Notes on Contract Theory

Daniel Barron
Northwestern University

Fall 2019

September 3, 2019

Why Contract Theory?

Reference:

- Coase, Ronald H. 1960. "The Problem of Social Cost." *The Journal of Law and Economics* 3, p 1-69.

Contract theory studies what happens when contracts are imperfect. We start with Coase (1960), which argues that parties will bargain to a Pareto efficient outcome when property rights as well-defined. To illustrate, suppose that party A can take an action that benefits herself and harms party B, and suppose that B has the ability to stop A from taking that action. Then A can induce B to allow the action by compensating him for the harm it causes. Party A is willing to pay this amount if, and only if, the benefits to A outweigh the harms to B. So Party A will take the action if, and only if, it is socially efficient: decentralized bargaining leads to (Pareto) efficient outcomes.

To be compelling, this argument requires a set of assumptions that do not hold in many applications. Contract theory has focused on two of these assumptions. First, the relevant action must be **contractible:** party A must be able to commit *not* to take the action following the agreement, or equivalently, B must be able to commit to prevent A from taking that action exactly when their agreement says so. Second, the parties' payoffs must be **common knowledge:** A and B must agree on the costs and benefits of an action in order to agree on a price that exceeds the cost but is less than the benefit.

Break the first assumption, and you have **moral hazard**: at least one party cannot commit to "private" actions and so cannot be prevented from taking their preferred action. Moral hazard arises wherever workers exert private effort to do their jobs, executives make strategic decisions that affect profits, researchers pursue innovative new products, and suppliers design inputs for their downstream partners. Break the second assumption, and you have **private information** (or **adverse selection**): parties can strategically misrepresent their payoffs, and in so doing, either prevent an agreement or distort its terms. Adverse selection arises wherever sellers knows more about their products than customers, patients know more about their health status than insurers, or workers know more about their ability than employers.

Once you break the assumptions the underpin the Coase theorem, outcomes are not necessarily Pareto efficient and new economic forces emerge. Contract theory is the study of those forces, their welfare effects, and how new tools and regulations might be used to improve economic outcomes.

What is the Goal of These Notes?

These notes have two goals. First, the field of contract theory has built on a set of relatively unified foundational models. These notes focus on these models, which were mostly written in the 1980s and 1990s. While the papers themselves are a few decades old, the lessons we learn from them, and their implications, still drive much of modern-day contract theory. These models are also an integral part of many other fields in economics that study contracting frictions, from macroeconomics, to industrial organization, to labor and personnel economics, to organizational economics.

The second goal of these notes is to discuss more recent topics in contract theory. Each chapter concludes with a "further reading" section that highlights recent papers that build on the basic ideas presented in the chapter. These sections are a useful way to understand how each topic has evolved since the fundamental papers covered in the rest of the chapter.

Most sections of the notes build on a specific paper or set of papers. These references are cited at the beginning of the section.

Useful References

These notes owe a deep debt to the lecture notes written by Richard Holden, which are available online on Richard's website: http://research.economics.unsw.edu.au/richardholden/assets/2060-rh-2016-v2.pdf. The topics covered here overlap substantially with Richard's notes, though the ordering and the approach differ. My hope is that these notes are complementary.

If you are looking for additional references for further study, there are several good textbooks that are the standards in the field:

- Bolton, Patrick, and Mathias Dewatripont. 2005. Contract Theory. MIT Press: Cambridge, MA.
- Stole, Lars. 1996. "Lectures on the Theory of Contracts and Organizations." Available online: https://pdfs.semanticscholar.org/5447/570daf54adebe883c4167367d947acec0199.pdf

Part I

Static Moral Hazard

Chapter 1

Insurance versus Incentives

1.1 Introduction

This chapter covers the classic moral hazard model. In it, a risk-neutral principal motivates a risk-averse agent to exert costly effort. Effort is not contractible, but a noisy signal of effort is. The principal can therefore commit to reward the agent as a function of this signal to motivate him to exert effort. But doing so exposes the agent to risk, leading to an outcome that is less efficient than what could be achieved if the principal could reward the agent based on effort.

This observation – that incentive pay based on a noisy signal exposes the agent to risk – means that there is a trade-off between motivating and insuring the agent, which is sometimes called the trade-off between insurance and incentives.

1.2 The Model

1.2.1 A Useful Framework

We begin by writing a general form of the contracting problem with a risk-averse agent, which we quickly specialize to a more familiar problem. We start with the general formulation for two reasons. First, the economics of the problem are actually clearer (though less tractable) in the more general model. Second, the general model is relatively easy to modify to study other types of incentives, a point that we will illustrate throughout the class.

Consider a **principal** ("she") who interacts with an **agent** ("he"). They play the following game:

1. The principal publicly offers a **contract** $s : \mathcal{Z} \to \mathbb{R}$.

2. The agent publicly **accepts or rejects** this contract. Rejection ends the game.

3. If the agent accepts the contract, he privately chooses an **action** $(F, c) \in \mathcal{A}$, where $F \in \Delta(\mathcal{Z})$ is a probability distribution and $c \in \mathbb{R}_+$ is a cost.

4. **Signal** $z \in \mathcal{Z}$ is realized according to F.

If the agent accepts the contract, then the principal's and agent's payoffs are $\pi((F, c), s(z))$ and $u(s(z)) - c$, respectively. If the agent rejects the contract, then the principal earns $\bar{\pi} \in \mathbb{R}$ and the agent earns $\bar{u} \in \mathbb{R}$. We typically take $u(\cdot)$ to be weakly concave, so that the agent is risk-averse, and consider settings in which the principal strictly prefers retaining the agent to her outside option. We consider subgame-perfect equilibria of this game.

The principal commits to the contract $s(\cdot)$, which pays the agent as a function of the signal. Given this contract, the agent takes two actions: he chooses whether or not to accept the contract, and conditional on accepting, he chooses an action that determines the principal's payoff and the distribution over z. The central contracting friction is that the agent's action is **non-contractible**: the contract cannot condition on (F, c), only on the signal z. In general, we say that a variable is **contractible** (or **verifiable**) if a formal contract can condition on it, and **non-contractible** (or **non-verifiable**) otherwise.

In practice, the agent's action might be very complicated, and could entail the agent making decisions, exerting effort, allocating mental resources, and so on. One advantage of this formulation is that it reduces that complicated action to two economically relevant features: the cost to the agent, c, and the resulting output distribution, $F(\cdot)$. By abstracting from the details of the action, the model highlights the important assumption: what matters is how the feasible distributions over contractible signals, z, relates to the benefit to the principal, $\pi(\cdot)$. In some settings, z is noisy but otherwise accurately reflects the principal's preferences. In other settings, such as multitasking models, z is misaligned with the principal's preferences.

1.2.2 The Optimal Program

We now specialize this problem to a more familiar form. Define the **incentives-insurance problem** as a special case in which:

1. $\mathcal{Z} \equiv \mathbb{R} \times \mathcal{X}$,with representative element (y, x),

2. $\pi((F, c), s) \equiv \mathbb{E}_F[y] - s$, and

3. the set of feasible (F, c) are parameterized by an **effort,** $a \in \mathbb{R}_+$. An agent who chooses a incurs the **cost** $c(a)$ and generates outcomes according to conditional distribution $F(\cdot|a)$.

We assume that $u(\cdot)$ is twice continuously differentiable, concave, and has range $\mathbb{R}$, while $c(\cdot)$ is twice continuously differentiable and convex. We also assume that $F(\cdot|a)$ takes values in an interval, has full support on that interval, and is at least twice continuously differentiable with density $f(\cdot|a)$.

Suppose that the agent participates in the optimal contract (rather than taking his outside option). Then the principal's optimal contract solves the following constrained maximization problem:

$$\max_{a, s(\cdot)} \mathbb{E}\left[y - s(y, x)|a\right] \tag{Obj}$$

subject to the **incentive compatibility** constraint

$$a \in \arg\max_{\tilde{a}} \left\{\mathbb{E}\left[u(s(y, x)) - c(\tilde{a})|\tilde{a}\right]\right\} \tag{IC}$$

and the **individual rationality** constraint

$$\mathbb{E}\left[u(s(y, x)) - c(a)|a\right] \geq \bar{u}. \tag{IR}$$

In this problem, the principal chooses both the contract $s(\cdot)$ and the desired effort a.[1] The contract must

[1] Implicitly, we are also allowing her to choose the agent's participation decision.

give the agent the incentive to actually choose a, which is guaranteed by (IC). Anticipating this effort, the agent must be willing to participate in the contract, which is guaranteed by (IR).

The advantage of formulating the optimal contract as the solution to a constrained maximization problem is that we can essentially ignore the initial timing of the model. We now have a well-defined problem to solve, and one that highlights how optimal incentives depend on the economic environment. This lesson is more general: when you are studying a contracting problem, always write down the constrained maximization problem. It will simplify your life and help you focus on the core economics of the setting.

1.2.3 First-Best

As a benchmark, we first consider what would happen if a itself were contractible. We refer to this case as the **first best**, in contrast the contracting problem with non-contractible a, which is the **second best.**

If a is contractible, then the principal can write a "forcing contract" that requires the agent to pay a large fine unless he chooses effort a. Such a contract guarantees that the agent is willing to choose a. We can therefore ignore (IC) in the first-best contracting program, which reduces to choosing a single payment s that is paid if, and only if, effort equals a.

The principal chooses s and a to solve

$$\max_{a,s} \mathbb{E}[y|a] - s$$

subject to (IR),

$$u(s) - c(a) \geq \bar{u}.$$

Any solution will set $u(s) = \bar{u} + c(a)$, which means that a is chosen to maximize the principal's payoff, $\mathbb{E}[y|a] - u^{-1}(\bar{u} + c(a))$. Assuming that $\mathbb{E}[y|a]$ is concave in a (and ignoring corner solutions, which can be ruled out by assuming $c'(0) = 0$ and $\frac{d}{da}\mathbb{E}[y|a] > 0$), the solution is defined by the first-order condition

$$\frac{d}{da}\mathbb{E}[y|a] - \frac{du^{-1}}{dv}|_{\bar{u}+c(a)} c'(a) = 0.$$

Define **first-best effort,** a^{FB}, as the effort that maximizes this payoff.

The key observation is that the insurance-incentives tradeoff disappears when a is contractible. The reason is that motivating the agent no longer requires exposing him to any risk. The optimal contract therefore conditions only on a, perfectly insuring the agent against risks while still motivating him to exert effort.

1.2.4 Risk-Neutral Agents and "Selling the Firm"

A risk-neutral agent can be given high-powered incentives, since he does not need to be insured. Consequently, even if a is non-contractible, the optimal contract with a risk-neutral agent attains first-best.

Suppose that $u(s) = s$. Then the first-best effort level satisfies

$$\frac{d}{da}\mathbb{E}[y|a^{FB}] - c'(a^{FB}) = 0.$$

That is, first-best maximizes the principal's benefit from higher output minus the agent's cost of choosing that effort.

Now, consider the contract that sets $s(y) = w + y$, where w is chosen so that (IR) binds if the agent chooses $a = a^{FB}$, $w = \bar{u} + c(a^{FB}) - \mathbb{E}[y|a^{FB}]$. This contract induces the agent to choose $a = a^{FB}$, since his payoff equals

$$\mathbb{E}[s(y)|a] - c(a) = \mathbb{E}[w + y|a] - c(a)$$

and w is a constant. Moreover, this contract gives the principal a payoff equal to $-w = \mathbb{E}[y|a^{FB}] - c(a^{FB}) - \bar{u}$, which equals her first-best payoff. Thus, while there is moral hazard in this setting, the moral hazard does not lead to inefficient outcomes.

The contract

$$s(y) = w + y$$

essentially **sells the firm** (or rather, sells the output of the firm) to the agent, whose effort incentives are therefore aligned with the social value of effort. We say that such a contract makes the agent the **residual claimant** on his effort: his marginal returns to effort equal the social marginal returns to effort.

1.3 The First-Order Approach

Reference:

- Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics* 10 (1), 74-91.

Now, we return to the case with a risk-averse agent. How might we solve the problem of maximizing (Obj) subject to (IC) and (IR)? One intuitively appealing approach is to replace (IC) with its first-order condition,

$$c'(a) = \frac{\partial}{\partial a}\mathbb{E}\left[u(s(y,x))|a\right]. \tag{IC-FOC}$$

By considering only local deviations, (IC-FOC) simplifies (IC) and relaxes the problem considerably. As we discuss later in this chapter, this replacement, which is typically called the **first-order approach**, should be used with caution.

The Lagrangian and the Modified Borch Rule

Suppose that $x \in \mathbb{R}$ and that for any effort a, the conditional outcome distribution $F(y,x|a)$ has a differentiable density $f(y,x|a)$ that is strictly positive on a common support. If we further assume that we can replace (IC) with (IC-FOC), then we can write the Langrangian for this relaxed problem:

$$\mathcal{L} = \int (y - s(y,x))\, f(y,x|a)dydx - \lambda\left(\bar{u} - \left(\int u(s(y,x)) f(y,x|a)dydx - c(a)\right)\right) - \mu\left(c'(a) - \int_y u(s(y,x)) f_a(y,x|a)dydx\right), \tag{1.1}$$

where λ and μ are the multipliers associated with (IR) and (IC-FOC), respectively.

Now, fix a and note that we can take the derivative of $\mathcal{L}$ with respect to $s(y,x)$ for *each* outcome (y,x). Doing so yields the first-order condition

$$-f(y,x|a) + \lambda u'(s(y,x)) f(y,x|a) + \mu u'(s(y,x)) f_a(y,x|a) = 0,$$

The agent's utility is strictly increasing in her pay, so we can divide by $u'(s(y,x))f(y,x|a)$ to yield

$$\frac{1}{u'(s(y,x))} = \lambda + \mu \frac{f_a(y,x|a)}{f(y,x|a)}. \tag{BORCH}$$

The expression (BORCH), which is sometimes called the **modified Borch rule**, gives the optimal contract $s(y,x)$ as a function of the agent's marginal utility of money ($u'(\cdot)$), the likelihood ratio on output at (f_a/f), and the multipliers associated with (IR) and (IC-FOC). Of course, we derived this expression under the big assumption that the first-order approach is valid. Nevertheless, (BORCH) is informative.

We can interpret $1/u'$ as the *cost to the principal of giving the agent a slightly higher utility,* which is increasing because $u(\cdot)$ is concave. Thus, (BORCH) tells us that the principal should pay the agent more when f_a/f is larger. Intuitively, a large f_a/f implies that (y,x) is *strong evidence that the agent choose effort no less than a.* Note that the agent's pay depends on the value of (y,x) as a signal of effort, *not* (necessarily) on the principal's benefit from (x,y). In particular, if f_a/f is decreasing in y over some range, then so is $s(y,x)$.

First-best

Note that we can recover the first-best optimal contract from (BORCH). If $\mu = 0$, then $u'(s(y,x)) = 1/\lambda$, and so $s(\cdot)$ is constant. In the absence of moral hazard, the principal offers perfect insurance to the agent, which makes sense because the principal is risk-neutral and the agent is risk-averse. In the presence of moral hazard, therefore, the optimal contract insures the agent as much as possible conditional on motivating her to choose the desired level of effort.

Monotonicity of $s(\cdot)$

For the moment, suppose that x is constant and independent of a, so that we can ignore it in the optimal contract. (With abuse of notation, we will suppress x in the discussion below.)

The contract that solves (BORCH) is increasing if $F(\cdot|a)$ satisfies the **Marginal Likelihood Ratio Property (MLRP).**

Definition 1 *Given a twice continuously differentiable distribution $F(\cdot)$, we say that $F(\cdot)$ satisfies **MLRP in** a if for all a,*

$$\frac{f_a(y|a)}{f(y|a)}$$

is (weakly) increasing in y.

MLRP says that the derivative of the log-density function with respect to a,

$$\frac{\partial}{\partial a} \ln f(y|a) = \frac{f_a(y|a)}{f(y|a)},$$

is increasing in a. This condition has a natural interpretation in the context of moral hazard problems: it says that for any outcomes $y' > y$, increasing effort increases the log density at y' more than at y. Roughly, more effort makes low outcomes less likely and high outcomes more likely. MLRP rules out situations in which, for example, effort increases the probability of "intermediate" outcomes but decreases the probability of exceptionally high outcomes.

Remark 1 *MLRP implies that $F(\cdot|a)$ is FOSD increasing in a. To see this, suppose that F has support on $[\underline{y},\bar{y}] \subseteq \mathbb{R}$ for all a. Then $F_a(\bar{y}|a) = 0$, because $F(\bar{y}|a) \equiv 1$ for* any *a. For any $y \in [\underline{y},\bar{y}]$*

$$F_a(y|a) = \int_{\underline{y}}^{y} f_a(z|a)dz = \int_{\underline{y}}^{y} \frac{f_a(z|a)}{f(z|a)} f(z|a)dz \leq 0.$$

The key step in this argument is the inequality. To see this step, note that

$$\int_{\underline{y}}^{y} \frac{f_a(z|a)}{f(z|a)} f(z|a)dz = \mathbb{E}_z\left[\frac{f_a(z|a)}{f(z|a)}|z \leq y\right] F(y)$$

and that

$$0 = F_a(\bar{y}|a) = \mathbb{E}_z\left[\frac{f_a(z|a)}{f(z|a)}|z \leq y\right] F(y) + \mathbb{E}_z\left[\frac{f_a(z|a)}{f(z|a)}|z > y\right](1 - F(y)).$$

Since f_a/f is increasing by MLRP,

$$\mathbb{E}_z\left[\frac{f_a(z|a)}{f(z|a)}|z \leq y\right] \leq \mathbb{E}_z\left[\frac{f_a(z|a)}{f(z|a)}|z > y\right].$$

Since $F(y) \geq 0$, f_a/f can have an expected value of 0 only if

$$\mathbb{E}_z\left[\frac{f_a(z|a)}{f(z|a)}|z \leq y\right] F(y) \leq 0.$$

1.4 The Informativeness Principle and its Implications

1.4.1 Which Variables are Included in an Optimal Contract?

While the principal does not directly care about the signal x, it might be informative about a. When should the optimal contract vary in x? Put another way, what signals should the optimal incentive contract include?

To address this question, we return to (BORCH), the first-order condition characterizing the optimal contract:

$$\frac{1}{u'(s(y,x))} = \lambda + \mu \frac{f_a(y,x|a)}{f(y,x|a)}.$$

As we noted previously, $u'(\cdot)$ is strictly decreasing. Therefore, $s(y,x)$ is (almost always) constant in x if and only if $\frac{f_a(y,x|a)}{f(y,x|a)}$ is (almost always) constant in x. We can interpret this condition using the language of sufficient statistics.

Definition 2 *Given a distribution $F(y,x,a)$, we say that y is a* ***sufficient statistic*** *for x if there exists functions $g(y,x)$ and $h(y,a)$ such that for all (y,x),*

$$f(y,x|a) = g(y,x)h(y,a).$$

Now, we discuss the **Informativeness Principle,** which gives necessary and sufficient conditions for the optimal contract to condition on a signal x.

Proposition 1 *Assume that the first-order approach is valid and that $\mu > 0$. Then the optimal $s(\cdot)$ is constant in x for all a if and only if y is a sufficient statistic for x for all a.*

Proof

We have already argued that $s(\cdot)$ is constant in x if and only if there exists a function $H(y,a)$ such that

$$\frac{f_a(y,x|a)}{f(y,x|a)} = H(y,a).$$

Noting that $\frac{f_a(y,x|a)}{f(y,x|a)} = \frac{\partial}{\partial a}\ln f(y,x|a)$, we can integrate both sides of this equality with respect to a to yield

$$\ln f(y,x|a) = \int_{\tilde{a}\leq a} H(y,\tilde{a})d\tilde{a} + G(y,x).$$

But then

$$f(y,x|a) = \exp\left\{\int_{\tilde{a}\leq a} H(y,\tilde{a})d\tilde{a} + G(y,x)\right\} = g(y,x)h(y,a),$$

where $g(y,x)\equiv\exp\{G(y,x)\}$ and $h(y,a)\equiv\exp\left\{\int_{\tilde{a}\leq a} H(y,\tilde{a})d\tilde{a}\right\}$. We conclude that y is a sufficient statistic for x.

Now, suppose y is a sufficient statistic for x for all a. There exists functions $h(y,a)$ and $g(y,x)$ such that for all (x,y,a),

$$f(y,x|a) = g(y,x)h(y,a).$$

For any sharing rule $s(y,x)$, define the alternative sharing rule $\tilde{s}(y)$ so that

$$u(\tilde{s}(y)) = \frac{\int_x u\left(s(y,x)\right)g(y,x)dx}{\int_x g(y,x)dx}.$$

Then for any a,

$$\begin{aligned}
\int_{y,x} u(s(y,x))f(y,x|a)dxdy &= \int_y h(y,a)\int_x u(s(y,x))g(y,x)dxdy \\
&= \int_y h(y,a)u(\tilde{s}(y))\int_x g(y,x)dxdy \\
&= \int_{y,x} u(\tilde{s}(y))f(y,x|a)dxdy.
\end{aligned}$$

The agent therefore earns the same expected utility for each effort; $s(y,x)$ satisfies (IC) and (IR), so $\tilde{s}(y)$ does too. Moreover, since $u(\cdot)$ is concave,

$$\int s(y,x)f(y,x|a)dxdy \leq \int \tilde{s}(y)f(y,x|a)dxdy$$

by Jensen's Inequality. ■

The Informativeness Principle says that a variable is included in the optimal contract if and only if the other variables are not a sufficient statistic for it. If a variable is not informative, then adding it to the contract simply adds noise, which is bad because the agent is risk-averse. If a variable is informative, on the other hand, the contract should vary at least a little bit. Such variables should be included *even if they are very noisy.* If $s(\cdot)$ is independent of x, and y is not a sufficient statistic for x, then we can perturb $s(\cdot)$ slightly so that it varies in x. Such a perturbation has a first-order effect on incentives and a second-order effect on the agent's exposure to risk.

One implication of the informativeness principle is that contracts should incorporate lots and lots of contingencies. The agent's pay should depend, at least a little bit, on anything that is even *slightly* (independently) informative of his effort. Of course, many real-world contracts are quite complicated and contain many terms. However, even those contracts are unlikely to include tiny payments associated with very noisy, not very informative outcomes. Therefore, one takeaway from the Informativeness Principle is that this contracting model might be missing a key friction that is present in the real world. We will return to this point when we discuss multitasking (Chapter 2) and gaming (Chapter 5).

1.4.2 Implication: Relative Performance Evaluation

Reference:

- Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *The Bell Journal of Economics* 13 (2), 324-340.

The Idea

What kinds of informative signals might the principal use in an optimal contact? One possibility – widespread in practice – is that workers are compensated on the basis of their performance *relative to* the performance of their peers. Relative performance evaluation is a feature of explicit monetary incentives, as with "employee of the month" bonuses, rank-order tournaments, and forced curves. Agents are also compared to each other for all kinds of other rewards, including promotions, task assignments, jobs, and so on.

The Informativeness Principle suggests one advantage to relative performance schemes: they filter *common shocks* that affect all agents' productivities. That is, suppose that the principal is actually contracting with *two* agents (one of whom will remain unmodeled). There is a state of the world that affects both agents: if the state is good, then effort leads to output, but if the state is bad, then it does not. To induce effort, an incentive scheme must create a link between effort and expected pay.

If each agent is rewarded or punished separately, then the agents are unavoidably exposed to the risk that they will be punished because the state is bad. On the other hand, a relative performance scheme can filter this common shock, for instance, by punishing an agent only when she does not produce output but the other agent does. Such a scheme limits risk and so allows the principal to provide higher-powered incentives without exposing the agent to more risk.

Linking to the Informativeness Principle

Formally, suppose that the state of the world is (θ_1, θ_2), jointly distributed according to some (twice continuously differentiable) distribution $H(\cdot)$ with density $h(\cdot)$. The agent's output depends on θ_1, $y = y(a, \theta_1)$, and the signal depends on θ_2, $x = x(\theta_2)$. Assume y and x are strictly increasing in θ_1 and θ_2, respectively.

Define $y^{-1}(z, a)$ so that $y(a, y^{-1}(z, a)) = z$. That is, y^{-1} gives the θ that results in each output z given a. Similarly, let $x^{-1}(x)$ satisfy $x^{-1}(x(\theta_2)) = \theta_2$. Both of these inverses are well-defined because y and x are strictly increasing in θ_1 and θ_2. Given these definitions, the distribution over (y, x) can be written

$$H(y^{-1}(y, a), x^{-1}(x)).$$

If there exist $H_1(\cdot)$ and $H_2(\cdot)$ such that $H(\theta_1, \theta_2) = H_1(\theta_1)H_2(\theta_2)$, then

$$F(y, x|a) = H(y^{-1}(y, a), x^{-1}(x)) = H_1(y^{-1}(y, a))H_2(x^{-1}(x)),$$

which is exactly the condition for x to not be informative. In that case, the Informativeness Principle says that x is *not* included in the optimal contract. Conversely, if no such $H_1(\cdot)$ and $H_2(\cdot)$ exist, then loosely, we cannot split apart $F(y, x|a)$ and so the optimal contract will condition on x. This result can be shown more fully by using the formula above to show that f_a/f varies in x.

What are the Practical Problems with RPE?

Let's step back from this model for a moment and think more broadly about the potential downsides of relative performance evaluation. Why might an organization not want to condition one agent's pay on another agent's performance?

- **Collusion:** the argument above implicitly assumes that agents can deviate only unilaterally, so that the *other* agent's output distribution is essentially exogenous from the perspective of *this* agent's effort incentive. In practice, however, agents might be able to collude with one another by jointly shirking and so evading punishment.

- **Sabotage:** under relative performance evaluation, an agent can increase her payoff by either increasing her own output or decreasing the output of her peer. This creates an incentive for agents to sabotage one another, or more mildly, to not help one another out.

- **The agents' interactions are close to "zero sum:"** this very nice point is made by Che and Yoo (2001) and follows from the sabotage incentives described above. By making one agent's gain the other agent's loss, relative performance evaluation ensures that agents have opposed interests. If we think that those agents are interacting repeatedly, that repeated interaction therefore resembles a repeated zero-sum game. In zero-sum games, repetition does not expand the set of equilibria. In contrast, if agents are given *joint* performance evaluation, so that each agent's payoff is *increasing* in the other agent's performance, then their repeated interaction looks like a Prisoner's Dilemma in which "cooperate, cooperate" corresponds to both agents exerting a lot of effort. Therefore, relative performance evaluation limits agents' incentive to motivate on another using informal relationships.
  - **Reference:** Che, Yeon-Koo, and Seung-Weon Yoo. 2001. "Optimal incentives for teams." *American Economic Review* 91 (3), 525-541.

1.5 When Does the First-Order Approach Hold?

In this section, we will ignore the signal x, so that output y is the only outcome.

1.5.1 A Counterexample

Reference:

- Mirrlees, J. A. 1999 (written 1975). "The Theory of Moral Hazard and Unobservable Behaviour: Part 1." *Review of Economic Studies* 66, 3-21.

The first-order approach assumes that we can replace (IC) with (IC-FOC). This is a substantial assumption. In particular, it requires that, e.g., (IC) is a concave maximization problem, which requires that

$$\int_y (u(s(y))f_{aa}(y|\tilde{a})dy) \leq c''(\tilde{a})$$

for *any* $\tilde{a}$. But this condition depends on the optimal contract $s(\cdot)$, which might in principle have a complicated shape. It is therefore a tall order for the agent's payoff to be concave for *any* contract.

In addition to this problem, Mirrlees (1975, published 1999) identified an even more substantial problem: in some natural settings, *no optimal contract exists.* In those cases, (IC-FOC) will be misleading even as a necessary condition.

Uniform example

To illustrate this point, we start with a somewhat contrived example. Suppose that $y \sim U[a, a+1]$, so that output is uniformly distributed with a support that depends on effort. We will show that there exists a contract that both perfectly insures *and* motivates the agent. For example, consider a contract $s(y)$ that gives the agent utility

$$u(s(y)) = \begin{cases} -K & y < a \\ \bar{u} & y \geq a \end{cases},$$

where K is a very large number. Such a contract is feasible if $u(\cdot)$ is unbounded from below. If the agent chooses effort a, this contract pays a constant amount that gives the agent exactly his outside option, $\bar{u}$. If the agent chooses any lower effort, however, then he faces the risk of earning the extremely bad payoff $-K$. We can choose $-K$ so that the agent would rather incur the effort cost a than deviate downward and face this risk. (Note that a marginal deviation downward decreases the agent's cost by $c'(a)$ and increases the probability of the penalty at rate 1.)

Normal example

In the uniform example, the outcome distribution has a shifting support, which means that a shirking agent is perfectly detected with positive probability. To attain first-best, however, we do not need shirking to be literally perfectly detectable; it is enough that the signal are occasionally arbitrarily informative about effort. To see this, take the classical example with normally-distributed output $y \sim N(a, \sigma^2)$. Recall that the density function of the normal distribution is

$$f(y|a) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-a)^2}{2\sigma^2}\right\},$$

so that

$$\frac{f_a(y|a)}{f(y|a)} = \frac{y-a}{\sigma^2}.$$

The normal distribution becomes *approximately perfectly informative in the tails*, in the sense that the likelihood ratio diverges to $-\infty$ as $y \downarrow -\infty$ (and to ∞ as $y \to \infty$). Now, consider the contract

$$u(s(y)) = \begin{cases} -K & y < \underline{y} \\ \bar{u} + \epsilon & y \geq \underline{y} \end{cases}.$$

We can take $\underline{y} \to -\infty$, $\epsilon \to 0$, and $K \to \infty$ "at the right rates," in order to expose the agent to arbitrarily little risk while still penalizing shirking arbitrarily harshly.

To make this point more precisely, fix $a > 0$ and consider the principal's optimal contract within the class described above:

$$\min \int_{-\infty}^{\underline{y}} u^{-1}(-K) f(y|a) dy + \int_{\underline{y}}^{\infty} u^{-1}(\bar{u}+\epsilon) f(y|a) dy$$

subject to the (IC-FOC) and (IR) constraints:

$$\int_{-\infty}^{\underline{y}} (-K) f_a(y|a) dy + \int_{\underline{y}}^{\infty} (\bar{u}+\epsilon) f_a(y|a) dy = c'(a)$$

and

$$\int_{-\infty}^{\underline{y}} (-K) f(y|a) dy + \int_{\underline{y}}^{\infty} (\bar{u}+\epsilon) f(y|a) dy \geq \bar{u}.$$

For any $\underline{y}$, we can choose K sufficiently large that (IC-FOC) holds. We can then make (IR) bind by setting $\epsilon > 0$ so that

$$\epsilon = \frac{F(\underline{y})}{1-F(\underline{y})}\left(K+\bar{u}\right).$$

Because $\lim_{y \to -\infty} \frac{f_a(y)}{f(y)} = -\infty$, for any $M > 0$, we can find an $\underline{y}$ such that $\frac{f_a(\underline{y})}{f(\underline{y})} = -M$. But since $-K < 0$, (IC-FOC) can be made to hold so long as

$$\int_{-\infty}^{\underline{y}} (-K)(-M) f(y|a) dy + \int_{\underline{y}}^{\infty} (\bar{u}+\epsilon) \frac{f_a(y|a)}{f(y|a)} f(y|a) dy \geq c'(a),$$

where here, we have multipled and divided through by $f(y|a)$ in the integral.

Now, consider sending $\underline{y} \to -\infty$. As we do so, $-M \to -\infty$ as well. Let $K = \frac{c'(a)}{MF(\underline{y}|a)}$, so that $K \to \infty$ as well. Then (IC-FOC) is satisfied, since we can plug in our expressions for ϵ and K and write it as

$$\lim_{\underline{y} \to -\infty} \left\{ \left(\frac{c'(a)}{MF(\underline{y}|a)}\right) MF(\underline{y}|a) + \int_{\underline{y}}^{\infty} \bar{u} f_a(y|a) dy + \int_{\underline{y}}^{\infty} \left(\frac{F(\underline{y})}{1-F(\underline{y})}\right)\left(\frac{c'(a)}{MF(\underline{y}|a)}\right) f_a(y|a) dy \right\} \geq c'(a).$$

Note that $\int f_a(y|a) dy = 0$, since $\int f(y|a) dy = 1$ for all a. Given that $M \to \infty$, this (IC-FOC) constraint simplifies in the limit to

$$c'(a) + 0 + 0 \geq c'(a),$$

which is clearly satisfied.

The (IR) constraint is satisfied by construction. Therefore, we need only show that the principal's utility approach first-best with this construction. Her expected payoff equals

$$\lim_{\underline{y} \to -\infty} \left\{ F(\underline{y}|a) u^{-1}\left(-\frac{c'(a)}{MF(\underline{y}|a)}\right) + (1 - F(\underline{y}|a) u^{-1}\left(\bar{u} + \frac{F(\underline{y}|a)}{1-F(\underline{y}|a)}\left(\frac{c'(a)}{MF(\underline{y}|a)} + \bar{u}\right)\right)\right\}.$$

Because $u^{-1}(\cdot)$ is convex, this payoff is bounded from below by

$$\lim_{\underline{y} \to -\infty} \left\{ u^{-1}\left(-\frac{c'(a)}{MF(\underline{y}|a)} F(\underline{y}|a) + (1 - F(\underline{y}|a))\left(\bar{u} + \frac{F(\underline{y}|a)}{1-F(\underline{y}|a)}\left(\frac{c'(a)}{MF(\underline{y}|a)} + \bar{u}\right)\right)\right)\right\},$$

which can be simplified to

$$\lim_{\underline{y} \to -\infty} \left\{ u^{-1}\left(-\frac{c'(a)}{M} + (1 - F(\underline{y}|a))\bar{u} + \frac{c'(a)}{M} + F(\underline{y}|a)\bar{u}\right)\right\}.$$

But this expression simply equals $u^{-1}(\bar{u})$, which is the principal's first-best surplus.

1.5.2 Sufficient Conditions for the First-Order Approach

Reference:

- Rogerson, William P. 1985. "The First-Order Approach to Principal-Agent Problems." *Econometrica* 53 (6), 1357-1367.
- Jewitt, Ian. 1988. "Justifying the First-Order Approach to Principal-Agent Problems." *Econometrica* 56 (5), 1177-1190.

MLRP and CDFC as Sufficient Conditions

Now, let's consider conditions under which the first-order approach is valid. As discussed above, finding such conditions is a daunting challenge because the first-order approach requires (IC) to be concave for *any* $s(\cdot)$. The early literature has focused on two conditions that guarantee the validity of the first-order approach but are very restrictive.

The first condition is MLRP, which we have already defined and is arguably reasonable. The second condition is called the "Convexity of the Distribution Function Condition," or CDFC, and is quite demanding.

Definition 3 *A distribution $F(\cdot|a)$ satisfies the* ***convexity of the distribution function condition****, or* ***CDFC****, if $F_{aa}(y|a) \geq 0$ for all (y, a).*

CDFC requires the distribution function to be convex in effort. This condition does not have a natural economic interpretation, and it is not satisfied by most of the "classic" families of distribution that you might consider. We discuss a practical class of distributions that *do* satisfy it below.

First, let's state the result that MLRP and CDFC guarantee the validity of the first-order approach.

Proposition 2 *Suppose $F(\cdot)$ satisfies MLRP and CDFC, $c(\cdot)$ is convex, and a is interior. Suppose that $u(\cdot)$ is twice continuously differentiable and several other technical assumptions from Rogerson (1985) hold. Then a solution to the first-order approach problem is also a solution to the full contracting problem.*

A rough intuition for the validity of the FOA: Suppose that $f(\cdot|a)$ has common support on $[\underline{y}, \bar{y}] \subseteq \mathbb{R}$ and $s(\cdot)$ is differentiable (this is a big restriction because $s(\cdot)$ is endogenous). We want to show that (IC) is concave. Integrating by parts, we can write the agent's utility

$$\int_y u(s(y))f(y|a)dy - c(a) = u(s(\bar{y})) - \int_y u'(s(y))s'(y)F(y|a)dy - c(a)$$

Differentiate this expression twice with respect to a to yield:

$$-\int_y u'(s(y))s'(y)F_{aa}(y|a)dy - c''(a).$$

MLRP implies that $s'(y) \geq 0$, while CDFC implies that $F_{aa} \geq 0$. If we assume that $c(\cdot)$ is convex, then this expression is negative, which means that the agent's utility is concave in a.

A practical class of distributions that satisfies MLRP and CDFC

While CDFC is generally difficult to satisfy, there is a special class of distributions that satisfy MLRP and CDFC and are very tractable for applications. These distributions satisfy the "spanning condition" proposed by Grossman and Hart (1983).

Consider the following class of distributions.

Definition 4 *A distribution $F(y|a)$ is a* ***mixture distribution*** *if there exist $F_L(y)$, $F_H(y)$ such that*

$$F(y|a) = (1-a)F_L(y) + aF_H(y)$$

In a mixture distribution, output is randomly drawn according to either $F_L(y)$ or $F_H(y)$, where the probability of being drawn from each is determined by a.

Proposition 3 *Consider a mixture distribution $F(y|a)$, and suppose that $F_H(y)$ is higher, in the MLRP sense, than $F_L(y)$. Then $F(y|a)$ satisfies MLRP and CDFC.*

Intuition for this result: For simplicity, assume that F_L and F_H have densities f_L and f_H. Then $F(y|a)$ satisfies MLRP if

$$\frac{f_a}{f} = \frac{f_H(y) - f_L(y)}{(1-a)f_L(y) + af_H(y)}$$

is increasing in y, or equivalently,

$$\frac{(1-a)f_L(y) + af_H(y)}{f_H(y) - f_L(y)} = a + \frac{f_L(y)}{f_H(y) - f_L(y)}$$

is decreasing in y. But this is true whenever $\frac{f_H}{f_L}$ is increasing, which is implied by F_H MLRP-dominating F_L. $F(y|a)$ satisfies CDFC because $\frac{\partial^2 F}{\partial a^2} = 0$.

Alternative conditions for the first-order approach: Jewitt (1988) presents a set of sufficient conditions that are substantially easier to satisfy in many applications. While MLRP and CDFC both constrain the *output distribution* while leaving the agent's utility to be unconstrained (apart from weak concavity). Jewitt's conditions, on the other hand, include a combination of (milder) restrictions on the output distribution and restrictions on the utility function. Jewitt (1988) is highly recommended for the reader who is interested in learning more about these conditions.

1.5.3 What if the First-Order Approach Fails? Useful Simplifications

Reference:

- Grossman, Sanford J., and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica* 51 (1), 7-46.

What can we say if the first-order approach fails? Grossman and Hart (1983) tackle this problem in a general model. While the resulting constrained optimization problem is very difficult to solve in general, that paper identifies two simple but powerful observations that are very useful in applications.

The first observation is that we can rewrite a contract in terms of the *agent's utility,* rather than the monetary payment. Given a contract $s(\cdot)$ and output y, define $v(y) \equiv u(s(y))$. That is, $v(y)$ gives the agent's utility (ignoring effort costs) under the contract $s(\cdot)$ when output y is realized. Then the optimal contracting problem is

$$\max_{a,v(\cdot)} \mathbb{E}\left[y - u^{-1}(v(y))|a\right] \tag{1.2}$$

subject to

$$a \in \arg\max_{\tilde{a}} \left\{\mathbb{E}\left[v(y) - c(\tilde{a})|\tilde{a}\right]\right\} \tag{1.3}$$

and

$$\mathbb{E}\left[v(y) - c(a)|a\right] \geq \bar{u}. \tag{1.4}$$

The constraints of this optimal contracting problem are linear in $v(y)$, and the objective function is concave in $v(y)$ because $u^{-1}(\cdot)$ is convex. This problem is therefore a standard convex optimization problem. While this reformulation does not solve the inherent problem that the first-order approach might be invalid, it does simplify the analysis and allow some progress to be made. It also greatly simplifies numerical simulations, since many math programs have built-in tools for solving convex optimization problems.

The second observation is that we can break the optimal contracting problem into two steps. First, *fix* an effort level a and ask: what is the *lowest-cost* way for the principal to induce the agent to choose a? This problem is given by

$$C(a) \equiv \min_{v(\cdot)} \mathbb{E}\left[u^{-1}(v(y))|a\right]$$

subject to the constraints (1.3) and (1.4). We call the resulting function $C(a)$, which is the cost the principal must incur to induce a, the *incentive cost* of inducing a.

Second, we can solve for the optimal effort level, which maximizes $\mathbb{E}[y|a] - C(a)$. Note that $C(\cdot)$ is not necessarily convex or particularly well-behaved. Nevertheless, breaking the contracting problem into two steps helps us understand exactly how a and $v(\cdot)$ matter in the principal's payoff. It is also frequently possible to say something about $v(\cdot)$, and hence something about $C(\cdot)$, even if we cannot fully characterize the optimal contract. Indeed, many contracting papers focus mainly on the first step of characterizing the optimal contract for a given effort level, rather than worrying about simultaneously optimizing over effort.

1.6 A simple example: CARA utility, linear contracts

It is sometimes useful to have a simple example that illustrates the trade-off between insurance and incentives that lies at the heart of this model. Unfortunately, the problem is inherently complicated, so we have to make some restrictive assumptions to get a really tractable model.

Suppose that the agent has **constant absolute risk aversion (CARA)** utility:

$$u(s) = -\exp\left\{-r(s - c(a))\right\}.$$

For reasons that will become clear, the agent's cost of effort is in the exponent. We assume that this cost is quadratic: $c(a) = \frac{\gamma a^2}{2}$. Assume that

$$y = a + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$. Finally, restrict the contract to be linear:

$$s(y) = t + by.$$

We know from above that linear contracts are not optimal in this setting, and indeed, that we can approximate first-best. So requiring the contract to be linear is a big restriction. However, it can be justified by a more complicated model; see the discussion of Holmstrom (1987) in Chapter 5.

A nice property of CARA utility is that

$$\mathbb{E}_y\left[-\exp\{-r(t+by-c(a))\right] = -\exp\{-r(t+ba-c(a)\}\mathbb{E}_y\left[\exp\{-rb\epsilon\}\right]$$

Defining

$$\hat{s}(a) = t + ba - \frac{r}{2}b^2\sigma^2 - \frac{1}{2}\gamma a^2,$$

we can write

$$\mathbb{E}_y\left[u(s(y))|a\right] = -\exp\{-r\hat{s}(a)\}.$$

Therefore, (IC) amounts to simply maximizing $\hat{s}(a)$.

The optimal a^* solves $a^* = \frac{b}{\gamma}$. If we write the agent's outside option in terms of money, $\bar{u} = -\exp\{-r\bar{U}\}$, then the optimal contract sets

$$\hat{s}(a^*) = \bar{U},$$

which pins down $t = \bar{U} + \frac{r}{2}b^2\sigma^2 - \frac{1}{2}\frac{b^2}{\gamma}$.

If the principal's payoff is just $\mathbb{E}_y[y - s(y)|a]$, then the optimal incentive scheme simply solves

$$\max_b \left\{\frac{b}{\gamma} - \frac{1}{2}\frac{b^2}{\gamma} - \bar{U} - \frac{r}{2}b^2\sigma^2\right\},$$

with first-order condition

$$\frac{1}{\gamma} - \frac{b^*}{\gamma} - rb^*\sigma^2 = 0.$$

Therefore, the slope of the optimal linear contracts satisfies

$$b^* = \frac{1}{1+\gamma r\sigma^2}.$$

This is a nice, simple solution that exhibits intuitive comparative statics: increasing the agent's risk aversion (r), the noisiness of output (σ^2), or the marginal cost of effort (γ) leads to lower-powered incentives.

1.7 Further Reading

1.7.1 Tournaments

References:

- Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimal Labor Contracts." *Journal of Political Economy* 89 (5), 841-864.

- Green, Jerry R., and Nancy L. Stokey. 1983. "A Comparison of Tournaments and Contracts." *Journal of Political Economy* 91 (3), 349-364.

All of the contracts we have considered so far condition pay on cardinal information – i.e., the *amount* of output produced, the *value* of a signal. But in some settings, it might be much easier to contract on ordinal rather than cardinal information, as, for example, when a supervisor can rank her workers more easily than describing the exact differences in their performances. In other settings, the structure of rewards makes ordinal information relevant, as, for example, when two workers are competing for a single promotion opportunity. These contracting problems are characterized by multiple agents who are engaged in a **tournament**.

We can model such tournaments by considering a contracting problem with one principal and multiple agents. Assume that output y itself is not contractible, but that the *ranking* of each agent's output is contractible. That is, a contract assigns a payment to the agent with the highest output, a (potentially different) payment to the agent with the second-highest output, and so on. Since ranking is an imperfect proxy for y, such contracts typically do not do as well as the optimal contract when y is contractible. However, they can perform quite well in some circumstances; indeed, if all parties are risk-neutral, then they can attain first-best.

Why might the ordinal ranking of outputs be contractible when output itself is not? Lazear and Rosen (1981) suggest one reason: it might be less costly to acquire information about an ordinal ranking than about precise outputs, since a ranking requires only that we know whether one agent performed better than another. A second reason is implicit in the example of a promotion tournament: it might be that the contractual reward is attached to a scarce instrument, such as a managerial position, that can be allocated to only a subset of the agents. This justification, however, raises the question of why the principal has control over the payments attached to the promotion but cannot pay anything *apart* from that promotion. A third reason is that it might be easier to induce a supervisor to reveal information about the rankings of agents rather than revealing information about their cardinal output. We will return to this point when we discuss subjective performance evaluation in the next chapter.

1.7.2 Recent Papers

The insurance-incentives trade-off has generated a gigantic literature over the past four decades. I focus on a few (relatively) recent directions that the literature has taken.

One intuitive conclusion from this class of models is that riskier jobs should have lower-powered pay. During the 1980s and 1990s, a number of empirical papers have tested this conclusion in various industries and job categories. For the most part, they do *not* find evidence consistent with this conclusion; indeed, many empirical analyses find that higher-risk jobs tend to have *higher-powered* incentives. This empirical puzzle spurred new theories in the early 2000s. Prendergast (2002) analyzes a richer environment to identify variables that might be omitted from these empirical exercises.

Prendergast (2002) argues that organizations facing risky economic environments have the incentive to delegate more responsibilities to their agents. In a risky environment, the organization *does not know* which actions an agent should optimally take, so it instead leaves those decisions up to the agent. Consistent with the empirical findings, such high-risk organizations have to offer high-powered incentive schemes so that their agents' incentives are aligned with its own. That is, risk leads to less pay based on (noiseless) *inputs* and more pay based on (noisy) *outputs*. Georgiadis and Powell (2019) studies other implications of the classic incentive-insurance model, with the goal of identifying the minimal data an organization must gather in order to learn how to improve their incentive schemes.

Other papers have embedded class contracting models into either input or output markets. Raith (2003) considers a principal-agent relationship in a model of (differentiated) output competition. That paper argues that, once we alow free entry and exit of firms, more competitive industries, defined as industries that can sustain a smaller number of firms in equilibrium for a fixed market size, should provide higher-powered incentives for cost-reducing effort. The reason is that more competitive industries can sustain fewer firms in equilibrium; each of the survivors captures a large fraction of the overall market, increasing total costs cQ and hence the gains from reducing per-unit costs c.

Chade and Swinkels (2019) consider input rather than output markets. In practice, incentive contracts serve multiple roles: they incentivize incumbent workers, but they also attract a particular selection of workers to the job. For example, Enterprise Rent-a-Car combines menial initial jobs with the promise of well-paying careers for those who stay, with the stated goal of attracting highly-motivated college graduates from the bottom half of the grade distribution. Chade and Swinkels (2019) model these dual considerations in an imperfectly competitive labor market that suffers from both adverse selection (agents have private information about their abilities) and moral hazard (agents must be motivated to work hard).