4 Hidden Action, Moral Hazard

In this chapter we analyze the following contracting problem between a principal and an agent: the principal hires the agent to perform a task; the agent chooses her "effort intensity" a, which affects "performance" q. The principal cares only about performance. But effort is costly to the agent, and the principal has to compensate the agent for incurring these costs. If effort is unobservable, the best the principal can do is to relate compensation to performance. This compensation scheme will typically entail a loss, since performance is only a noisy signal of effort.

This class of principal-agent problems with moral hazard has been widely used as a representation of various standard economic relations. Among the most well-known applications are the theory of insurance under "moral hazard" (Arrow, 1970, and Spence and Zeckhauser, 1971, provide early analyses of this problem); the theory of the managerial firm (Alchian and Demsetz, 1972; Jensen and Meckling, 1976; Grossman and Hart, 1982); optimal sharecropping contracts between landlords and tenants (Stiglitz, 1974; Newbery and Stiglitz, 1979); efficiency wage theories (Shapiro and Stiglitz, 1984); and theories of accounting (see Demski and Kreps, 1982, for a survey). There are of course many other applications, and for each of them specific principal-agent models have been considered.

The basic moral hazard problem has a fairly simple structure, yet general conclusions have been difficult to obtain. As yet, the characterization of optimal contracts in the context of moral hazard is still somewhat limited. Very few general results can be obtained about the form of optimal contracts. However, this limitation has not prevented applications that use this paradigm from flourishing, as the short list in the preceding paragraph already indicates. Typically, applications have put more structure on the moral hazard problem under consideration, thus enabling a sharper characterization of the optimal incentive contract. This chapter begins by outlining the simplest possible model of a principal-agent relation, with only two possible performance outcomes. This model has been very popular in applications. Another simple model, which we consider next, is that of a normally distributed performance measure together with constant absolute risk-averse preferences for the agent and linear incentive contracts. Following the analysis of these two special models, we turn to a more general analysis, which highlights the central difficulty in deriving robust predictions on the form of optimal incentive contracts with moral hazard. To do so, we build in particular on the classical contributions by Mirrlees (1974, 1975, 1976), Holmström (1979), and Grossman and Hart (1983a). We then turn to two applications, relating to managerial incentive schemes and the optimality of debt as a financial instrument, respectively.

4.1 Two Performance Outcomes

Suppose for now that performance, or output q, can take only two values: $q \in \{0, 1\}$. When $q = 1$ the agent's performance is a "success," and when $q = 0$ it is a "failure." The probability of success is given by $\Pr(q = 1 | a) = p(a)$, which is strictly increasing and concave in a. Assume that $p(0) = 0$, $p(\infty) = 1$, and $p'(0) > 1$. The principal's utility function is given by

$$V(q - w)$$

where $V'(\cdot) > 0$ and $V''(\cdot) \le 0$. The agent's utility function is[1]

$$u(w) - \psi(a)$$

where $u'(\cdot) > 0$, $u''(\cdot) \le 0$, $\psi'(\cdot) > 0$, and $\psi''(\cdot) \ge 0$. We can make the convenient simplifying assumption that $\psi(a) = a$, which does not involve much loss of generality in this special model.

4.1.1 First-Best versus Second-Best Contracts

When the agent's choice of action is observable and verifiable, the agent's compensation can be made contingent on action choice. The optimal compensation contract is then the solution to the following maximization problem:[2]

$$\max_{a, w_i} p(a)V(1 - w_1) + [1 - p(a)]V(-w_0)$$

subject to

$$p(a)u(w_1) + [1 - p(a)]u(w_0) - a \ge \bar{u}$$

where $\bar{u}$ is the agent's outside option. Without loss of generality we can set $\bar{u} = 0$. Denoting by λ the Lagrange multiplier for the agent's individual-

1. It is standard in the literature to assume that the agent's utility is separable in income and effort. This assumption conveniently eliminates any considerations of income effects on the marginal cost of effort. In the general model without wealth constraints it guarantees that the agent's individual rationality constraint is always binding at the optimum.

rationality constraint, the first-order conditions with respect to w_1 and w_0 yield the following optimal coinsurance, or so-called Borch rule, between the principal and agent (see Borch, 1962):

$$\frac{V'(1 - w_1)}{u'(w_1)} = \lambda = \frac{V'(-w_0)}{u'(w_0)}$$

The first-order condition with respect to effort is

$$p'(a)[V(1 - w_1) - V(-w_0)] + \lambda p'(a)[u(w_1) - u(w_0)] - \lambda = 0$$

which, together with the Borch rule, determines the optimal action a.

Example 1: Risk-Neutral Principal $[V(x) = x]$ The optimum entails full insurance of the agent, with a constant wage w^* and an effort level a^* such that

$$u(w^*) = a^* \quad \text{and} \quad p'(a^*) = \frac{1}{u'(w^*)}$$

That is, marginal productivity of effort is equated with its marginal cost (from the perspective of the principal, who has to compensate the agent for her cost a^*).

Example 2: Risk-Neutral Agent $[u(x) = x]$ The optimum entails full insurance of the principal, with

$$w_1^* - w_0^* = 1 \quad \text{and} \quad p'(a^*) = 1$$

Once again, marginal productivity of effort is equated with its marginal cost for the principal.

When the agent's choice of action is unobservable, the compensation contract cannot be made contingent on action choice. Then the agent's output-

2. Alternatively, the maximization problem can be formulated as

$$\max_{a, w_i}\{p(a)u(w_1) + [1 - p(a)]u(w_0) - a\}$$

subject to

$$p(a)V(1 - w_1) + [1 - p(a)]V(-w_0) \ge \bar{V}$$

By varying $\bar{V}$ or $\bar{u}$ it is possible to allow for any division of surplus between the principal and agent. This is a useful shortcut that allows us to separate the analysis of the form of the optimal contract from the bargaining game between principal and agent.

contingent compensation induces her to choose an action to maximize her payoff:

$$\max_{a} p(a)u(w_1)+[1-p(a)]u(w_0)-a$$

The second-best contract is then obtained as a solution to the following problem:

$$\max_{a,w_i} p(a)V(1-w_1)+[1-p(a)]V(-w_0)$$

subject to

$$p(a)u(w_1)+[1-p(a)]u(w_0)-a\geq 0$$

and

$$a\in\arg\max_{\hat{a}} p(\hat{a})u(w_1)+[1-p(\hat{a})]u(w_0)-\hat{a} \qquad \text{(IC)}$$

The first-order condition of the agent's optimization problem is given by

$$p'(a)[u(w_1)-u(w_0)]=1 \qquad (4.1)$$

Given our assumptions on $p(\cdot)$ and $u(\cdot)$, there is a unique solution to this equation for any compensation contract $(w_0; w_1)$. We can therefore replace the agent's incentive constraint (IC) by the solution to equation (4.1). In general, replacing the agent's incentive constraint by the first-order conditions of the agent's optimization problem will involve a strict relaxation of the principal's problem, as we shall see. However, in this special two-outcome case the principal's constrained optimization problem remains unchanged following this substitution. This substitution simplifies the analysis enormously, as will become clear subsequently. We begin by analyzing the second-best problem in two classical cases: in the first case, principal and agent are risk neutral, but the agent faces a resource constraint; in the second case, at least one of the contracting parties is risk averse.

4.1.2 The Second Best with Bilateral Risk Neutrality and Resource Constraints for the Agent

When the agent is risk neutral, so that $u(x)=x$, first-best optimality requires that $p'(a^*)=1$. The first-order condition of the agent's optimization problem then also becomes $p'(a)(w_1-w_0)=1$, so that the first-best action could be

implemented with $w_1^*-w_0^*=1$. This solution can be interpreted as an upfront "sale" of the output to the agent for a price $-w_0^*>0$.

If $w_0=0$ and $w_1-w_0=1$, the agent would obtain an expected payoff equal to

$$p(a^*)-a^*$$

which is strictly positive, since $p''(a)$ is strictly negative and $p'(a^*)=1$. By contrast, the principal would obtain a zero payoff, since he is selling the output upfront at a zero price. A risk-neutral principal faced with the constraint $w_0\geq 0$ would therefore choose $w_0=0$ and, faced with $w_1=1/p'(a)$ in the second-best problem, would choose a to solve the following problem:

$$\max_{a} p(a)(1-w_1)$$

subject to

$$p'(a)w_1=1$$

Solving this problem yields

$$p'(a)=1-\frac{p(a)p''(a)}{[p'(a)]^2}$$

As can be readily checked, the solution to this equation is smaller than a^*. This result is intuitive, since inducing more effort provision by the agent here requires giving her more surplus.

4.1.3 Simple Applications

The results just presented can be interpreted in several ways. For example, one can think of the agent as a manager of a firm and the principal an investor in this firm. Then w_1 is interpreted as "inside" equity and $(1-w_1)$ as "outside" equity (Jensen and Meckling, 1976); or, still thinking of the agent as a manager and w_1 as inside equity, $(1-w_1)$ can be thought of as the outstanding debt of the firm (Myers, 1977); an alternative interpretation is that the agent is an agricultural laborer under a sharecropping contract (Stiglitz, 1974). Under all these interpretations a lower w_1 reduces incentives to work, and it can even become perverse for the principal, generating a form of "Laffer curve" effect. This can mean, for example, that reducing the face value of debt can increase its real value by reducing the debt burden "overhanging" the agent (Myers, 1977).

The standard agency problem assumes that actions are unobservable but output is observable. But for some contracting problems in practice even performance may be difficult to observe or describe. To capture these types of situations some applications involving moral hazard assume that contracting on q is too costly but actions can be observed at a cost through monitoring. One important example of such applications is the efficiency wage model (Shapiro and Stiglitz, 1984). In this model the focus is on trying to induce effort through monitoring. Assume, for example, that effort can be verified perfectly at a monitoring expense M. The full monitoring optimum then solves

$$\max_{a,w} p(a) - w - M$$

subject to

$$w - a \geq 0$$

and

$$w \geq 0$$

which yields $w^* = a^*$ and $p'(a^*) = 1$.

But suppose now that the principal is able to verify the agent's action with probability 0.5 without spending M. If the agent is found shirking, it is optimal to give her the lowest possible compensation. In this case, with a limited wealth constraint, $w \geq 0$, the variable w is set equal to 0. If the principal decides not to spend M, his problem is

$$\max_{a,w} p(a) - w$$

subject to

$$w - a \geq 0.5w$$

The LHS of the incentive constraint is the agent's payoff if she chooses the prescribed action a. The RHS is the agent's maximum payoff if she decides to shirk. In that case it is best not to exert any effort and gamble on the possibility that the agent will not be caught. Now the principal has to give the agent a compensation that is twice her effort level. In other words, the principal gives the agent rents that she would lose when caught shirking.

Having to concede rents once again lowers the principal's desire to induce effort, so that optimal effort is lower than a^*. Finally, the choice of whether or not to monitor depends on the size of M.

4.1.4 Bilateral Risk Aversion

Let us now return to our simple theoretical example. In the absence of risk aversion on the part of the agent and no wealth constraints, the first best can be achieved by letting the agent "buy" the output from the principal. In contrast, if only the agent is risk averse, the first-best solution requires a constant wage, independent of performance. Of course, this completely eliminates effort incentives if effort is not observable. Optimal risk sharing under bilateral risk aversion does not provide for full insurance of the agent, but risk aversion on the part of the agent still prevents first-best outcomes under moral hazard in this case. Indeed, the principal then solves

$$\max_{a,w_i} p(a)V(1 - w_1) + [1 - p(a)]V(-w_0)$$

subject to

$$p(a)u(w_1) + [1 - p(a)]u(w_0) - a \geq 0 \tag{IR}$$

and

$$p'(a)[u(w_1) - u(w_0)] = 1 \tag{IC}$$

Letting λ and μ denote the respective Lagrange multipliers of the (IR) and (IC) constraints and taking derivatives with respect to w_0 and w_1 yields

$$\frac{V'(1 - w_1)}{u'(w_1)} = \lambda + \mu \frac{p'(a)}{p(a)}$$

and

$$\frac{V'(-w_0)}{u'(w_0)} = \lambda - \mu \frac{p'(a)}{1 - p(a)}$$

When $\mu = 0$, we obtain the Borch rule. However, at the optimum $\mu > 0$ under quite general conditions, as we shall show, so that optimal insurance is distorted: the agent gets a larger (smaller) share of the surplus relative to the Borch rule in case of high (low) performance. Specifically, in order to induce effort, the agent is rewarded (punished) for outcomes whose frequency rises (falls) with effort. In our two-outcome setting, this result is particularly simple: $q = 1$ is rewarded, and $q = 0$ is punished.

Paradoxically, the principal perfectly predicts the effort level associated with any incentive contract and realizes that the reward or punishment corresponds only to good or bad luck. Nevertheless, the incentive scheme is needed to induce effort provision.

4.1.5 The Value of Information

Assume now that the contract can be made contingent not only on q but also on another variable $s \in \{0, 1\}$. This variable may be independent of effort (think of the state of the business cycle) or may depend on it (think of consumer satisfaction indices, if the agent is a salesperson). However, the variable s does not enter directly into the agent's or the principal's objective functions. Specifically, assume $\Pr(q = i, s = j \mid a) = p_{ij}(a)$. The principal can now offer the agent a compensation level w_{ij}, to solve

$$\max_{a, w_{ij}} \sum_{i=0}^{1} \sum_{j=0}^{1} p_{ij}(a) V(i - w_{ij})$$

subject to

$$\sum_{i=0}^{1} \sum_{j=0}^{1} p_{ij}(a) u(w_{ij}) \geq a \tag{IR}$$

$$\sum_{i=0}^{1} \sum_{j=0}^{1} p'_{ij}(a) u(w_{ij}) = 1 \tag{IC}$$

Again, letting λ and μ denote the respective Lagrange multipliers of the (IR) and (IC) constraints, and taking derivatives with respect to w_{ij}, one obtains the following first-order conditions with respect to w_{ij}:

$$\frac{V'(i - w_{ij})}{u'(w_{ij})} = \lambda + \mu \frac{p'_{ij}(a)}{p_{ij}(a)}$$

Hence the variable s drops out of the incentive scheme (i.e., $w_{ij} = w_i$ for $i = 0, 1$) if for all a

$$\frac{p'_{i0}(a)}{p_{i0}(a)} = \frac{p'_{i1}(a)}{p_{i1}(a)} \quad \text{for } i = 0, 1$$

where a denotes the second-best action choice.

Integrating these conditions yields

$$p_{i0}(a) = k_i p_{i1}(a) \quad \text{for } i = 0, 1$$

where the k_i's are positive constants. The condition for s to be absent from the agent's incentive scheme is thus that, for each q, changing effort yields the same change in relative density whatever s may be. When this is satisfied, one says that q is a "sufficient statistic" for (q, s) with respect to a, or that s is not "informative" about a given q.

When q is not a sufficient statistic for a, taking s into account improves the principal's payoff by allowing for a more precise signal about effort, and thus a more favorable trade-off between effort provision and insurance.

The sufficient statistic result is due to Holmström (1979) (see also Shavell, 1979a, 1979b; and Harris and Raviv, 1979). It extends to general settings, as will be shown later on in this chapter and in Chapters 6 and 8.

4.2 Linear Contracts, Normally Distributed Performance, and Exponential Utility

Next to the two-performance-outcome case, another widely used special case involves linear contracts, normally distributed performance, and exponential utility. Performance is assumed to be equal to effort plus noise: $q = a + \varepsilon$, where ε is normally distributed with zero mean and variance σ^2. The principal is assumed to be risk neutral. The agent has constant absolute risk-averse (CARA) risk preferences represented by the following negative exponential utility function:

$$u(w, a) = -e^{-\eta(w - \psi(a))}$$

where w is the amount of monetary compensation and $\eta > 0$ is the agent's coefficient of absolute risk aversion ($\eta = -u''/u'$). Note that in contrast with the earlier formulation, effort cost here is measured in monetary units. For simplicity the cost-of-effort function is assumed to be quadratic: $\psi(a) = \frac{1}{2}ca^2$.

Suppose that the principal and agent can write only linear contracts of the form

$$w = t + sq$$

where t is the fixed compensation level and s is the variable, performance-