) under asymmetric information (see Figure 2). The reason is that the cost of implementing \( e = l \) under symmetric and asymmetric information are the same while they differ for implementing effort level \( e = h \). Hence, the expected profit for effort level \( e = h \) may be lower than those for \( e = l \) under asymmetric information even though under symmetric information the principal’s expected profit is maximized with \( e = h \). This is clearly not Pareto-optimal since the principal has lower expected profits than under symmetric information while the agent is held at the same expected utility level of \( \bar{U} \).

**Proposition 2**

The second-best allocation obtained under asymmetric information may fail to be Pareto optimal:

- if high effort \( e = h \) is implemented risk-sharing is inefficient since the risk-averse agent is not fully insured by the risk-neutral principal;

- low effort \( e = l \) might be implemented even though the high effort \( e = h \) is socially optimal.
A For comparison with Lecture 7: The linear programming approach

As we have seen, in the simple two-outcome model it is possible to solve directly for the optimal wages that implement effort \( e = h \) analytically. In contrast, with more than two outcomes this requires specifying numerically the parameter values. However, the linear programming approach allows for fairly detailed characterizations of the properties of the solution to the contracting problem even without specifying numerical parameter values. To allow comparisons the two-outcome case is outlined below.

We can move from the program in \([P:AI]\) to the Lagrangian, noting that the IR and IC constraints bind.

\[
\mathcal{L} = p_h (1 - w_1) + (1 - p_h) (-w_0) - \lambda \left[ \psi - p_h u(w_1) - (1 - p_h) u(w_0) \right] \\
- \mu \left[ \psi - (p_h - p_l) [u(w_1) - u(w_0)] \right],
\]

where \( \lambda \) is the Lagrange multiplier for the IR constraint and \( \mu \) that for the IC constraint. The
first-order conditions are:

\[
\frac{\partial L}{\partial w_1} = -p_h + \lambda p_h u'(w_1) + \mu (p_h - p_l) u'(w_1) = 0, \tag{21}
\]

\[
\frac{\partial L}{\partial w_0} = -(1 - p_h) + \lambda (1 - p_h) u'(w_0) - \mu (p_h - p_l) u'(w_0) = 0, \tag{22}
\]

\[
\frac{\partial L}{\partial \lambda} = p_h u(w_1) + (1 - p_h) u(w_0) = \psi, \tag{23}
\]

\[
\frac{\partial L}{\partial \mu} = (p_h - p_l) [u(w_1) - u(w_0)] = \psi. \tag{24}
\]

Again, it is easy to check that the second-order conditions satisfy the requirements for the solution to be a maximum.

Letting \( p^e_i \) denote the probability that an agent receives wage \( w_i \), \( i \in \{0, 1\} \) given effort \( e \in \{l, h\} \), we can rewrite the conditions (21) and (22):

\[
\text{(21)} \quad \Leftrightarrow \quad -p^h_1 + \lambda p^h_1 u'(w_1) + \mu (p^h_1 - p^l_1) u'(w_1) = 0 \quad \text{and} \quad \frac{1}{u'(w_1)} = \lambda + \mu \left(1 - \frac{p^l_1}{p^h_1}\right),
\]

\[
\text{(22)} \quad \Leftrightarrow \quad -p^h_0 + \lambda p^h_0 u'(w_0) - \mu \left[(1 - p^h_0) - (1 - p^l_0)\right] u'(w_0) = 0 \quad \text{and} \quad \frac{1}{u'(w_0)} = \lambda + \mu \left(1 - \frac{p^l_0}{p^h_0}\right).
\]

That is, in optimum the wage scheme must satisfy

\[
\frac{1}{u'(w_i)} = \lambda + \mu \left(1 - \frac{p^l_i}{p^h_i}\right), \quad \forall i \in \{0, 1\}. \tag{25}
\]

Since the Lagrange multipliers \( \lambda \) and \( \mu \) are positive constants, the wages \( w_i \) are proportional to the Likelihood Ratio \( \frac{p^l_i}{p^h_i} \). The Likelihood Ratio measures how informative output is about the underlying probability distribution \( p^e_i \) (see Section ??). Note, \( \frac{p^l_1}{p^h_1} = \frac{p_l}{p_h} < 1 \) and \( \frac{p^l_0}{p^h_0} = \frac{1-p_l}{1-p_h} > 1 \). Hence,

\[
\frac{1}{u'(w_1)} > \frac{1}{u'(w_0)} \Rightarrow u'(w_0) > u'(w_1)
\]

Thus, \( w_0 < w_1 \) since \( u''(\cdot) < 0 \). Indeed, high (low) output is more (less) likely with high effort than with low effort. Therefore, a high (low) profit is also statistical information indicating that it is likely that the agent exerted high (low) effort and should be rewarded (punished).