Lecture Note 1

Agency Theory, Part I: Pay for Performance

This note considers the simplest possible organization: one boss (or “Principal”) and one worker (or “Agent”). One of the earliest applications of this Principal-Agent model was to sharecropping, where the landowner was the Principal and the tenant farmer the Agent, but in this course we will typically talk in terms of more familiar organization structures. For example, we might consider a firm’s shareholders to be the Principal and the CEO to be the Agent. One can also enrich the model to analyze a chain of command (i.e., a Principal, a Supervisor, and an Agent), or one Principal and many Agents, or other steps towards a full-fledged organization tree.

The central idea behind the Principal-Agent model is that the Principal is too busy to do a given job and so hires the Agent, but being too busy also means that the Principal cannot monitor the Agent perfectly. There are a number of ways that the Principal might then try to motivate the Agent. This note describes the basic Principal-Agent model of incentive contracts (similar to profit sharing or sharecropping); later notes discuss richer and more realistic models. Taken literally and alone, the basic Principal-Agent model may seem too abstract to be useful. But we begin here because this model is an essential building block for many discussions throughout the course—concerning not only managing the incentives of individuals but also managing the incentives of organizational units (such as teams or divisions) and of firms themselves (such as suppliers or partners).

Before turning to the specifics of the Principal-Agent model, we briefly consider the nature and use of economic models more generally.

1. An Introduction to Economic Modeling

We will use several economic models in this course, so it may be helpful to begin by describing what an economic model is and what it can do. We will defer discussion of whether such models are useful until after we have a few under our belts!
An economic model is a simplified description of reality, in which all assumptions are explicit and all assertions are derived. Such a model can produce qualitative and/or quantitative predictions. A qualitative prediction is that “x goes up when y falls.” A quantitative prediction is that $x = \frac{1}{y}$. A model’s (qualitative or quantitative) predictions are useful when they are robust within the environment(s) of interest.

Quantitative predictions often hinge on specific assumptions from the model. If the model will be applied in one particular environment (such as a queuing model for bank tellers, or the Black-Scholes model for option pricing) then the specific assumptions need to match the environment fairly closely, otherwise the quantitative predictions will not be useful in that environment. One might call this “engineering modeling” rather than “economic modeling.”

Qualitative predictions are often more robust, in two senses. First, qualitative predictions may continue to hold if one makes small changes in the model’s specific assumptions. For example, a model’s quantitative predictions might depend on whether a particular probability distribution is normal, exponential, or uniform, but the model’s qualitative predictions might hold for any single-peaked (i.e., hill-shaped) distribution, including the three mentioned above as well as others.

Qualitative predictions can also be robust in a second (and, for our purposes, more important) sense: a simple model’s qualitative predictions may be preserved even if one adds much more richness to the model. The major points we will derive from the economic models in this course are robust predictions in this latter sense. That is, adding greater richness and realism to these models will certainly change the models’ quantitative conclusions, but the major points we derive from the simple models will still be part of the package of qualitative conclusions from the richer models.

2. Elements of the Basic Principal-Agent Model

During this course we will frequently use the term “incentives.” In some settings we will mean a cash payment for a measured outcome, but in other settings our use of this term will be much broader. Lest anyone be misled or disaffected by the narrowness of the former meaning, we will start our discussion of the basic Principal-Agent model by attempting some broader definitions: let “rewards” mean outcomes that people care about (not just dollars), let “effort” mean actions that people won’t take without rewards (not just hours worked), and let “incentives” mean links between rewards and effort (not just
compensation contracts). We will refine these definitions throughout the course. For now we simply note that, according to these definitions, there are clearly lots of incentives out there, even if there are many fewer dollar-denominated incentive contracts.

To be more precise about rewards, effort, and incentives, we turn now to the elements of the basic Principal-Agent model: (a) the technology of production, (b) the set of feasible contracts, and (c) the payoffs to the parties.

A. The Technology of Production

In this simple model, the production process is summarized by just three variables: (1) the Agent’s contribution to firm value (or, loosely, the Agent’s “output”), denoted by $y$; (2) the action the Agent takes to produce output, denoted by $a$; and (3) events in the production process that are beyond the Agent’s control (i.e., “noise”), denoted by $\varepsilon$.

(1) The Agent’s contribution to firm value, $y$: In the sharecropping context, the Agent’s contribution is simply the harvest. In the CEO context, one definition of the Agent’s contribution is the change in the wealth of the shareholders through appreciation in the firm’s stock price. For workers buried inside an organization, it is sometimes very difficult to define and measure a contribution to firm value; in later notes we will discuss alternative performance measures (both objective and subjective).

(2) The action the Agent takes to produce output, $a$: The most straightforward interpretation is that the Agent’s action is effort. This interpretation may be reasonably accurate in the sharecropping context and for low-level workers in large organizations. For a CEO, however, one should think of “effort” not in terms of hours worked but rather in terms of paying attention to shareholders’ interests—does the CEO take actions that increase shareholder value (versus taking actions that indulge pet projects)? As noted above, in the CEO context and elsewhere, we will define “effort” to be actions that the firm values but that the employee will not take without rewards.

(3) Events beyond the Agent’s control, $\varepsilon$: In the sharecropping context, one event beyond the Agent’s control is the weather. In the CEO context, “animal spirits” in the stock market are similarly beyond the Agent’s control.

In the basic Principal-Agent model, these three variables arise in the following order:

1. The Principal and the Agent sign a compensation contract (see below).
2. The Agent chooses an action ($a$), but the Principal cannot observe this choice.
3. Events beyond the Agent’s control (ε) occur.

4. Together, the action and the noise determine the Agent’s output (y).

5. Output is observed by the Principal and the Agent (and by a Court, if necessary).

6. The Agent receives the compensation specified by the contract, as a function of the realized value of y.

To keep the exposition simple, we will make two very specific assumptions.

ASSUMPTION 1: The production function is y = a + ε.

(Some readers may wonder about the units in this production function: how can hours of effort plus inches of rain equal bushels of corn? As an antidote to this concern, consider the slightly more general production function y = g a + h ε, where g is the number of bushels of corn produced per hour of effort and h similarly translates inches of rain into bushels of corn. We have simply set g = h = 1 for notational convenience.)

ASSUMPTION 2: The noise term ε has a normal distribution with mean zero and variance σ²; a higher value of σ² indicates that there is more risk in the production process.

These two assumptions (and others we impose below) lead to some quantitative predictions that are not robust. One would get different answers, for example, if the production function were y = a ε or if the distribution of ε were exponential. The point of making these assumptions, however, is that they greatly facilitate our derivation of several qualitative predictions that are robust.

B. Contracts

We will focus on contracts in which the Agent’s total compensation for the period of the contract, denoted by w, is a linear function of output:

\[ w(y) = s + b y. \]

In such a contract, s can be thought of as salary and b as the Agent’s bonus rate (so that the Agent’s bonus is b y). We sometimes call w the Agent’s “wage,” but this should be understood to mean total compensation, not an hourly wage.

Linear contracts are simple to analyze, are observed in some real-world settings, and have an appealing property: they create uniform incentives, in the following sense. Think of
output, $y$, as aggregate output over (say) a year, but think of the Agent as taking lots of little actions over the course of the year—such as one per day. A non-linear contract may create unintended or unhelpful incentives over the course of the year, depending on how the Agent has done so far.

As an example of unintended or unhelpful incentives, consider a contract that pays a low wage if output is below a target level and a high wage if output meets or exceeds the target level—such as $50,000 for the year if output is below 1000 widgets, but $100,000 for the year if output meets or exceeds 1000 widgets. In this note, we ignore future considerations such as the Agent’s reputation in the labor market. In this case, once the Agent reaches the target level, he or she will stop working; also, if the end of the year draws near and the Agent is still far from reaching the target level, then he or she will stop working. Alternatively, if the Agent were paid (say) a salary of $s = 50,000$ and a bonus rate of $b = 50$ per widget then the Agent would earn $100,000$ for producing 1000 widgets but would have constant incentives regardless of performance to date: every extra widget earns the Agent $50$.

Note that a stock option creates uniform incentives on the upside, in its linear portion, but potentially unintended or unhelpful incentives if it is underwater (or even nearly so). If the option is severely underwater then there are essentially no incentives, because the Agent’s payoff is constant (at zero). More perversely, when the option is at the money, the Agent’s payoff is convex (flat below and linear above), which creates an incentive for risk-taking behavior. Similar incentives and behavior have been documented in several settings, including high-risk portfolio choices by managers of ostensibly conservative mutual funds (see Chevalier and Ellison, 1997).

C. Payoffs

The Principal receives the Agent’s contribution to firm value, $y$, but has to pay the Agent’s wage, $w$, so the Principal’s profit is the difference between the value received and the wage paid:

$$y - w.$$ 

For simplicity, we assume Principal is risk neutral—that is, the Principal simply wants to maximize the expected payoff, namely $E(y - w)$, where the notation $E(x)$ denotes the expected value of the random variable $x$. 
The Agent receives the wage \( w \) but has to take a costly action (e.g., supply effort) to produce any output. Let \( c(a) \) be the dollar amount necessary to compensate the Agent for taking a particular action, \( a \). Think of \( a = 0 \) as no action at all, so \( c(0) = 0 \). (A bit more precisely, think of the action \( a = 0 \) as the action the Agent would take without any financial inducements, hence \( c(0) = 0 \).)

The Agent’s payoff is the difference between the wage received and the cost of the action taken:

\[
w - c(a) .
\]

In this note (and throughout this course), we will focus on the simple case in which the Agent is risk-neutral—that is, the Agent simply wants to maximize the expected payoff \( E(w) - c(a) \). (Since \( \epsilon \) is the only uncertainty in the model and does not affect the Agent’s cost function, no expectation is necessary in the second term of the Agent’s payoff.) See the Appendix of this note for the alternative case in which the Agent is risk-averse.

We will assume that the Agent’s cost function has the (intuitive) shape shown in Figure 1.

There are really two assumptions being made here: (1) bigger actions are more costly (i.e., the cost function \( c(a) \) is increasing), and (2) a small increase in the Agent’s action is more costly starting from a big action than starting from a small action (i.e., the cost function \( c(a) \) is convex—or, equivalently, the marginal cost of actions is increasing). The latter assumption implies that an extra five hours of work per week is tougher for someone currently working 80 hours a week than for someone currently working 40.
3. A Risk-Neutral Agent’s Response to a Linear Contract

A risk-neutral Agent wants to choose the action that maximizes the expected value of the payoff \( w - c(a) \). Since \( w = s + b \cdot y \) and \( y = a + \epsilon \), the Agent wants to maximize the expected value of \( s + b(a + \epsilon) - c(a) \). Since the only uncertainty in the Agent’s payoff arises from the productivity shock \( \epsilon \), and since the expected value of \( \epsilon \) is zero, the Agent’s problem is

\[
\max_a \quad s + b \cdot a - c(a) .
\]

Starting from an arbitrary action \( a_0 \), the marginal benefit of choosing a slightly higher action is \( b \), and the marginal cost of choosing a slightly higher action is the slope of the cost function at \( a_0 \), denoted \( c'(a_0) \). Thus, if \( c'(a_0) \) is less than \( b \) then it is optimal for the Agent to choose a higher action than \( a_0 \), whereas if \( c'(a_0) \) is greater than \( b \) then it is optimal for the Agent to choose a lower action than \( a_0 \). Therefore, the optimal action for a risk-neutral Agent to choose in response to a contract with slope \( b \) is the action at which the slope of the cost function equals \( b \), as shown in Figure 2.

As an illustration, in the simple case where the cost function is the quadratic \( c(a) = \frac{1}{2} a^2 \) we have that the marginal cost function is \( c'(a) = a \) and so the Agent’s optimal action in response to a contract with slope \( b \) is \( a^*(b) = b \).

Note that in this model the salary \( s \) does not affect the Agent’s optimal action: the slope of the contract completely determines the Agent’s incentives; the salary is just a transfer of wealth that does not affect incentives.
Because the slope of \( c(a) \) increases as the Agent’s action increases, a larger value of \( b \) will increase the Agent’s optimal action, \( a^*(b) \). That is, \textit{steeper slopes create stronger incentives}. This result is intuitive: a profit-sharing plan that gives a worker 50% of the firm’s profit is more likely to get the worker’s attention than a plan paying 1%. But does this imply that the Principal would prefer a contract with a very large value of \( b \)?

Definitely not. First, holding the Agent’s salary constant, steeper slopes create stronger incentives but also give away more of the output to the Agent. For example if \( b = 1 \) then the Principal is sure to lose money (assuming a positive salary) because the Agent gets all the output. But in some settings it may make sense to combine a steep slope with a \textit{negative} salary: the slope does good things for the Agent’s effort, the negative salary is interpreted as a fee the Agent pays the Principal for the privilege of getting the job, and the Agent expects to recover the fee by keeping most of the output. Most franchise contracts combine an up-front fee (\( s < 0 \)) and a steep slope (\( b \) near one) in this way. To pay such a fee up front, the Agent would need plenty of wealth or access to credit. There could also be disagreements about the appropriate size for such a fee: a Principal with a poor production process might try to claim otherwise to justify a high fee.

A second potential problem with a contract with a steep slope is that it imposes a lot of risk on the Agent. This risk doesn’t matter if the Agent is risk-neutral, but would matter if the Agent were risk-averse. Interested readers can find more on the interplay between risk and incentives in the Appendix of this note, but this material is \textit{not} required elsewhere in this course.

4. Lessons and Limits of the Basic Principal-Agent Model

The basic Principal-Agent model is extremely simple. On its own it tells us only a little about managing incentives. Among its lessons are (1) contract shape matters (\textit{e.g.}, linear versus kinked or jumped), (2) steeper slopes create stronger incentives, and (3) steeper slopes are not always better.

Beyond these simple lessons, the basic Principal-Agent model has two main values. First, the model gives us a language for expressing and analyzing abstract concepts such as reward, effort, and incentives in terms of more concrete model elements such as production (\( y = a + \varepsilon \)), contract (\( w = s + by \)), and payoffs (\( w - c(a) \) to the Agent and \( y - w \) to the Principal). Second, and probably more important, the model teaches us much by what it leaves out. One instructive exercise is to compare this stick-figure model to the rich
incentive issues in the case study on Lincoln Electric (Fast and Berg, 1975). And even in its own abstract terms, the model clearly omits important issues. For example, in this model there would be no effort if \( b = 0 \), but in the real world we sometimes see great effort even if there is no direct link between pay and performance. Also, the model gives no hint of the agonies that many firms experience when they attempt to link pay to performance. We address both of these omissions in future notes, starting with the latter in Lecture Note 2, “Getting What You Pay For.”

References


Appendix: Risk and Incentives

A risk-averse person doesn’t like risk: such a person would prefer to receive the expected payoff from a gamble for sure rather than allow the gamble’s outcome to be realized and receive whatever payoff results. Such a person has a concave utility function, U(x). That is, the slope of the utility function decreases as the payoff x increases, as shown in Figure 3 below. To be more precise about the definition and consequences of risk aversion, we introduce three related ideas: EV = the expected dollar value of the payoff from a gamble; EU = the expected utility from the payoff from a gamble; and CE = the certainty equivalent of a gamble—the dollar amount such that one is indifferent between receiving CE for sure versus receiving the realized payoff from the gamble, so U(CE) = EU.

To illustrate these ideas, we consider three examples. The first example is shown in Figure 3: a gamble pays $H or $L with equal probability. The expected value of the payoff is EV = (L + H)/2. The expected utility is EU = [U(H) + U(L)]/2. Finally, the certainty equivalent satisfies U(CE) = EU, as always.

![Figure 3](image)

Saying that a person is risk-averse (i.e., that the person would prefer to receive the payoff EV with certainty rather than face the gamble) is equivalent to saying that U(EV) > EU, as
shown in the figure. We can also state this preference in terms of dollars rather than utility: EV > CE.

Two other examples may help clarify these concepts:

**Example 2:** a gamble pays $100 or $0 with equal probability, and so has an EV of $50.

**Example 3:** the payoff from a gamble has a normal distribution with mean m and variance v, so the gamble has an EV of m.

Given a utility function and a gamble, we can compute the associated expected utility, EU. Suppose U(x) = \( \sqrt{x} \) in Example 2. Then the expected utility is

\[
EU = \frac{1}{2} U(100) + \frac{1}{2} U(0) = \frac{1}{2} \cdot 10 + \frac{1}{2} \cdot 0 = 5.
\]

The certainty equivalent must satisfy U(CE) = EU, or \( \sqrt{CE} = 5 \), so CE = $25. Thus, a person with utility function U(x) = \( \sqrt{x} \) would rather receive, say, $25.01 than receive the realized payoff from the gamble in Example 2, even though the expected value of that payoff is $50. Such a person is quite risk-averse!

Consider the utility function U(x) = -e^{-rx} in conjunction with Example 3. (We will discuss and apply this utility function below. For now, suffice it to say that the parameter r \( \geq 0 \) measures the person’s risk-aversion, where r = 0 corresponds to risk-neutrality and higher values of r indicate greater risk-aversion. Under this utility function, the person has the same level of risk-aversion, described by r, regardless of the size of the payoff, x.) With a little calculus (available on request), one can show that

\[
EU = -e^{-r[m - \frac{1}{2} rv]}.
\]

To compute the certainty equivalent of the gamble in Example 3 given the utility function U(x) = -e^{-rx}, we must find the dollar value CE that satisfies U(CE) = EU, or

\[
-e^{-r} CE = -e^{-r[m - \frac{1}{2} rv]},
\]

Recall that a single utility value means nothing: utilities are only meaningful in comparisons, such as U(A) > U(B), which means that the individual prefers A to B. As we will see below, the fact that we use utility values only in comparisons means that we have no problem with negative utility values. In the example above, for instance, U(A) might equal -5 and U(B) might equal -10.
so \( CE = m - \frac{1}{2} rv \). Thus, although the expected value of the payoff from this gamble is \( m \), a person with this utility function would be indifferent between receiving \( m - \frac{1}{2} rv \) for sure and receiving the realized payoff from the gamble. The difference between EV and CE (namely, \( EV - CE = \frac{1}{2} rv \)) is called the *risk premium* for the gamble—the cost (for someone with this utility function) of bearing the gamble’s risk. If someone is more risk averse (*i.e.*, if \( r \) is higher) or if the gamble involves more risk (*i.e.*, if \( v \) is higher) then this risk-bearing cost is higher.

Given this brief introduction to risk-aversion, we can now return to the basic Principal-Agent model described above, this time under the assumption that the Agent is risk-averse. In particular, suppose the Agent has the utility function \( U(x) = -e^{-rx} \). In this case, a single value of the contract slope \( b \) is *efficient*. That is, if the parties were to sign a contract with a slope other than this efficient value then they could both be made better off by switching to a contract with the efficient slope (and a new value of \( s \) chosen to make both parties better off after the switch).

As an illustration, suppose that the Agent’s cost function is \( c(a) = \frac{1}{2} a^2 \). Then one can show (see Gibbons (2001) for details) that the efficient slope is

\[
b^* = \frac{1}{1 + r\sigma^2}.
\]

Two points are worth noting about \( b^* \). First, \( b^* \) responds to changes in \( r \) or \( \sigma^2 \) in sensible ways: if the Agent is very risk averse (*i.e.*, \( r \) is large) then \( b^* \) is small; likewise, if there is lots of risk in the production process (*i.e.*, \( \sigma^2 \) is large) then \( b^* \) is again small. Put more colloquially, \( b^* \) is determined in part by the Agent’s need for insurance (which is in turn determined by both risk aversion and noise).

Second, \( b^* \) is between zero and one. Setting \( b = 0 \) would perfectly insure the Agent but would provide no incentives. Setting \( b = 1 \) effectively sell the firm to the Agent but would provide no insurance. The fact that \( b^* \) is between zero and one reflects the trade-off between providing the Agent with incentives and forcing the Agent to bear risk.