LECTURES ON THE THEORY OF CONTRACTS IN CORPORATE FINANCE:
FROM DISCRETE-TIME TO CONTINUOUS-TIME MODELS

Jaeyoung Sung

Department of Finance (M/C 168)
University of Illinois at Chicago
601 South Morgan Street
Chicago, Illinois 60607-7124
jsung@uic.edu

This work is partially supported by Com^{2}MaC-KOSEF, Korea.

Preface

This manuscript is based on lectures that I have given to graduate students at the University of Illinois at Chicago and at Pohang University of Science and Technology, Pohang, Korea. The objective is to introduce methodological aspects of contracting theories and their applications in corporate finance problems. The target audience for this manuscript are students and researchers who have basic background in mathematics including introductory stochastic calculus with a limited exposure to finance.

Chapter 1 provides a brief overview of various areas of finance, and describes how agency problems can arise in corporate management. Then we examine moral hazard and adverse selection problems using discrete-time and continuous-time models in chapters 2 to 4. In particular, the moral hazard problems are investigated with a discrete-time formulation in chapter 2 and with a continuous-time formulation in chapter 3. Chapter 4 introduces interactions between moral hazard and adverse selection problems.

Contents

Preface **iii**

1 Introduction **1**
- 1.1 Corporate Investment Decisions . . . 2
- 1.2 Principal-Agent Problems . . . 3
  - 1.2.1 Examples: . . . 4
- 1.3 How to Control Moral Hazard . . . 7
  - 1.3.1 Monitoring . . . 7
  - 1.3.2 Bonding . . . 7
  - 1.3.3 Explicit incentive contracts . . . 7

2 Moral Hazard and Incentive Contracts: a Discrete-Time Approach **13**
- 2.1 The Basic Model . . . 13
- 2.2 The First-Best Solution: the case of full information . . . 14
- 2.3 The Second-Best Solution: the case of moral hazard . . . 17
- 2.4 The Shape of the Second-best Contract . . . 20
- 2.5 The Value of Informative Signal in Contracting . . . 21
  - 2.5.1 Application to Insurance Deductibles . . . 22
- 2.6 The Validity of the First-order Approach . . . 23
- 2.7 Normally Distributed Outcome . . . 27

3 Moral Hazard and Incentive Contracts: a Continuous-Time Approach **29**
- 3.1 The First-best Contract . . . 30
- 3.2 The Second-best Contract . . . 33
- 3.3 Applications . . . 40
  - 3.3.1 Unobservable Project Selection . . . 43

4 Adverse Selection and Moral Hazard **45**
4.1 The Model: Pure Adverse Selection 47
4.2 Managerial Contracting under Adverse Selection and Moral Hazard 52
4.2.1 Applications 57

A Dynamic Programming Equation with Exponential Utility **63**

Bibliography **67**

Index **70**

Chapter 1

Introduction

There are many areas in finance. Finance in general is a discipline to study investors' behaviors in their investing and financing activities in product and capital markets. Basic areas include corporate finance, investments and financial intermediation.

Corporate finance is concerned with investors' investing and financing behaviors in product markets. Investments deals with the same issues in capital markets. In product markets, real assets are traded. Examples of real assets are factories, machinery, lands, production technology, etc. Major issues in corporate finance include capital budgeting (choice of right projects), capital structure (choice of financing methods), dividend policy, and mergers/acquisitions.

In capital markets, financial assets such as stocks, bonds, and derivatives are traded. Major issues in investments include asset pricing, consumption/investment decisions, and market efficiency. A popular subarea of investments is derivatives which is concerned with pricing derivative assets such as options, futures and swaps

Financial intermediation is also an important area to facilitate investors' various investing and finance activities in both product and capital markets. This area is concerned with the efficiency of security trading systems (market microstructure), and brokerage functions of various financial institutions such as investment banking and insurance.

In typical MBA programs, the major three areas of finance are covered in many different finance courses. such as corporate finance, investments, options and futures (financial engineering), mergers and acquisitions, international finance, money and banking, and insurance. Sometimes, real estate is also included in the finance program.

In this manuscript, we mainly focus on the agency relationships between the manager and investors in corporate management. The agency theories provide a major paradigm to study modern corporate finance.

1.1 Corporate Investment Decisions

The values of financial assets fundamentally depend on the value of underlying assets. For example, the fundamental determinant of corporate securities such as stocks, bonds and their derivatives is the economic value of the corporation. The firm creates its value by choosing right projects, financing with right kinds of capital market instruments, and managing for the best outcomes (profits).

In order to examine how the value is created, let us consider a simple example of corporate investment decision-making problem. Suppose that the current riskfree interest rate is 5%, and that the firm has an opportunity to invest in a project with the following prospect of future cashflows:

$t = 0$	$t = 1$	$t = 2$
$-1{,}000$	500	800

Then the firm's problems is to decide whether it should accept this project. Since the opportunity cost of taking on this project (the cost of capital) is the riskfree interest rate, the decision rule can be as simple as follows: If the project yields higher return than the riskfree interest rate, then accept the project.

Since

$$\frac{500}{1.05} + \frac{800}{1.05^2} - 1000 = 201.81 > 0,$$

the firm should accept the project. In other words, since the present value (PV) of the future cashflows from the project is larger than the initial required investment, the project should be accepted. By accepting this project, the manager increases the value of the firm by $201.81.

The above practice for the project selection decision is called "positive-NPV rule." The NPV (net present value) of the project is defined as follows:

$$\text{NPV} := \text{PV of future cashflows} - \text{initial investment}$$

NPV is the amount of value created by the project, and it is also sometimes called "economic value added."

The optimal project selection rule is as follows: Take the project in question if its NPV is positive; and otherwise reject. Basically, the positive-NPV rule can lead to the maximization of shareholders' wealth.

However, note that typically the decision maker is the manager or the CEO of the firm, not shareholders. The positive-NPV rule implicitly assumes that the manager always behave in the best interest of shareholders. Then one may ask whether the assumption can be valid in reality.

Given that the manager is an independent economic agent, there is no *a priori* reason why the manager should behaves to maximize the shareholders' wealth unless the manager has incentives to do so. Then the question is: What kind of incentive systems are needed to motivate the manager to work for shareholders' best interests?

To get an insight into this question, we need to look at the relationship between the manager and shareholders.

1.2 Principal-Agent Problems

The manager and shareholders are related to each other *via* a managerial compensation contract. In the literature, the party like the manager is called the agent and the party like shareholders the principal.

A problem arises because the agent acts in his own best interest, not in the best interest of shareholders. If the principal can costlessly monitor the agent's actions, then the principal will specify in the contract all courses of actions the agent must follow. If the agent deviates from the specification, then the agent will be punished so severely that he has no incentives to go against it. If the principal cannot costlessly monitor the agent's actions, the principal cannot enforce the detailed action plan to the agent. As a result, the agent may have discretion on some actions.

The structure of the contract affects the agent's behavior, particularly on actions that are unobservable to the principal. The agent carries out the unobservable actions to his own interest, not to the principal's best interest. This kind of problem is called "moral hazard problem." Note however that the term is a misnomer. Instead, it should be called "morale hazard problem."

In general, the principal-agent problems can arise because of the following reasons:

- Postcontractual opportunism
  Hidden action: moral hazard
  Hidden information (information asymmetry after contracting)

- Precontractual opportunism
  Adverse selection (information asymmetry before contracting)

In brief, the principal-agent problem arises because of information asymmetry between the two parties before or after the contract is signed. Consequently, both the principal and agent play different kinds of games against each other depending upon different kinds of information asymmetry problems. The structures of different games are illustrated in Figure 1.1 through 1.5.

1.2.1 Examples:

It is instructive to go over a few real life examples of agency problems. The examples in this subsection are adapted from Milgrom and Roberts [1992, p170].

The U.S. Savings and Loan Crisis

Savings and loan Associations (S&Ls) are financial institutions that borrow money from the public in the form of deposits and invest it primarily in residential mortgage loans. The deposits of individual investors in an S&L are insured by the Federal Savings and Loan Insurance Corporation (FSLIC), a U.S. federal government agency, until 1990. Each S&L purchases the deposit insurance. The size of the insurance premiums were not linked to the riskiness of the S&Ls' portfolio of loans and other investments. In case of default by an S&L, the depositors will get paid by the FSLIC.

In 1980s, the S&Ls turned to riskier investments, including loans on commercial real estate and "junk bonds" (high-yielding but risky corporate bonds). Unfortunately, a lethal combination of events occured: (1) Commercial real estate market collapsed in several parts of the country; and (2) defaults by some corporations on their junk bonds. As a result, over 500 S&Ls slipped into bankruptcy. Because of this massive scale of the bankruptcy problem, the FSLIC's reserves were inadequate to cover its promises to protect depositors and U.S. tax payers ended up with paying the bill for hundreds of billions of dollars.

Who or what is to blame? Moral hazard. The deposit insurance program was designed so poorly that its low capital requirement gave S&Ls managers incentives to take risks aggressively at the expense of depositors' and U.S. tax payers' wealth. In 1980s, S&L capital requirements (the amount of the S&L owners' own money at risk) was as low as 3% of its total investments.

To see how this low capital requirement creates moral hazard problems, let us consider the following example. Suppose a S&L can choose between to possible investments, "safe" and "risky." Both investments require an initial outlay of $100, which consists of $97 from deposits and $3 from the S&L's own capital. Assume that the interest rate on the deposits is zero. The safe investment returns either $100 or $110 with equal probabilities. The risky investment returns either $125 or $65 with equal probabilities.

In this case, although the safe investment is optimal for the society, the S&L has a very strong incentive to choose the risky investment. To make matters worse, competition among S&Ls intensified managers' risk-taking behaviors.

The above case is a moral hazard problem arisen in financial institutions. The same problem can also occur in corporations.

Managerial Moral Hazard

Consider a widely-held public firm which is characterized by disperse ownership across small investors. Thus no individual shareholder has any real incentives to monitor managers in order to ensure that they are running the firm in shareholders' interests.

In this case, there can be many potential sources of corporate moral hazard problems as follows.

- Corporate executives may invest firm's earnings in low-value projects to expand their empires.
- They may pay themselves exorbitantly and lavish expensive perquisites upon themselves.
- They may run ongoing operations in a way to pursue their own personal goals other than maximizing the value of the firm.
- They may resist attempts to force more profitable operations, especially by resisting takeovers that threaten their jobs.

To see why manager's takeover defense can be viewed as a moral hazard problem, let us look at the history. During the 1980s, there was a wave of hostile takeovers in the U.S. A hostile takeover is the acquisition of enough number of shares in a company to give a controlling ownership in the firm, where the offer to acquire the firm is opposed by the target company's executives and directors. Successful hostile takeover attempts generally resulted in the replacement of target firms' senior management and the naming of new board of directors. Many observers have interpreted the hostile takeovers as a corrective response to managerial moral hazard or managerial incompetence.

The prices paid for the stock of firms in hostile takeovers in this period on average represented a 50% premium over the target's original market value. This indicates that manager's takeover defense can pose a great obstacle in creating extra wealth for shareholders. A major form of takeover defense is poison pills. An example of poison pills is "Shareholder Rights Plans." For example, existing managers sell "Shareholder Rights Plans" to their shareholders, where the plans give rights to their holders to buy shares of the (target) firm at very low prices if the takeover occurs. By selling these plans, the managers in effect remove the ability of the owners of the firm to sell their shares to a corporate raider.

Current systems allow boards of directors to adopt poison pills without shareholders' approval. The empirical evidence is that adopting a poison pill typically reduces firm's share price. It suggests that the adoption does not serve shareholder's interests.

1.3 How to Control Moral Hazard

Two major conditions for a moral hazard problem to arise between the principal and agent are: (1) conflicts of interests, and (2) inability to write enforceable contracts covering all crucial elements of transactions.

There are several ways to mitigate a moral hazard problem.

1.3.1 Monitoring

Shareholders may devote resources to monitoring and verification, and use the results of monitoring as the basis for rewards and penalties. They may use monitoring in preventing inappropriate behavior directly catching it before it occurs. E.g. U.S. firms are not allowed to publish financial statements until they have been verified by independent auditors.

1.3.2 Bonding

In some industries, it is common to require the posting of bonds to guarantee performance. The bond is a sum of money to be forfeited in the event that inappropriate behavior is detected.

For example, the capital provided by the owners of S&L acts like a bond because in the event of losses the capital must be paid out to meet obligations.

1.3.3 Explicit incentive contracts

Sometimes monitoring is too costly. It may be impossible to monitor individual managerial actions and efforts in various managerial activities, but it may still possible to measure the outcome of managerial actions and efforts. The outcome may be the realized (or reported) annual profit of the firm.

Unfortunately, perfect connections between unobservable actions and observed outcome are rare. Nevertheless, the principal may use the outcome to provide the agent incentives to work toward the principal's interests by rewarding good outcomes. For example, in practice, explicit incentives contracts are constructed so that the bonus depends on various quantities including accounting earnings per share (EPS), return on equity (ROE), return on asset (ORA), economic value added (EVA, net operating profit after tax - cost of capital), stock options with fixed exercise prices, stock options with fixed exercise prices, stock options with adjustable exercise prices (exercise price increases at a rate equal to cost of capital minus dividend yield minus compensation-risk premium), and stock ownerships.

The bonus plans based on above quantities however may not elicit desired managerial behaviors, because the plans can sometimes induce the manager to go against shareholders' interests. For example, the managerial bonus plans in Xerox were based on the EPS. According to the W.S.J. article (June 1, 2001, C1) by Maremont and Bandler, "Concession by Xerox May Not Satisfy the SEC," the SEC found that Xerox inflated its revenues by using a low discount rate, say 6%, in valuing lease contracts with companies in Latin America when the local interest rates were as high as 30% in Brazil. It was argued that Xerox managers used lower discount rates in an attempt to receive large bonuses for high revenues. However, this kind of misuses of accounting rules costs shareholders dearly in higher corporate taxes as well as larger executive bonuses.

The bonus plans based on the other quantities can also sometimes produce undesirable managerial behaviors.

Exercise 1.1 *Stock options and ownerships are popular ways used in practice to reduce managerial moral hazard problems. Discuss why they can effectively or ineffectively serve the purpose.*

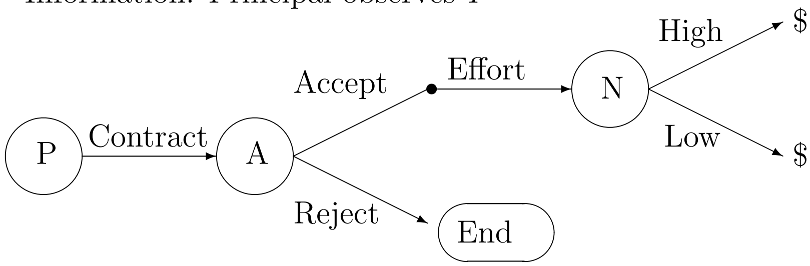
Then what will be the right quantity for the bonus plan to be based on? The bonus plan should be depend on the outcome of managerial effort and decisions. Note that typically the outcome is interpreted as the overall profit of the firm. However, the overall profit can be a sum of market-dependent profit and firm-specific profit. Of the two components, the market-dependent profit has nothing to do with managerial effort and decisions. Suppose that the firm's profit in a particular year has increased or decreased because of favorable or unfavorable market situations. Then the manager should not be rewarded or penalized because of incremental or decremental market-dependent portion of the overall profit.

What the manager can affect by his effort and decisions is the firm-specific portion of the overall profit. Thus the managerial bonus plan has to depend only on the the firm-specific portion of the overall profit, not on the overall profit. In this sense, a bonus plan tied with stock price is not efficient, because the manager can be rewarded or penalized for market movements.

In this manuscript, we focus on the design of incentive contracts based on firm-specific portion of the overall profit.

Figure 1.1: Moral hazard with hidden action

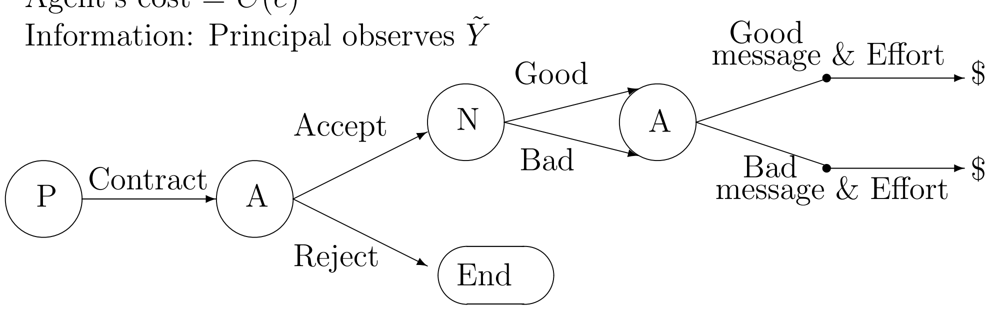
Outcome: $\tilde{Y} = e + \tilde{\varepsilon}$
Agent's cost $= C(e)$
Information: Principal observes $\tilde{Y}$



Example: After managerial contracting, the principal cannot observe managerial effort.

Figure 1.2: Moral hazard: hidden information

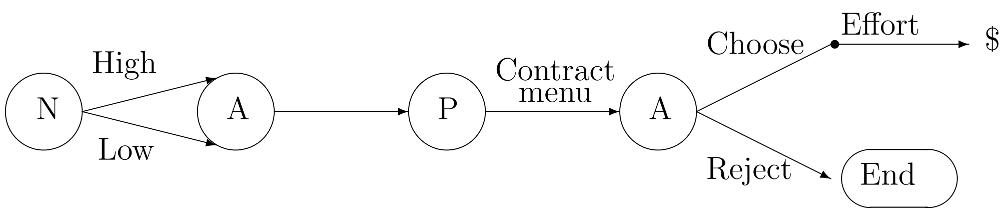
Outcome: $\tilde{Y} = e + \tilde{I}$
Agent's cost $= C(e)$
Information: Principal observes $\tilde{Y}$



Example: After the contracting, the manager knows better about the firm than the principal.

Figure 1.3: Adverse Selection

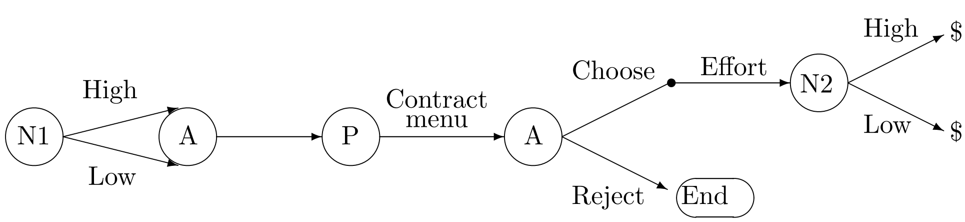
Outcome:$\tilde{Y} = e + \tilde{a}$
Agent's cost $= C(a, e)$
Information: Principal observes $\tilde{Y}$



Example: Before contracting, the manager knows better about his ability.

Figure 1.4: Moral Hazard (hidden action) and Adverse Selection

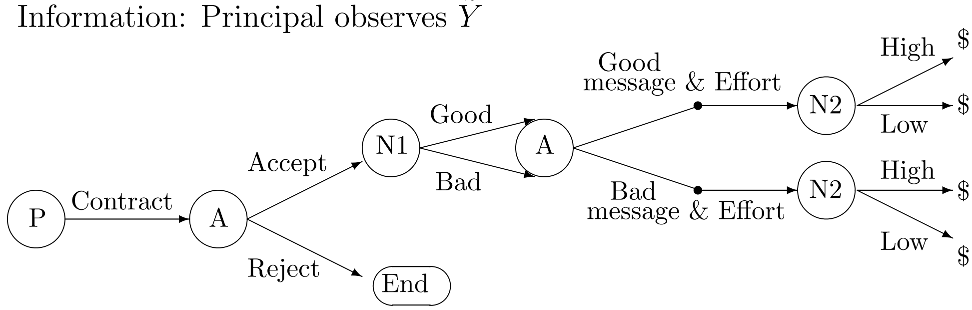
Outcome:$\tilde{Y} = e + \tilde{a} + \tilde{\varepsilon}$
Agent's cost $= C(a, e)$
Information: Principal observes $\tilde{Y}$



Example: Before contracting, the manager knows his ability.
After contract, the managerial effort is unobservable.

Figure 1.5: Moral hazard: hidden action + hidden information

Outcome:$\tilde{Y} = e + \tilde{I} + \tilde{\varepsilon}$
Agent's cost $= C(e)$
Information: Principal observes $\tilde{Y}$



Example: After contracting, the manager knows more about the firm than the principal, and exerts unobservable effort.

Chapter 2

Moral Hazard and Incentive Contracts: a Discrete-Time Approach

Although the economic significance of contractual relationships between economic agents have long been recognized since Adam Smith in the 18th century, it was not until only recently that economists started formal rigorous economic analyses. Pioneers in modern contracting theories include Ross [1973], Mirrlees [1974, 1976] and Holmstrom [1979]. These researchers tried to formulate contracting problems by using discrete-time models. In this manuscript, we also start with discrete-time models, and then in the next chapter we discuss continuous-time models.

2.1 The Basic Model

There are two discrete time periods, 0 and 1. The contract between the principal and agent can only be written on verifiable information. Assume that the principal and the agent are endowed with utilities of wealth given by $U(w)$ and $V(w)$, respectively where $U, V : \mathcal{R} \to \mathcal{R}$ are continuously twice differentiable, strictly increasing, and concave. Both the principal and agent are expected utility maximizers.

The principal precommits to contract S, i.e. he makes a take-it-or-leave-it offer to the agent whose reservation utility is exogenously given as $\theta \in \mathcal{R}$. θ is the level of the agent's utility that he can achieve by rejecting the offer and taking a job elsewhere. Contract S can be enforced costlessly and accurately (by court). There is no financial markets for the contract, and thus the manager cannot trade his contract in the market in an attempt to undo the incentives implied by the contract.

If the agent accepts the offer at time 0, the principal gives him an asset to manage until time 1. Then the agent exerts effort (a hidden action) $a \in A$ which affects the outcome $x \in X$. (Typically, $A, X \subset \mathcal{R}$.) In particular, let $X = [x_l, x_h]$, where x_l and x_h are independent of a. The outcome x will be realized at time 1. One may view x as the corporate economic profit realized at the end of year. Note that economic profit $\neq$ accounting profit.

For effort a, the agent incurs the disutility of effort $D(a)$, where $D'(a) \geq 0$ and $D''(a) > 0$. We further assume that the agent's net utility is separable in wealth and effort and given by $V(w) - D(a)$. Assume that $x = h(a, \varepsilon)$, where ε is a real random variable. Let the pdf of x given by $f(x, a)$. $f(x, a) > 0$, $\forall (x, a)$. f_a and f_{aa} exist and are continuous.

Assume that a is unobservable to the principal, but x is observable and verifiable. Thus, the contract S can only be based on observables such as x, but not on unobservable a. We call $S(x)$ the salary scheme or the sharing rule. At time 1, the agent will be compensated for his effort according to the salary scheme $S(x)$.

Thus, the principal's problem is to design $S(x)$ in order to maximize his expected utility, or $\max E[U(x - S(x))]$, while making $S(x)$ acceptable by the agent. In designing the contracts, the following questions arise: (1) How does the optimal sharing rule look like? (2) What will be the effect of moral hazard to the principal's welfare? Or what will be the size of agency cost? (3) What is the agent's optimal response or behavior given the optimal sharing rule in place?

2.2 The First-Best Solution: the case of full information

Before tackling the above problem, let us look at the benchmark case in which both a and x are observable and verifiable. The principal's problem can be stated as follows:

Problem 2.1 *Choose S and a to*

$$\begin{aligned} \max \quad & \int_{x_l}^{x_h} U(x-S) f(x,a) dx \\ s.t. \quad & \int_{x_l}^{x_h} V(S) f(x,a) dx - D(a) \geq \theta \end{aligned}$$

The constraint is called the agent's participation or individual rationality (IR) constraint. The Lagrangian is

$$L = \int_{x_l}^{x_h} \{U(x-S)) + \lambda V(S)\} f(x,a) dx - \lambda D(a) - \lambda \theta$$

Assuming an interior solution, the first order conditions are

$$\begin{aligned} & \forall x \in X, \quad \frac{U'(x-S)}{V'(S(x))} = \lambda \\ & \int_{x_l}^{x_h} \{U(x-S)) + \lambda V(S)\} f_a(x,a) dx - \lambda D'(a) = 0 \\ & \int_{x_l}^{x_h} V(S) f(x,a) dx - D(a) = \theta \end{aligned}$$

The first condition implies that the salary scheme S has to be designed so that the ratios of marginal utilities of wealth are equalized across states.

Remark 1: Suppose that $U' = 1$, i.e. the principal is risk-neutral. Then the first FOC implies that $V'(S(x))$ should be constant for all states x. That is, $S(x)$ has to be constant across all x's, i.e. the principal pays a fixed salary to the agent for his level of effort a, and then receives $x - S$. The intuition is straightforward: Since the principal is risk-neutral and the agent is risk-averse, the principal can take risks more easily than the agent. Therefore, the principal bears all risks and the agent receives a riskfree payoff.

For optimal level of effort choice, let us look at the second FOC.

$$\begin{aligned}
&\lambda D'(a) \\
&= \int_{x_l}^{x_h} \{U(x - S(x))) + \lambda V(S(x))\} f_a(x,a)dx \\
&= \{U(x - S(x))) + \lambda V(S(x))\} F_a(x,a)|_{x_l}^{x_h} \\
&- \int_{x_l}^{x_h} \{U'(x - S(x)))(1 - S'(x))) + \lambda V'(S(x))S'(x)\} F_a(x,a)dx,
\end{aligned}$$

where F is the distribution function. Note that since $\forall a \in A$, $F(x_l, a) = 0$ and $F(x_h, a) = 1$, we have $F_a(x_l, a) = F_a(x_h, a) = 0$. Therefore, substituting the first FOC, we have

$$\lambda D'(a) = -\int_{x_l}^{x_h} U'(x - S(x)))F_a(x,a)dx,$$

or

$$D'(a) = -\int_{x_l}^{x_h} V'(S(x)))F_a(x,a)dx.$$

When the principal is risk-neutral and the agent is risk-averse, S is constant. Thus

$$\begin{aligned}
D'(a) &= -V'(S)\int_{x_l}^{x_h} F_a(x,a)dx \\
&= V'(S)\int_{x_l}^{x_h} x f_a(x,a)dx \\
&= V'(S)\frac{\partial}{\partial a}E[x|a]
\end{aligned}$$

The last quantity is the agent's marginal (monetary) value of incremental output for an additional unit of effort, and $D'(a)$ the agent's marginal (monetary) cost of effort.

Thus the optimal level of effort is determined such that the agent's marginal value of an additional output is equal to the agent's marginal disutility of effort. Moreover, when the principal is risk-neutral and the agent is risk-averse, the last FOC implies

$$V(S) - D(a) = \theta.$$

2.3 The Second-Best Solution: the case of moral hazard

In this section, the agent's effort a is unobservable to the principal. The principal's problem is stated as follows:

Problem 2.2 *Choose S and a to*

$$\begin{aligned} \max \quad & \int_{x_l}^{x_h} U(x - S(x)) f(x,a) dx \\ s.t. \quad & \int_{x_l}^{x_h} V(S(x)) f(x,a) dx - D(a) \geq \theta \\ & a \in \arg\max_{a' \in A} \int_{x_l}^{x_h} V(S(x)) f(x,a') dx - D(a'). \end{aligned}$$

The last condition is called "the incentive compatibility" (IC) constraint for effort. Although its statement appears to be simple, the solution to Problem 2.2 is not easy to obtain. In some cases, Problem 2.2 can be solved under highly restrictive assumptions on U, V, D, and f. Even then, solutions are algebraically too complex to interprete.

To transform the problem into a manageable form, consider the IC constraint. Again assuming an interior solution, the IC implies that

$$\int_{x_l}^{x_h} V(S(x)) f_a(x,a) dx - D'(a) = 0, \tag{2.1}$$

$$\int_{x_l}^{x_h} V(S(x)) f_{aa}(x,a) dx - D''(a) \leq 0. \tag{2.2}$$

The above two are the local first- and second-order conditions.

To solve Problem 2.2, one may try to replace the IC constraint with the first order condition in (2.1). The problem with the IC replaced with the FOC is called the principal's relaxed problem, and such a practice is called "the first order approach" to incentive contracting problems. Since it is possible that the FOC in (2.1) is not even a necessary condition, it is not guaranteed that the relaxed problem yields an optimal solution. That is, the validity of the first order approach is not guaranteed. We will come back to this point shortly.

For the time being, we ignore this validity issue. With the first order approach, the principal's problem yields the following Lagrangian:

$$
\begin{aligned}
L \;=\; & \int_{x_l}^{x_h} U(x - S(x)) f(x,a) dx + \lambda \left\{ \int_{x_l}^{x_h} V(S) f(x,a) dx - D(a) - \theta \right\} \\
& + \mu \left\{ \int_{x_l}^{x_h} V(S(x)) f_a(x,a) dx - D'(a) \right\}
\end{aligned}
$$

With the pointwise maximization, the first order condition with respect to S is

$$\frac{U'(x - S(x))}{V'(S(x))} = \lambda + \mu \frac{f_a(x,a)}{f(x,a)}, \quad \forall x \in X,$$

where $\lambda \geq 0$. As compared with the first-best case, it can be seen that, roughly speaking, if $\mu \neq 0$, then the second-best solution is different, and the risk sharing between the principal and the agent becomes less efficient.

Theorem 2.1 *(Holmstrom, BJE 1979) Assume that $\lambda > 0$, $V'' < 0$, $F_a(x,a) < 0$ for all $x \in (x_l, x_h)$, and the first order approach is valid with an interior solution. Then $\mu > 0$, i.e. the second-best solution $\neq$ the first-best solution.*

Proof: Suppose not, i.e $\mu \leq 0$. ¿From the FOC with respect to a,

$$\int_{x_l}^{x_h} U(x - S(x))) f_a(x,a) dx + \mu \left\{ \int_{x_l}^{x_h} V(S(x)) f_{aa}(x,a) dx - D''(a) \right\} = 0.$$

By using the agent's second order condition in (2.2), the above equation implies that

$$\int_{x_l}^{x_h} U(x - S(x)) f_a(x,a) dx \leq 0. \tag{2.3}$$

However, the principal's FOC implies otherwise as can be seen below. Contradiction. Therefore $\mu > 0$.

To see this, define $\varphi(x,S)$ as follows:

$$\varphi(x,S) := \frac{U'(x - S)}{V'(S)}.$$

Then since U', $V' > 0$, $V'' < 0$, and $U' \leq 0$, φ is strictly increasing in S. Let us also define $S_\lambda(x)$ through the following equation:

$$\varphi(x, S_\lambda(x)) = \lambda.$$

Then it can be shown that $S'_\lambda(x) \in [0, 1)$.

Now recall that the principal's FOC with respect to S is

$$\frac{U'(x - S(x))}{V'(S(x))} = \lambda + \mu \frac{f_a(x,a)}{f(x,a)}, \quad \forall x \in X.$$

Comparing $S_\lambda(x)$ with $S(x)$, we have two cases: $\mu < 0$ and $\mu = 0$. When $\mu < 0$, note that since $\mu \leq 0$ by hypothesis, $S(x) \leq S_\lambda(x)$ if and only if $f_a(x,a) \geq 0$. Thus, (check yourself)

$$U(x - S(x)) f_a(x,a) \geq U(x - S_\lambda(x)) f_a(x,a), \quad \forall x \in X.$$

That is

$$\int_{x_l}^{x_h} U(x - S(x)) f_a(x,a)\, dx \geq \int_{x_l}^{x_h} U(x - S_\lambda(x)) f_a(x,a)\, dx$$

However

$$\begin{aligned} &\int_{x_l}^{x_h} U(x - S_\lambda(x)) f_a(x,a)\, dx \\ &= U(x - S_\lambda(x)) F_a(x,a)\big|_{x_l}^{x_h} \\ &\qquad - \int_{x_l}^{x_h} U'(x - S_\lambda(x))(1 - S'_\lambda(x)) F_a(x,a)\, dx > 0. \end{aligned}$$

The last inequality is from the fact that $F_a(x_l, a) = F_a(x_h, a) = 0$, $S'_\lambda(x) \in [0, 1)$, and $F_a(x,a) < 0$ for all $x \in (x_l, x_h)$. Therefore, it contradicts (2.3).

When $\mu = 0$, $S(x) \equiv S_\lambda(x)$ for all x. Thus we reach the same contradiction. □

Remark 1: Theorem 2.1 tells us that under a set of very reasonable conditions, moral hazard problems can be serious enough to result in losses in social welfare.

Remark 2: The assumption that the supports x_l and x_h are independent of a is important. If they shift as a changes, some outcomes can be definitely informative about effort and the principal can enforce the first best.

Remark 3: Note that f_a/f is the derivative of the log-likelihood function of a, $\ln f(x,a)$, given a realization of x.

2.4 The Shape of the Second-best Contract

The next question is: What should be the shape of the 2nd-best optimal sharing rule in general?

Definition 2.1 *The pdf f satisfies the* monotone likelihood ratio property *(MLRP) if and only if f_a/f is increasing in x.*

Theorem 2.2 *Assume that $\lambda > 0$, $V'' < 0$, $F_a(x,a) < 0$ for all $x \in (x_l, x_h)$, and the first order approach is valid with an interior solution. If f satisfies the MLRP, then $S'(x) > 0$, i.e. the higher the outcome, the higher the compensation.*

Proof: By redifferentiating the FOC with respect to x,

$$\frac{U''V' \cdot (1 - S'(x)) - U'V'' \cdot S'(x)}{V'^2} = \mu \frac{\partial}{\partial x}\left(\frac{f_a}{f}\right)$$

Or

$$(-U''V' - U'V'')S'(x) = \mu \frac{\partial}{\partial x}\left(\frac{f_a}{f}\right) V'^2 - U''V' > 0.$$

Therefore $S'(x) > 0$. □

Remark 1: Under the MLRP, a high value of x implies that a high value of a is more likely.

Remark 2: (See Stole [1997].) If the MLRP does not hold, sometimes the optimal contract may punish the agent for some high outcomes. Suppose that there are two possible levels of effort a_l and a_h, and three possible outcomes x_l, x_m, and x_h, and that the pdf's $f(x,a)$ are given as follows:

	x_l	x_m	x_h
a_h	0.4	0.1	0.5
a_l	0.5	0.4	0.1
f_a/f	–0.25	–3	0.8

In this case, if the principal wishes to induce the high effort a_h, it may be necessary to punish the agent when x_m is realized because x_m is most indicative of the low effort a_l.

Suppose that the principal is risk neutral and the agent is risk averse. Then the first best risk sharing rule would be as follows. The principal takes all the risk and the agent takes a riskfree share. That is, the first-best optimal contract is a complete insurance contract for the agent. However, with such a contract, the agent has no incentives to work. Therefore, to make the agent to work, the principal has to impose some risks to the agent. But then, the first best cannot be achieved. Therefore, the second-best contract has to be based on *a tradeoff between risk sharing and incentives.*

Exercise 2.1 *Suppose both the principal and agent are risk neutral. Explain why the second-best contract should be designed in such a way that the agent takes all the risks and the principal is given a fixed income.*

2.5 The Value of Informative Signal in Contracting

The value of informative signal in contracting was studied by Holmstrom [1979] and Shavell [1979].

As we have seen above, the shape of the contract depends on the information content of the outcome about the effort level. With the MLRP, a high outcome means a high effort level with a high probability.

What if there is another signal s about effort in addition to x? Will it be useful in contracting? First, recall the concept of sufficiency encountered in mathematical statistics.

Definition 2.2 *$T(x, s)$ is sufficient for $a \in A$ in the family of distributions given by $\{f(x, s, a), a \in A\}$, if the conditional distribution of (x, s) given $T(x, s)$ is independent of a.*

Lemma 2.1 *(Fisher-Neyman Factorization Theorem) $T(x, s)$ is sufficient for $a \in A$ in the family of distributions given by $\{f(x, s, a), a \in A\}$, if and only if the joint pdf of (x, s) is factorized as below:*

$$f(x, s, a) = g(T(x, s), a)h(x, s).$$

Definition 2.3 *An additional signal s is informative about effort if and only if x is not sufficient.*

In the absence of additional signals other than the performance metric x, the optimal salary scheme should satisfy

$$\frac{U'(x - S)}{V'(S)} = \lambda + \mu \frac{f_a(x, a)}{f(x, a)}, \quad \forall x \in X.$$

When there is an additional signal s about the agent's effort a, the optimal salary scheme solves

$$\frac{U'(x - S)}{V'(S)} = \lambda + \mu \frac{f_a(x, s, a)}{f(x, s, a)}, \quad \forall x \in X,$$

If f_a/f is independent of s, then the signal s is useless. If not, the signal in general is useful in improving the agency problem.

Consider two signals s and s'. Signal s' is uninformative if

$$f(x, s, s', a) = g(T(x, s), a)h(x, s, s').$$

In this case, f_a/f is independent of s', and $T(x, s)$ is a sufficient statistic. When there is a sufficient statistic $T(x, s)$, the optimal salary scheme depends only on $T(x, s)$ and nothing else.

2.5.1 Application to Insurance Deductibles

Suppose that the probability of an insured accident depends on the effort of the insured, but that conditional on an accident occuring, the size of the damage from the occured accident is independent of the effort of the insured. Then what should be a right form of insurance contract?

Let x be the size of an accident. Assume that $f(0,a) = 1 - p(a)$, $f(x,a) = p(a)g(x)$ for $x < 0$, and that $p'(a) < 0$. In this setup, the probability of an accident depends on effort, but the loss from it is completely independent of effort. Note that for $x \leq 0$,

$$\begin{aligned} f(x,a) &= \left\{(1-p(a))\chi_{\{x=0\}} + p(a)(1-\chi_{\{x=0\}})\right\} \\ &\quad \cdot \left\{\chi_{\{x=0\}} + g(x)(1-\chi_{\{x=0\}})\right\}, \end{aligned}$$

where $\chi_{\{x=0\}}$ is an indicator function for event $\{x = 0\}$. Thus $\chi_{\{x=0\}}$ is sufficient for a. Therefore, the optimal insurance contract in this case should only depend on $\chi_{\{x=0\}}$. To see this, note that $f_a/f = p'(a)/p(a) < 0$, for $x < 0$, independent of x, and $f_a/f = -p'(a)/(1-p(a)) > 0$ for $x = 0$, again independent of x. To find an optimal insurance contract, let us assume that the principal (the insurer) is risk neutral. The FOC is

$$\frac{1}{V'(S)} = \lambda + \mu \times \begin{cases} \frac{p'(a)}{(1-p(a))} & \text{if } x = 0 \\ \frac{p'(a)}{p(a)} & \text{if } x < 0 \end{cases}$$

Thus, the optimal contract will be of the following form:

$$S = \alpha\chi_{\{x=0\}} + \beta(1-\chi_{\{x=0\}}), \quad \alpha,\, \beta \in \mathcal{R}.$$

By rewriting, $S = \alpha + (\beta - \alpha)(1 - \chi_{\{x=0\}})$. α can be viewed as an insurance premium and $\beta - \alpha$ as a deductible amount conditional on an accident occuring. In other words, the insured (agent) pays to the insurance company (principal) $-\alpha$ as an insurance premium, and when an accident occurs, the agent pays to the insurance company an amount of $-(\beta-\alpha)$. (Note that α and $(\beta-\alpha)$ should be negative in reasonable settings. In fact, to make sense out of this insurance model, for $x < 0$, $x < \beta - \alpha$.)

2.6 The Validity of the First-order Approach

The first-order approach is valid if the agent's problem given a contract is concave in effort.

Definition 2.4 *A distribution function is said to have a convexity distribution function property (CDFC) if and only if* $F_{aa}(x, a) \geq 0$.

Remark 1: An interesting special case of the CDFC is the linear distribution function condition (LDFC): for $a \in (0, 1)$,

$$f(x, a) = a g_1(x) + (1 - a) g_2(x)$$

where $g_1(x)$ first-order stochastically dominates $g_2(x)$. That is, the pdf f is a convex combination of two distributions, one of which dominates the other in the sense of the first-order stochastic dominance.

Theorem 2.3 *(Rogerson [1985]) Assume that* $\lambda > 0$, $V'' < 0$, *and* $F_a(x, a) < 0$ *for all* $x \in (x_l, x_h)$. *The first-order approach is valid if the distribution function* $F(x, a)$ *satisfies the MLRP and CDFC.*

Heuristic Proof: Given S, the agent's expected utility is

$$\begin{aligned}
&\int_{x_l}^{x_h} V(S(x)) f(x, a) dx - D(a) \\
&= V(S(x)) F(x, a)|_{x_l}^{x_h} - \int_{x_l}^{x_h} V(S(x)) S'(x) F(x, a) dx - D(a) \\
&= V(S(x_h)) - \int_{x_l}^{x_h} V(S(x)) S'(x) F(x, a) dx - D(a).
\end{aligned}$$

We need to show that the agent's SOC is satisfied. But

$$-\int_{x_l}^{x_h} V(S(x)) S'(x) F_{aa}(x, a) dx - D''(a) \leq 0,$$

because $F_{aa}(x, a) \geq 0$ and $S'(x) < 0$ under MLRP and $\mu > 0$. □

The above heuristic proof is circular, because it makes use of $\mu > 0$, which holds under the assumption that the first-order approach is valid. To avoid this kind of circular argument, Rogerson proposes the "doubly relaxed principal's problem by replacing the first-order equality condition with an (artificial) inequality condition as follows:

Problem 2.3 *(Doubly Relaxed Principal's Problem) Choose S and a to*

$$\begin{aligned} \max \quad & \int_{x_l}^{x_h} U(x - S(x))f(x,a)dx \\ s.t. \quad & \int_{x_l}^{x_h} V(S(x))f(x,a)dx - D(a) \geq \theta \\ & \frac{\partial}{\partial a}\int_{x_l}^{x_h} V(S(x))f(x,a)dx - D'(a) \geq 0. \end{aligned}$$

The last constraint is weaker than the first-order condition with equality. If the problem yields a solution satisfying the last condition with inequality, then it does not solve the original principal's problem. If the solution satisfies the constraint with equality and the agent's SOC is satisfied, then one can say the solution solves the original problem. In both cases, Rogerson shows that the SOCs are satisfied.

Proof: The Lagrangian for the above doubly relaxed problem is

$$\begin{aligned} L &= \int_{x_l}^{x_h} U(x - S(x))f(x,a)dx + \lambda\left\{\int_{x_l}^{x_h} V(S)f(x,a)dx - D(a) - \theta\right\} \\ & \quad +\mu\left\{\frac{\partial}{\partial a}\int_{x_l}^{x_h} V(S(x))f(x,a)dx - D'(a)\right\}, \end{aligned}$$

and we now know that $\mu \geq 0$. If $\mu > 0$, then the constraint is binding and thus the agent's FOC has to be satisfied with equality. Suppose that $\mu = 0$. Then $\lambda > 0$. The principal's FOC with respect to S is

$$\frac{U'(x - S(x))}{V'(S(x))} = \lambda, \quad \forall x \in X.$$

Thus we have $S'(x) \geq 0$.

On the hand, let us rewrite the Lagrangian

$$\begin{aligned} L &= \int_{x_l}^{x_h} U(x - S(x))f(x,a)dx + \lambda\left\{\int_{x_l}^{x_h} V(S)f(x,a)dx - D(a) - \theta\right\} \\ &= U(x_h - S(x_h)) - \int_{x_l}^{x_h} U'(x - S(x))S'(x)F(x,a)dx \\ & \quad +\lambda\left\{\int_{x_l}^{x_h} V(S)f(x,a)dx - D(a) - \theta\right\}. \end{aligned}$$

Since $\lambda > 0$, the principal's FOC is

$$\begin{aligned}\frac{\partial L}{\partial a} &= -\int_{x_l}^{x_h} U'(x - S(x))S'(x)F_a(x,a)dx \\ &\quad +\lambda \frac{\partial}{\partial a}\left\{\int_{x_l}^{x_h} V(S)f(x,a)dx - D(a) - \theta\right\} = 0.\end{aligned}$$

Since $S'(x) > 0$ and $F_a(x,a) \leq 0$, the FOC implies that we must have

$$\frac{\partial}{\partial a}\left\{\int_{x_l}^{x_h} V(S)f(x,a)dx - D(a) - \theta\right\} \leq 0.$$

This relationship can be consistent with the last constraint of the doubly relaxed problem only if

$$\frac{\partial}{\partial a}\left\{\int_{x_l}^{x_h} V(S)f(x,a)dx - D(a) - \theta\right\} = 0.$$

Thus whether the last constraint is binding or not, the agent's FOC has to be satisfied with equality.

Now, looking at the local SOC of the agent's problem in (2.2),

$$\begin{aligned}&\int_{x_l}^{x_h} V(S(x))f_{aa}(x,a)dx - D''(a) \\ &= V(S(x))F_{aa}(x,a)|_{x_l}^{x_h} - \int_{x_l}^{x_h} V'(S(x))S'(x)F_{aa}(x,a)dx - D''(a) \\ &= -\int_{x_l}^{x_h} V'(S(x))S'(x)F_{aa}(x,a)dx - D''(a) \leq 0.\end{aligned}$$

The last inequality is satisfied by the MLRP and CDFC. (Recall $S'(x) > 0$ by the MLRP.) □

Remark 1: Note that the MLRP and CDFC conditions are so restrictive that popular distributions like normal distribution cannot satisfy this conditions.

Remark 2: The CDFC is particularly restrictive. Jewitt [1988] shows that under CARA utility assumption, one can avoid CDFC. Examples are

1. a Gamma distribution with a mean of αa such that $f(x,a) = a^{-\alpha}x^{\alpha-1}e^{-x/\alpha}\Gamma(\alpha)^{-1}$.
2. a Poisson distribution with a mean of a such that $f(x,a) = a^x e^{-a}/\Gamma(1+x)$.
3. a Chi-square distribution with a degrees of freedom such that $f(x,a) = \Gamma(2a)^{-1}2^{-2a}x^{2a-1}2^{-x/2}$.

2.7 Normally Distributed Outcome

What can happen to the optimal contract if the first-order approach cannot be justified? Mirrlees [1974] constructs a very interesting case.

Theorem 2.4 *(Mirrlees [1974]) Suppose $f(x,a)$ is a pdf form a normal distribution with a mean of zero and a variance of σ^2, and that the principal is risk neutral. Then the second best solution does not exist, and the first-best solution can be achieved arbitrarily closely.*

Proof: Note that

$$f(x,a) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad \frac{f_a(x,a)}{f(x,a)} = \frac{x-a}{\sigma^2}.$$

Let S^* and a^* be the first-best salary and effort, respectively. Since S^* is constant, $V(S^*) - D(a^*) = \theta$. Now consider a compensation scheme given in a following form of a step function:

$$S = \begin{cases} S^* & \text{if } x \geq \eta \\ \epsilon(\eta) & \text{if } x < \eta, \end{cases}$$

where $\epsilon(\eta) \in \mathcal{R}$ is determined to satisfy the agent's incentive constraint with effort level a^* such that given any $\eta < a^*$,

$$\int_{\eta}^{\infty} V(S^*)f_a(x,a^*)dx + \int_{-\infty}^{\eta} V(\epsilon)f_a(x,a^*)dx = D'(a^*).$$

That is

$$\{V(\epsilon) - V(S^*)\}\int_{-\infty}^{\eta} f_a(x,a^*)dx = D'(a^*).$$

Since $f_a(x, a^*) < 0$ for all $x < \eta$, we have for all $x < \eta$,

$$V(S^*) - V(\epsilon) > 0, \quad S^* > \epsilon.$$

Thus one can interprete ϵ as punishment for a bad outcome.

On the other hand, we claim that this step-function contract can also satisfy the participation constraint arbitrarily closely as η approaches $-\infty$. Given this contract, the agent's utility is

$$\int_{\eta}^{\infty} V(S^*) f(x, a^*) dx + \int_{-\infty}^{\eta} V(\epsilon) f(x, a^*) dx - D(a^*)$$
$$= \theta - \int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon)\} f(x, a^*)\, dx.$$

It suffices to show that $\int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon)\} f(x, a^*)\, dx$ approaches zero as $\eta \to -\infty$. Let

$$M(\eta) := \frac{f_a(\eta, a^*)}{f(\eta, a^*)}.$$

Since $\eta < a^*$, $M(\eta) < 0$; and since $M'(\eta) > 0$, for all $x < \eta$,

$$\frac{f_a(x, a^*)}{f(x, a^*)} \leq M(\eta) < 0, \quad \text{or} \quad \frac{1}{M(\eta)} f_a(x, a^*) > f(x, a^*).$$

Thus,

$$\int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon(\eta))\} f(x, a^*)\, dx$$
$$< \frac{1}{M(\eta)} \int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon(\eta))\} f_a(x, a^*)\, dx$$

The LHS of the above inequality is always greater than zero, approaching zero as $\eta \to 0$. However, by the incentive constraint and using the fact that $V(S^*) \int_{-\infty}^{\infty} f_a(x, a^*) dx = 0$, we have

$$\int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon)\} f_a(x, a^*) dx = D'(a^*).$$

Hence

$$0 < \int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon(\eta))\} f(x, a^*)\, dx < \frac{1}{M(\eta)} D'(a^*)$$

Since the RHS of the inequality in (2.4) approaches zero as $\eta \to -\infty$, and since $V(S^*) - V(\epsilon(\eta))$ is positive but bounded by $V(S^*)$, $\int_{-\infty}^{\eta} \{V(S^*) - V(\epsilon(\eta))\} f(x, a^*)\, dx$ converges to zero as $\eta \to -\infty$. □

Chapter 3

Moral Hazard and Incentive Contracts: a Continuous-Time Approach

As we have seen in the discrete-time model, we have made very little progress because of methodological obstacles. Now we assume that having signed a contract, the agent exerts effort continuously for a certain period. Then the solution sometimes becomes surprisingly simple. This chapter is based on Holmstrom and Milgrom [1987], Schättler and Sung [1993], and Sung [1990,1995].[1]

The continuous-time model is set up as follows: The time period of interest is a unit interval $[0, 1]$. Both the principal and agent exhibit constant absolute risk aversion (CARA). In particular, the principal and the agent utility functions are $-e^{-Rw}$ and $-e^{-rw}$, where R and r are the CARA coefficients for the principal and the agent respectively.

The filtered probability space is given by $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$. The outcome of the agent's effort is driven by a Brownian motion with a drift

[1]There are many other related continuous-time agency models. See Sung [1997] for the outcome driven by point (jump) processes; Sung [1991] and Müller [1998] for the first-best solutions; Ou-Yang [1999] for portfolio management applications; Arora and Ou-Yang [2000] for managerial career concerns; Detemple, Govindaraj and Loewenstein [2001] for a general class of utility functions; and Schättler and Sung [1997], Hellwig and Schmidt [1998], and Müller [2000] for relationships between discrete-time and continuous-time models.

as follows:

$$dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t, \quad Y_0 \in \mathcal{R},$$

where $\{B_t, \mathcal{F}_t\}$ is the standard Brownian motion (standard Wiener) process. In fact, we assume that $\mathcal{F}_t$ is the argumented sigma algebra generated by $\{B_t\}$ up to t, and that $\mathcal{F}_1 = \mathcal{F}$.

The agent exerts effort to control $\mu_t \in U$ and $\sigma_t \in \Sigma$ at an instantaneous monetary cost of $c(\mu_t, \sigma_t)$. Thus, the cumulative cost up to time t is $\int_0^t c(\mu_s, \sigma_s)ds$. Assume that U and Σ are bounded, open and positive real intervals; that c_μ is strictly increasing and convex in μ; and that f_μ is strictly increasing and concave in μ.

Control laws are $\mathcal{F}_t$-predictable mappings such that $(\boldsymbol{\mu}, \boldsymbol{\sigma}) : \Omega \times [0, t] \to \mathcal{R}^2$, and $\boldsymbol{\mu}(\omega, t) = \mu_t$, $\boldsymbol{\sigma}(\omega, t) = \sigma_t$. Admissible sets of control laws for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are denoted by $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$, respectively.

The principal can only observe the whole outcome process $\{Y_t\}$, but not $\{B_t\}$. (Therefore, from his computation of the quadratic variation of $\{Y_t\}$, the principal can infer $\{\sigma_t\}$ with an arbitrary precision.)

Major technical conditions: We assume $\boldsymbol{\sigma}$ satisfies a uniform Lipschitz condition in $Z, \bar{Z} \in C[0, 1]$: There exists a constant K such that

$$|\boldsymbol{\sigma}(t, Z) - \boldsymbol{\sigma}(t, \bar{Z})| \leq K \sup_{0 \leq s \leq t} \|Z(s) - \bar{Z}(s)\|$$

We also assume that the admissible salary function S is an $\mathcal{F}_1$-measurable random variable of the following structure:

$$S(Y) = F(Y) + \int_0^1 \alpha(t, Y)dt + \int_0^1 \beta(t, Y)dY_t,$$

where α and β are bounded $\mathcal{F}_t$-predictable processes, and F is an $\mathcal{F}_1$-measurable random variable that is bounded below.

All other aspects of the model are the same as those of the previous discrete-time model.

3.1 The First-best Contract

For the sake of simplicity, we assume that the principal is risk neutral. The principal's problem is

Problem 3.1 *Choose control laws for the drift and diffusion rates* $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ *and a compensation scheme* S *to maximize*

$$
\begin{aligned}
& E\left[Y_1 - S\right] \\
& s.t.\ 1)\, dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t \\
& \qquad 2)\, E\left[-\exp\left\{-r\left(S - \int_0^1 c(\mu_t, \sigma_t)dt\right)\right\}\right] \geq -\exp\left\{-r\mathcal{W}_0\right\}.
\end{aligned}
$$

Since investors prefer more to less, then the second constraint should hold with equality. For the above problem, we have the following result.

Theorem 3.1 *Let* (μ^*, σ^*) *be a point which maximizes* $f(\mu, \sigma) - c(\mu, \sigma)$ *over* $U \times \Sigma$*. Then investors will have the manager choose* (μ^*, σ^*) *for all* $t \in [0, 1]$*, i.e.* $(\boldsymbol{\mu}, \boldsymbol{\sigma}) = (\mu^*, \sigma^*)$*, and pay to the agent a constant amount of salary* S^* *where* $S^* = \mathcal{W}_0 + c(\mu^*, \sigma^*)$*.*

Remark 1: If the maximum of $f(\mu, \sigma) - c(\mu, \sigma)$ is attained at an interior point (μ^*, σ^*), then

$$
\begin{aligned}
c_\mu(\mu^*, \sigma^*) &= f_\mu(\mu^*, \sigma^*), \quad \text{and} && (3.1) \\
c_\sigma(\mu^*, \sigma^*) &= f_\sigma(\mu^*, \sigma^*). && (3.2)
\end{aligned}
$$

Remark 2: Note that condition (3.1) implies that the optimal level of the agent's effort is determined so that the marginal cost of effort is equal to marginal increase in the mean of the output. Conditions (3.1) and (3.2) indicate that if the managerial control of σ_t is costly, then investors have the manager put the minimum possible effort (or no effort) on the idiosyncratic variance. If it is not costly, then $c_\sigma \equiv 0$ for any σ, and so the investors' wealth is not affected by the idiosyncratic variance. That is, the first best contract instructs the manager to control σ to maximize the mean of the output.

Proof: Define $Z := S - \int_0^1 c(\mu_t, \sigma_t)dt$. Then investors' problem becomes:

$$
\begin{aligned}
\max_{Z, (\mu_t, \sigma_t) \in U \times \Sigma} \quad & E\left[Y_1 - Z - \int_0^1 c(\mu_t, \sigma_t)dt\right] \\
\text{s.t.} \quad & dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t \\
& E\left[-e^{-rZ}\right] = -e^{-r\mathcal{W}_0}.
\end{aligned}
$$

This can be broken down into following two independent maximization problems:

$$\max_{(\mu_t,\sigma_t)\in U\times\Sigma} E\left[Y_1 - \int_0^1 c(\mu_t,\sigma_t)dt\right]$$
$$\text{s.t.}\;\; dY_t = f(\mu_t,\sigma_t)dt + \sigma_t dB_t$$

and

$$\max_{Z} \quad E[-Z]$$
$$\text{s.t.}\;\; E\left[-e^{-rZ}\right] = -e^{-r\mathcal{W}_0}.$$

¿From the constraint of the second maximization problem and by using the Jensen's inequality, we have

$$-e^{-r\mathcal{W}_0} = E\left[-e^{-rZ}\right] \leq -e^{-rE[Z]}.$$

This implies that $E[-Z] \leq -\mathcal{W}_0$. It is always possible to choose choosing Z such that $E[Z] = \mathcal{W}_0$. Therefore at the optimum we have

$$-e^{-r\mathcal{W}_0} = E\left[-e^{-rZ}\right] = -e^{-rE[Z]}. \tag{3.3}$$

But this relationship can hold if and only if $Z = \mathcal{W}_0$ almost everywhere. 'If' part is obvious. For the 'only if' part, suppose that Z is not a constant a.e. Define $\varphi := Z - E[Z]$, $A := \{\omega \in \Omega \,|\, Z - E[Z] > 0\}$, $B := \{\omega \in \Omega \,|\, Z - E[Z] = 0\}$, and $C := \{\omega \in \Omega \,|\, Z - E[Z] < 0\}$. Then $E\left[-e^{-r\varphi}\right] = -1$ by (3.3), and $E[\varphi\chi_A] = -E[\varphi\chi_C]$, where χ is the indicator function. Furthermore $E[\varphi\chi_A] \neq 0$ because Z is not constant a.e. by hypothesis. But

$$\begin{aligned}
-1 = E[-e^{-r\varphi}] &= \int_A -e^{-r\varphi}dP + \int_B -e^{-r\varphi}dP + \int_C -e^{-r\varphi}dP \\
&= \int_\Omega -e^{-r\varphi\chi_A}dP + \int_\Omega -e^{-r\varphi\chi_C}dP + 1 \\
&\leq -e^{-rE[\varphi\chi_A]} - e^{rE[\varphi\chi_A]} + 1.
\end{aligned}$$

For the inequality above, we used Jensen's inequality and the fact that $E[\varphi\chi_A] = -E[\varphi\chi_C]$. This implies that

$$-1 \leq (\frac{1}{2})\left\{-e^{-rE[\varphi\chi_A]} - e^{rE[\varphi\chi_A]}\right\}.$$

But we know the right hand side of the above inequality is strictly less than -1 because $-e^{-rx} < rx - 1$ when $x \neq 0$. Therefore if Z is not a constant a.e., then (3.3) cannot hold.

Since $Z = \mathcal{W}_0$ a.e., the first best contract S can be represented in the following form:

$$S = \mathcal{W}_0 + \int_0^1 c(\mu_t, \sigma_t)dt.$$

Furthermore the investors' problem can be reduced to the following:

$$\begin{array}{cl} \max\limits_{(\mu_t,\sigma_t)\in U\times\Sigma} & E\left[Y_1 - \mathcal{W}_0 - \int_0^1 c(\mu_t, \sigma_t)dt\right] \\ \text{s.t.} & dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t. \end{array}$$

Now this is a standard stochastic control problem. For this particular case, the principal's problem is

$$\begin{array}{cl} \max\limits_{(\mu_t,\sigma_t)\in U\times\Sigma} & E\left[\int_0^1 \left(f(\mu_t, \sigma_t) - \mathcal{W}_0 - c(\mu_t, \sigma_t)\right) dt\right] \\ \text{s.t.} & dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t. \end{array}$$

Thus, for $t \in [0, 1]$ a.e., $(\mu_t^*, \sigma_t^*) \in \arg\max_{(\mu,\sigma)\in U\times\Sigma}\{f(\mu, \sigma) - c(\mu, \sigma)\}$.
□

3.2 The Second-best Contract

The principal's problem is stated as follows:

Problem 3.2 *Choose control laws for the drift and diffusion rates* $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ *and a compensation scheme* S *to maximize*

$$E\left[Y_1 - S\right]$$

s.t. 1) $dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t$

2) *Given* S *and* $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$,

$$\boldsymbol{\mu} \in \arg\left\{\begin{array}{rl} \max_{\boldsymbol{\mu}'\in\boldsymbol{U}} & E\left[-\exp\left\{-r\left(S - \int_0^1 c(\mu_t', \sigma_t)dt\right)\right\}\right] \\ s.t. & dY_t = f(\mu_t', \sigma_t)dt + \sigma_t dB_t \end{array}\right.$$

3) $E\left[-\exp\left\{-r\left(S - \int_0^1 c(\mu_t, \sigma_t)dt\right)\right\}\right] \geq -\exp\left\{-r\mathcal{W}_0\right\}.$

Condition 2) is for the incentive compatibility, and 3) for the individual rationality.

Let us first look at the agent's problem.

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\boldsymbol{U}} \quad & E\left[-\exp\left\{-r\left(S-\int_0^1 c(\mu_t,\sigma_t)dt\right)\right\}\right] \\ \text{s.t.} \quad & dY_t = f(\mu_t,\sigma_t)dt + \sigma_t dB_t \end{aligned}$$

With an admissible salary function, the agent problem is

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\boldsymbol{U}} \quad & E\left[-\exp\left\{-r\left(F(Y)+\int_0^1 \alpha(t,Y)dt + \int_0^1 \beta(t,Y)dY_t \right.\right.\right. \\ & \left.\left.\left. -\int_0^1 c(\mu_t,\sigma_t)dt\right)\right\}\right] \\ \text{s.t.} \quad & dY_t = f(\mu_t,\sigma_t)dt + \sigma_t dB_t \end{aligned}$$

This is a stochastic control problem. For simplicity, we assume $F(Y) = F(Y_1)$, $\alpha(t,Y) = \alpha(t,Y_t)$, and $\beta(t,Y) = \beta(t,Y_t)$. (See Schättler and Sung [1993] for more general treatment.) For further analysis, we need dynamic programming equations and their verification theorems, all of which are derived in the Appendix.

Theorem 3.2 *(Holmstrom and Milgrom [1987], Schättler and Sung [1993]) Given an admissible $\boldsymbol{\sigma}\in\boldsymbol{\Sigma}$, an admissible salary function is necessarily represented as follows: $S(Y) = R(\boldsymbol{\mu},\boldsymbol{\sigma})$, where*

$$\begin{aligned} R(\boldsymbol{\mu},\boldsymbol{\sigma}) \;=\; & \mathcal{W}_0 + \int_0^1 c(\mu_t,\sigma_t)dt + \frac{r}{2}\int_0^1 \frac{c_\mu^2(\mu_t,\sigma_t)}{f_\mu^2(\mu_t,\sigma_t)}\sigma_t^2 dt \\ & + \int_0^1 \frac{c_\mu(\mu_t,\sigma_t)}{f_\mu(\mu_t,\sigma_t)}dY_t - \int_0^1 \frac{c_\mu(\mu_t,\sigma_t)}{f_\mu(\mu_t,\sigma_t)} f(\mu_t,\sigma)dt. \qquad (3.4) \end{aligned}$$

Remark 1: Eq. (3.4) reveals that the structure of an optimal second-best contract consists of four parts: (1) the reservation utility determined in the labor market; (2) the cost of effort; (3) the compensation risk imposed on the agent in order to give him incentives to work; and (4) the risk premium to be paid to the agent for imposing the compensation risk.

A Simplified Proof: (Here we provide a shortcut version of the proof under much more restrictive assumptions than the theorem needs. For a full proof, see Schättler and Sung [1993].) For simplicity, we assume $F(Y) = F(Y_1)$. Given an admissible salary function S, let us define the agent's value function V such that

$$\begin{aligned} V(t,.) \;:=\; & E\left[-\exp\left\{-r\left(S - \int_t^1 c(\mu_s^*, \sigma_s)ds\right)\right\} \,\Big|\, \mathcal{F}_t\right] \\ &= E\Bigg[-\exp\Bigg\{-r\Bigg(F(Y_1) \\ &\quad + \int_t^1 \left(\alpha(s,Y_s) + \beta(s,Y_s)f(\mu_s^*,\sigma_s) - c(\mu_s^*,\sigma_s)\right) ds \\ &\quad\quad + \int_t^1 \beta(s,Y_s)\sigma_t dB_s\Bigg)\Bigg\} \,\Bigg|\, \mathcal{F}_t\Bigg] \end{aligned}$$

where $\{\mu_s^*, t \le s \le 1\}$ are optimal controls over the remaining time period $[t,1]$. Let us further assume that $V(t,.) = V(t,Y_t)$ and $V(t,Y_t)$ satisfies the assumptions of A.1. (Note that we are assuming our filtration $\{\mathcal{F}_t\}$ is Markovian.) Then by Lemma A.1, the agent's dynamic programming equation is given as follows:

$$\begin{aligned} 0 \;\equiv\; & \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}\sigma^2 \\ & + \max_{\mu}\Bigg\{\frac{\partial V}{\partial y}(f(\mu,\sigma) - r\beta\sigma^2) \\ & \quad + r\left(c(\mu_t,\sigma_t) - \alpha - \beta f(\mu_t,\sigma_t) + \frac{r}{2}\beta^2\sigma^2\right)V\Bigg\} \end{aligned}$$

with the terminal condition $V(1,Y_1) = -e^{-rF(Y_1)}$. The first order condition for μ is

$$\frac{\partial V}{\partial y}f_\mu + rV(c_\mu - \beta f_\mu) = 0 \tag{3.5}$$

Substituting the first-order condition back into the dynamic programming equation, we obtain the following:

$$\begin{aligned} 0 \;\equiv\; & \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}\sigma^2 \\ & + rV\left\{-\left(\frac{c_\mu}{f_\mu} - \beta\right)(f(\mu,\sigma) - r\beta\sigma^2) + c(\mu) - \alpha - \beta f + \frac{r}{2}\beta^2\sigma^2\right\} \end{aligned}$$

On the other hand, let us define the certainty equivalent wealth process $\mathcal{W}_t$ as follows:

$$\mathcal{W}_t = -\frac{1}{r}\log(-V(t, Y_t)).$$

Then we have by using the Itô's rule,

$$\mathcal{W}_\tau = \mathcal{W}_0 - \int_0^\tau \frac{1}{rV} dV + \frac{1}{2}\int_0^\tau \frac{1}{rV^2} d\langle V, V\rangle_t.$$

But

$$\begin{aligned} dV &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial y}dY_t + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}\sigma^2 dt \\ &= \left[\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}\sigma^2\right]dt + \frac{\partial V}{\partial y}dY_t \\ &= rV\left\{\left(\frac{c_\mu}{f_\mu} - \beta\right)(f(\mu,\sigma) - r\beta\sigma^2) - c(\mu) + \alpha + \beta f - \frac{r}{2}\beta^2\sigma^2\right\} dt \\ &\qquad -rV\left(\frac{c_\mu}{f_\mu} - \beta\right) dY_t \end{aligned}$$

Thus,

$$d\langle V, V\rangle_t = r^2 V^2 \left(\frac{c_\mu}{f_\mu} - \beta\right)^2 \sigma^2 dt$$

Therefore,

$$\begin{aligned} &\mathcal{W}_\tau - \mathcal{W}_0 \\ &= -\int_0^\tau \left\{\left(\frac{c_\mu}{f_\mu} - \beta\right)(f(\mu,\sigma) - r\beta\sigma^2) - c(\mu) + \alpha + \beta f - \frac{r}{2}\beta^2\sigma^2\right\} dt \\ &\quad + \int_0^\tau \left(\frac{c_\mu}{f_\mu} - \beta\right) dY_t + \frac{r}{2}\int_0^\tau \left(\frac{c_\mu}{f_\mu} - \beta\right)^2 \sigma^2 dt \end{aligned} \tag{3.6}$$

Note that

$$S = F(Y_1) + \int_0^1 \alpha dt + \int_0^1 \beta dY_t$$

and that by construction, $\mathcal{W}_1 = F(Y_1)$. However, rearranging Eq.(3.6), we have

$$F(Y_1) + \int_0^1 \alpha dt + \int_0^1 \beta dY_t$$

$$= \mathcal{W}_0 - \int_0^1 \left\{ \left(\frac{c_\mu}{f_\mu} - \beta \right) (f(\mu, \sigma) - r\beta\sigma^2) - c(\mu) + \beta f - \frac{r}{2}\beta^2\sigma^2 \right\} dt$$
$$+ \int_0^1 \frac{c_\mu}{f_\mu} dY_t + \frac{r}{2} \int_0^\tau \left(\frac{c_\mu}{f_\mu} - \beta \right)^2 \sigma^2 dt$$
$$= \mathcal{W}_0 - \int_0^1 \left\{ \frac{c_\mu}{f_\mu} f(\mu, \sigma) - c(\mu) - \frac{r}{2}\frac{c_\mu^2}{f_\mu^2}\sigma^2 \right\} dt + \int_0^1 \frac{c_\mu}{f_\mu} dY_t.$$

This proves the statement of the theorem. □

The next theorem establishes the validity of the first order approach. Before that, we need the definition of implementability.

Definition 3.1 *We call an admissible control law* $\boldsymbol{\mu}$ *"implementable" if given a salary function* $S[\boldsymbol{\mu}]$ *in the form of (3.4), then* $\boldsymbol{\mu}$ *is an optimal control law for the agent's problem.*

Theorem 3.3 *(Schättler and Sung [1993], Implementability Theorem) An admissible control law* $\boldsymbol{\mu}$ *can be implemented by the salary function represented by (3.4) if and only if for almost every* $(\omega, t) \in \Omega \times [0, 1]$, μ_t *minimizes over* U *the function* $h(\mu', \sigma_t; \mu_t)$, *i.e.,*

$$\mu_t \in \arg\min_{\mu' \in U} h(\mu', \sigma_t; \mu_t) := c(\mu', \sigma_t) - \frac{c_\mu(\mu_t, \sigma_t)}{f_\mu(\mu_t, \sigma_t)} f(\mu', \sigma_t). \tag{3.7}$$

Proof: Suppose that the salary function S given to the agent in the form of (3.4) with $\{\hat{\mu}_t; 0 \le t \le 1\}$. Then the agent's expected utility is

$$E\left[- \exp\left\{ -r \left(\mathcal{W}_0 \right.\right.\right.$$
$$+ \int_0^1 \{\hat{c} - \frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f} - c(\mu_t, \sigma_t) + \frac{\hat{c}_\mu}{\hat{f}_\mu} f(\mu_t, \sigma_t) + \frac{r}{2}\frac{\hat{c}_\mu^2}{\hat{f}_\mu^2}\sigma^2\} dt$$
$$\left.\left.\left. + \int_0^1 \frac{\hat{c}_\mu}{\hat{f}_\mu}\sigma_t \, dB_t \right) \right\} \right],$$

where the hat on the top of each function is to indicate the omittion of its argument $(\hat{\mu}_t, \sigma_t)$. The dynamic programming equation is:

$$\begin{aligned} 0 \quad \equiv \quad & \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}\sigma^2 + \max_{\mu}\left\{ \frac{\partial V}{\partial y}(f(\mu,\sigma) - r\sigma^2 \frac{\hat{c}_\mu}{\hat{f}_\mu}) \right. \\ & \left. + rV\left(c(\mu_t,\sigma_t) - \frac{\hat{c}_\mu}{\hat{f}_\mu} f(\mu_t,\sigma_t) - \hat{c} + \frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f} \right) \right\} \qquad (3.8) \end{aligned}$$

with the terminal condition being $V(1,.) = -e^{-r\mathcal{W}_0}$. Consider a function V as follows:

$$V_t = -e^{-r\mathcal{W}_0}.$$

Then V satisfies the dynamic programming equation because $\hat{\mu}_t \in \arg\max\{c(\hat{\mu}_t, \sigma_t) - \frac{\hat{c}_\mu}{\hat{f}_\mu} f(\hat{\mu}_t, \sigma_t)\}$. Therefore, by the verification theorem given in Lemma A.2, when a salary function S with $\{\hat{\mu}_t; 0 \le t \le 1\}$ is assigned, the agent optimally choose $\{\hat{\mu}_t; 0 \le t \le 1\}$. This proves the "if" part of the theorem.
For "only if" part, suppose that $\{\hat{\mu}_t; 0 \le t \le 1\}$ is implementable. That is, the agent's optimal controls are $\{\hat{\mu}_t; 0 \le t \le 1\}$. Thus, the agent optimal expected utility is

$$E\left[-\exp\left\{ -r\left(\mathcal{W}_0 + \frac{r}{2}\int_0^1 \frac{\hat{c}_\mu^2}{\hat{f}_\mu^2}\sigma^2\, dt + \int_0^1 \frac{\hat{c}_\mu}{\hat{f}_\mu}\sigma_t\, dB_t \right) \right\} \right] = -e^{-r\mathcal{W}_0}.$$

The last equality is by the Novikov theorem. Therefore, with implementable controls, the agent's value function is $-e^{-r\mathcal{W}_0}$. Since this function satisfies all assumptions of Lemma A.1, it should satisfy the dynamic programming equation in (3.8). Hence we must have

$$0 \equiv \max_{\mu}\left(c(\mu_t,\sigma_t) - \frac{\hat{c}_\mu}{\hat{f}_\mu} f(\mu_t,\sigma_t) - \hat{c} + \frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f} \right).$$

Thus we have $\hat{\mu}_t \in \arg\min\{c(\mu'_t, \sigma_t) - \frac{\hat{c}_\mu}{\hat{f}_\mu} f(\mu'_t, \sigma_t)\}$. □

Theorem 3.3 suggests that the implementable control set $\hat{U}$ is given by

$$\hat{U} = \left\{ \hat{\mu} \in U \;\middle|\; \hat{\mu} \in \arg\min_{\mu' \in U} c(\mu', \sigma) - \frac{c_\mu(\hat{\mu},\sigma)}{f_\mu(\hat{\mu},\sigma)} f(\mu', \sigma) \right\}. \qquad (3.9)$$

It can be shown that investors only need to consider implementable controls in choosing a salary function of the form as in (3.4). The reason is that investors do not gain by considering other admissible salary functions than the ones in (3.4) with implementable controls chosen from (3.9). See Sung [1997, Proposition 1].

Corollary 3.1 *If f is linear and c is increasing and convex, then $\hat{U} = U$.*

Under the assumptions of Corollary 3.1, all controls in U can be implemented by the represented salary function $R(\boldsymbol{\mu}, \boldsymbol{\sigma})$, and thus the issue of implementability can be safely ignored as long as we choose a contract from the family of salary functions in the form of (3.4). ¿From now on, we assume these assumptions hold. Then the principal's problem can now be equivalently restated as follows:

Problem 3.3 *Choose control laws for the drift and diffusion rates $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ to maximize*

$$
\begin{aligned}
&E\Bigg[Y_1 - \mathcal{W}_0 - \int_0^1 c(\mu_t, \sigma_t)dt - \frac{r}{2}\int_0^1 \frac{c_\mu^2(\mu_t, \sigma_t)}{f_\mu^2(\mu_t, \sigma_t)}\sigma_t^2 dt \\
&\qquad - \int_0^1 \frac{c_\mu(\mu_t, \sigma_t)}{f_\mu(\mu_t, \sigma_t)}dY_t + \int_0^1 \frac{c_\mu(\mu_t, \sigma_t)}{f_\mu(\mu_t, \sigma_t)} f(\mu_t, \sigma)dt\Bigg] \\
&s.t.\ \ dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t
\end{aligned}
$$

Or the principal's problem is to choose control laws for the drift and diffusion rates $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ to maximize

$$
\begin{aligned}
&E\left[Y_1 - \mathcal{W}_0 - \int_0^1 c(\mu_t, \sigma_t)dt - \frac{r}{2}\int_0^1 \frac{c_\mu^2(\mu_t, \sigma_t)}{f_\mu^2(\mu_t, \sigma_t)}\sigma_t^2 dt\right] \\
&\text{s.t.}\ \ dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t
\end{aligned}
$$

Clearly, the optimal controls are for $t \in [0, 1]$ a.e.,

$$
(\mu_t, \sigma_t) \in \underset{(\mu_t', \sigma_t') \in U \times \Sigma}{\arg\max} \left(f(\mu_t', \sigma_t') - c(\mu_t', \sigma_t') - \frac{r}{2}\frac{c_\mu^2(\mu_t', \sigma_t')}{f_\mu^2(\mu_t', \sigma_t')}{\sigma_t'}^2 \right).
$$

Thus the optimal control pair is constant over time. Therefore the optimal contract is given by

$$\begin{aligned} S &= \mathcal{W}_0 + c(\mu^*, \sigma^*) + \frac{r}{2}\frac{c_\mu^2(\mu^*, \sigma^*)}{f_\mu^2(\mu^*, \sigma^*)}\sigma^{*2} \\ &\quad + \frac{c_\mu(\mu^*, \sigma^*)}{f_\mu(\mu^*, \sigma^*)}(Y_1 - Y_0) - \frac{c_\mu(\mu^*, \sigma^*)}{f_\mu(\mu^*, \sigma^*)}f(\mu^*, \sigma^*). \end{aligned} \tag{3.10}$$

Note that in equilibrium, Y_1 is normally distributed. One may wonder why the linear contract is optimal, whereas it is clearly suboptimal in the analogous discrete-time setting with normally distributed outcome. It is particularly puzzling because we know that with a normally distributed outcome, the first-best solution is achievable in the limit by using the Mirrlees forcing solution.

An intuitive reason is as follows: In a continuous-time framework, the Mirrlees forcing solution (to his discrete-time model) does not work. For his solution is not implementable in continuous time. Consider a Mirrlees penalty-reward contract that implements the first-best solution in discrete time. Suppose that the same contract is given to the agent in continuous time. Then the agent will not exert the first-best level of effort over time. Instead, when an intermediate outcome is deeply into the penalty zone, the agent loses the hope of bringing the outcome process back to the reward zone, and thus abandons making any effort at all. The same kind of agent reaction will occur when the intermediate outcome is deeply into the reward zone. The agent may find no effort is necessary because even with zero effort, it is highly unlikely that the final outcome will drop down into the penalty zone.

3.3 Applications

This section is based on Sung [1995].

We examine problems of project selection under moral hazard in a widely publicly held corporation. Assume that investors (the principal) are well diversified and thus exhibit risk-neutrality over idiosyncratic risks of the firm. The manager (the agent) however exhibits risk-aversion over the firm-specific risks (perhaps because he has invested his nondiversifiable human capital in the firm and the income generated from the capital takes up a significant portion of his total wealth).

The manager performs two types of functions: exerting effort to increase the profit from ongoing operations, and making a project choice decision given a project opportunity set. The managerial effort is to control $\{\mu_t\}$ the drift rate of the profit of ongoing operations at an instantaneous cost of controlling the drift rate is $c(\mu)$.

The managerial effort is unobservable to investors, but his project decision may or may not be observable. To capture the spirit of managerial decision making, we assume that the manager's project selection is costless to the manager. Further we assume that the marginal profit contribution of a project to the final outcome cannot be observed separately.

We assume that both investors and the manager know that the firm has investment opportunities. The efficient frontier of the opportunities is represented in idiosyncratic risk-return space by $g(\sigma)$, a real valued function of σ. A project is represented by $(g(\sigma), \sigma)$. If the manager chooses σ', his choice is equivalent to the choice of project $(f(\sigma'), \sigma')$.

The outcome of the managerial effort and project decision is realized as follows:

$$dY_t = (\mu_t + g(\sigma_t))dt + (\sigma_t + b)dB_t,$$

where b represent the risk level of ongoing operations without taking on an additional project.

Give the above model setup, the principal's problem can be stated as follows:

$$\begin{aligned} &\max\ E\left[Y_1 - S\right] \\ &\text{s.t. }\ dY_t = \{\mu_t + g(\sigma_t)\}dt + (\sigma_t + b)dB_t, \end{aligned} \tag{3.11}$$

where S is represented as in (3.4), i.e.,

$$\begin{aligned} S(Y) \;=\; & \mathcal{W}_0 + \int_0^1 c(\mu_t)dt + \frac{r}{2}\int_0^1 c_\mu^2(\mu_t)(\sigma_t + b)^2 dt + \int_0^1 c_\mu(\mu_t)dY_t \\ & - \int_0^1 c_\mu(\mu_t)(\mu_t + g(\sigma_t))dt. \end{aligned}$$

The optimal controls are for $t \in [0,1]$ a.e.,

$$(\mu_t, \sigma_t) \in \underset{(\mu_t', \sigma_t') \in U \times \Sigma}{\arg\max} \left(\mu_t' + g(\sigma_t') - c(\mu_t') - \frac{r}{2}c_\mu^2(\mu_t')(\sigma_t' + b)^2\right).$$

Thus the optimal controls are constant over time, and the FOCs are

$$1 - c_\mu(\mu) - rc_\mu(\mu)c_{\mu\mu}(\mu)(\sigma + b)^2 = 0$$
$$g'(\sigma) - rc_\mu^2(\mu)(\sigma + b) = 0.$$

Remark 1: Recall from Proposition (3.1) that $c_\mu(\mu) = 1$ in the first-best case. As compared with the first best, the above FOC suggests that in the second best, $c_\mu(\mu) < 1$ in general. That is, the manager in the second best does not work so hard as in the first best.

Suppose that the cost function is given as $c(\mu) = (K/2)\mu^2$, where K is a constant that depends on managerial ability; the more competent the manager, the lower the value of K. Therefore, the FOC suggests that the sensitivity of the contract c_μ as K decreases. In other words, a manager with high ability will be given a high-powered incentive contract. To be more specific, by the FOC, we have

$$\mu^* = \frac{1}{1 + rK(\sigma + b)^2}.$$

Holding σ constant, μ^* increases as r or K decreases.

The second FOC basically tells us about the project selection rule. A direct interpretation of the second FOC suggests that the project is selected in such a way that the marginal NPV improvement along the frontier is equal to the marginal increase in the risk premium of the managerial compensation contract.

For an additional insight, Fig 3.1 is constructed based on the second FOC. The figure illustrates how an optimal project is determined when a contract S is given with sensitivity $c_\mu(\mu)$. Since the sensitivity is higher for the manager with high ability, the slope of the straight line in the lower graph of Fig 3.1 is steeper, and thus the second-best project for the high-ability manager is less risky than that for a manager with lower ability. That is, the manager with high ability chooses less risky projects at the expense of lower NPVs.

In sum, when a high powered incentive contract is given, the manager works harder, but chooses less risky project at the expense of lower NPVs. This suggests that given a project opportunity $g(\sigma)$, the sensitivity of the managerial contract is based on a tradeoff between the benefit of incentives in managerial effort and the cost of disincentives of choosing risky projects.

The sensitivity is also affected by the shape of $g(\sigma)$. The flatter the frontier $g(\sigma)$, the lower the sensitivity. One may impagine that the frontier becomes flatter when the firm has fewer profitable project opportunities. In the U.S., empirical evidence by Jensen and Murphy [1990] for large publicly held corporations indicates that CEO wealth changes \$3.25 for every \$1,000 change in shareholder wealth. That is, the sensitivity of the contract is 0.00325 on average. Why low sensitivities? In theory, low sensitivities can occur because of high risk aversion, high idiosyncratic variance of ongoing operations, high marginal cost of managerial effort, or an efficient frontier with fewer profitable project opportunities.

3.3.1 Unobservable Project Selection

What if the principal cannot observe the manager's project selection? In this case, the managerial project selection rule becomes

$$g'(\sigma) - rc_\mu(\mu)(\sigma + b) = 0.$$

It can be shown that the principal's problem reduces to

$$\begin{aligned} \max_{\mu,\sigma} \quad & \mu + g(\sigma) - c(\mu) - \frac{1}{2}c_\mu^2(\mu)(\sigma + b)^2 \\ \text{s.t.} \quad & g'(\sigma) - rc_\mu(\mu)(\sigma + b) = 0. \end{aligned}$$

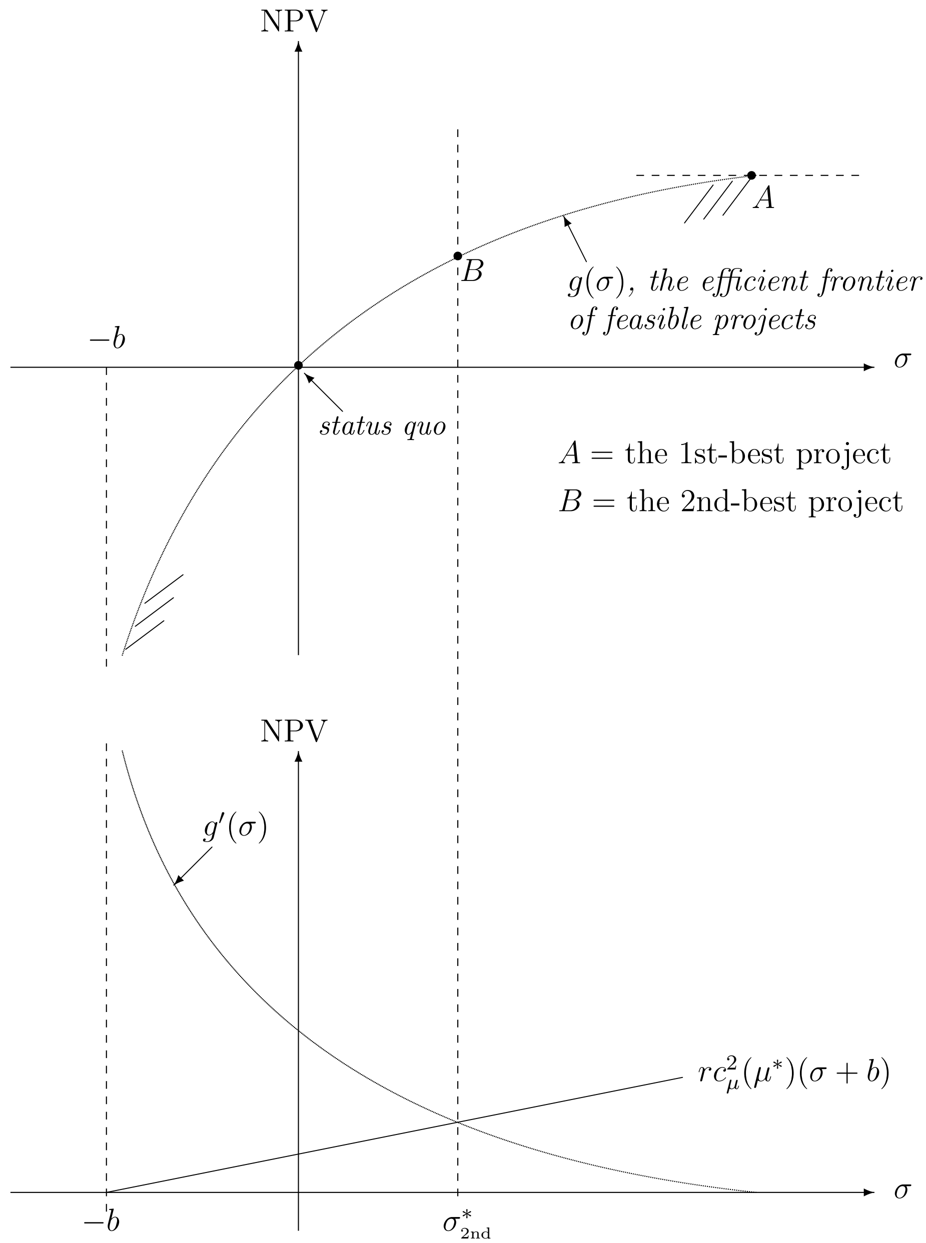


Figure 3.1: Project selection rules under moral hazard

Chapter 4

Adverse Selection and Moral Hazard

Next obvious questions are: (1) What if the managerial ability cannot be observed by the principal at the time of contract?; and (2) what if after the contract is signed, the manager knows much more than the principal about the firm?

In the literature, the first type is called the "adverse selection." Although the second problem is classified differently from the first type, from a mathematical perspective, the two problems are similar to each other.

To be specific, consider two types of managers; one with high ability and the other with low ability. For example, the cost of managerial effort may be given by $(\theta/2)\mu^2$. Different managers have different values in θ; θ_h for the high type and K_l for the low type, where $\theta_h < \theta_l$. Thus θ represents the managerial ability. Although individual managers know their own θ values, the principal (investors) cannot observe these values.

Ideally, the principal wish to write different contracts for different managers: If he can, he wants to assign $S(Y, \theta_h)$ for the high type and $S(Y, \theta_l)$ for the low type, where Y is the observable outcome of managerial effort. However, since he cannot observe the type, the principal has to rely on the manager's report on the manager's own type. That is, the contract has to depend on the manager's reported type as well as the outcome of effort. Thus, we write the contract as $S(Y, R)$, where R is the managerial report.

Knowing that the principal cannot observe the type, the manager has incentives to strategically send the principal a confusing message about her type. As a result, the manager may sometimes find that a confusing message would yield a better contract for her than a direct truthful message. Given $S(Y, R)$, type-θ manager chooses reporting strategy R, and effort e to

$$\max_{R,e} E_\theta[U(S(Y, R), e)].$$

Thus in order to solve an adverse selection problem, the principal appears to deal with the manager's opportunistic choice of reporting strategies as well as that of effort levels.

Fortunately, the following result, known as "the Revelation Principle," tells us that the principal does not need to consider the agent's various (direct/indirect) reporting strategies other than the truthful and direct reporting strategy. That is, the principal only need to consider compensation schemes that will induce the agent to reveal her type truthfully.

Theorem 4.1 *(Revelation Principle) If a salary scheme $S^*(Y, R)$ implements $e^*(\theta)$ and a reporting strategy $R^*(\theta)$ for $\theta \in \Theta$, then there exists another salary scheme $\bar{S}(Y, \theta)$ that also implements $e^*(\theta)$ and a direct truthful reporting strategy. In particular, $\bar{S}(Y, \theta) \equiv S^*(Y, R^*(\theta))$.*

Proof: The proof is straightforward from the definition of optimality. Since $S^*(Y, R)$ implements, for each θ, an effort level of $e^*(\theta)$ and a reporting strategy of $R^*(\theta)$, we have

$$(R^*(\theta), e^*(\theta)) \in \arg\max_{R,e} E_\theta[U(S^*(Y, R), e)].$$

Consider another compensation scheme $\bar{S}(Y, \theta)$ such that $\bar{S}(Y, \theta) \equiv S^*(Y, R^*(\theta))$. We claim that $\bar{S}(Y, \theta)$ also implements, for each θ, $e^*(\theta)$ (without affecting the agent's and the principal's utilities) but under a direct truthful reporting strategy, i.e.,

$$(\theta, e^*(\theta)) \in \arg\max_{\theta',e} E_\theta[U(\bar{S}(Y, \theta'), e)].$$

This claim however follows immediately from the definition of the optimality of $(R^*(\theta), e^*(\theta))$. To be complete, recall that $\forall(R, e)$,

$$E_\theta[U(S^*(Y, R^*(\theta)), e^*(\theta))] \geq E_\theta[U(S^*(Y, R), e)].$$

In particular, the above inequality holds for $(R^*(\theta'), e)$, $\forall\theta' \in \Theta$. Therefore, $\forall(\theta', e)$,

$$E_\theta[U(\bar{S}(Y, \theta), e^*(\theta))] \geq E_\theta[U(\bar{S}(Y, \theta'), e)].$$

Finally note that since $\bar{S}(Y, \theta) \equiv S^*(Y, R^*(\theta))$, both salary schemes give identical utility levels for both the principal and agent. □

4.1 The Model: Pure Adverse Selection

The model in this section is from Laffont and Tirole [1986, JPE].

Both the principal and agent are risk neutral. In this model, one can conveniently interprete the principal as a firm and the agent as a contractor.

The agent produces goods for the principal at a cost of C. The cost C depends on the agent's effort is $e \in \mathcal{R}_{++}$, and the agent's ability $\theta \in \Theta$. In particular, we assume $C = \theta - e$. The structure of the outcome suggests that the agent with high ability has a low value in θ.

The principal observes C, but cannot observe e and θ separately. However, the principal knows that θ is distributed with a pdf $f(\theta)$ and its cumulative distribution function $F(\theta)$. We assume $f(\theta) > 0$ for all $\theta \in \Theta$.

Upon receiving goods from the agent, the principal reimburses the cost C and pays S to the agent. Let q be a constant, representing the value of the agent's work to the principal. The principal's utility is $q - C - S$.

The agent's utility is $S(\theta) - D(e)$, where $D : A \to \mathcal{R}^+$ is the disutility of effort. (The agent exerts costly effort to reduce the cost.) We assume $D' > 0$ and $D'' > 0$. The agent's reservation utility is zero, i.e. in equilibrium

$$U(\theta) = S(\theta) - D(e(\theta)) \geq 0.$$

The principal's expected utility is

$$\int_{\Theta} (q - C(\theta) - S(\theta))\, f(\theta) d\theta.$$

The principal's problem is to design a menu of contracts for the agent specifying the compensation $S(\theta)$ and $C(\theta)$.

Problem 4.1 *Choose* $S(\theta)$ *and* $C(\theta)$ *to*

$$\begin{aligned} \max \quad & \int_{\Theta} (q - C(\theta) - S(\theta))\, f(\theta) d\theta \\ s.t. \quad & S(\theta) - D(\theta - C(\theta)) \geq 0 \\ & \theta \in \arg\max_{\theta' \in \Theta} U(\theta'|\theta) := S(\theta') - D(\theta - C(\theta')). \end{aligned}$$

The second is the incentive constraint *via* the revelation principle.

The IC implies for all $\theta' \in \Theta$,

$$U(\theta) \equiv U(\theta|\theta) \geq U(\theta'|\theta)$$

Theorem 4.2 *S is incentive compatible if and only if* $\forall\theta,\, \theta' \in \Theta$,

$$U(\theta') - U(\theta) = -\int_{\theta}^{\theta'} D'(\theta - C(\theta)) d\theta,$$

and $\dot{C}(\theta) \geq 0$ *a.e.*

Proof: *Necessity.* By the IC,

$$U(\theta) \geq U(\theta'|\theta) = U(\theta') + \{U(\theta'|\theta) - U(\theta'|\theta')\}.$$

But,

$$\begin{aligned} & U(\theta'|\theta) - U(\theta'|\theta') \\ & = D(\theta' - C(\theta')) - D(\theta - C(\theta')) \end{aligned}$$

Thus we have

$$U(\theta) - U(\theta') \geq D(\theta' - C(\theta')) - D(\theta - C(\theta'))$$

Also by symmetry,

$$U(\theta') - U(\theta) \geq D(\theta - C(\theta)) - D(\theta' - C(\theta))$$

Therefore,

$$D(\theta - C(\theta')) - D(\theta' - C(\theta')) \geq U(\theta') - U(\theta) \geq D(\theta - C(\theta)) - D(\theta' - C(\theta)) \tag{4.1}$$

By letting $\theta' \to \theta$, we have

$$\frac{dU}{d\theta} = -D'(\theta - C(\theta)).$$

Eq. (4.1) also implies that

$$0 \leq \int_{\theta}^{\theta'} \int_{C(\theta)}^{C(\theta')} D''(\hat{\theta} - C) dC d\hat{\theta}$$

Thus, if $\theta' > \theta$, then $C(\theta') \geq C(\theta)$, which implies $\dot{C}$ exists a.e. and $\dot{C} \geq 0$ a.e.

Sufficiency. Suppose not. That is, suppose that when $U'(\theta) = -D'(\theta - C(\theta))$ and $\dot{C} \geq 0$, it is sometimes possible that there exist θ and θ' such that

$$U(\theta'|\theta) > U(\theta|\theta).$$

Without loss of generality, assume that $\theta' > \theta$. Then,

$$U(\theta'|\theta) - U(\theta'|\theta') > U(\theta) - U(\theta').$$

Thus

$$D(\theta' - C(\theta')) - D(\theta - C(\theta')) > -\int_{\theta'}^{\theta} D'(\hat{\theta} - C(\hat{\theta})) d\hat{\theta},$$

or

$$\int_{\theta}^{\theta'} D'(\hat{\theta} - C(\theta')) d\hat{\theta} - \int_{\theta}^{\theta'} D'(\hat{\theta} - C(\hat{\theta})) d\hat{\theta} > 0.$$

But the above inequality cannot be true because

$$-\int_{\theta}^{\theta'} \int_{C(\hat{\theta})}^{C(\theta')} D''(\hat{\theta} - C)) dC \, d\hat{\theta} < 0$$

Contradiction. □

By Theorem 4.2, we can rewrite Problem 4.1 as follows.

Problem 4.2 *Choose U and e to*

$$\begin{aligned} \max \quad & \int_{\Theta} (q - \theta + e - U(\theta) - D(e))\, f(\theta) d\theta \\ s.t. \quad & \dot{U}(\theta) = -D'(e) \\ & U(\theta) \geq 0 \\ & 1 - \dot{e}(\theta) \geq 0 \end{aligned}$$

Since $U(\theta)$ is decreasing, the second constraint can be replaced with $U(\theta_h) \geq 0$. Since the higher the U, the lower the principal's expected profit, we set $U(\theta_h) = 0$. We first ignore the last constraint, and later on we will see that this constraint is not binding. By treating U as the state variable and e as the decision variable, The Hamiltonian is

$$H = (q - \theta + e - U - D(e))\, f(\theta) - \mu D'(e).$$

By the Pontryagin maximum principle, the necessary conditions are

$$\begin{aligned} & \dot{\mu}^*(\theta) = -\frac{\partial H}{\partial U} = f(\theta) \\ & \mu^*(\theta_l) = 0 \\ & e^*(\theta) \in \arg\max_{e'} H(e', \theta, U, \mu) \end{aligned}$$

Thus, assuming an interior solution

$$\begin{aligned} & \mu^*(\theta) = F(\theta) \\ & D'(e^*) = 1 - \frac{F(\theta)}{f(\theta)} D''(e^*). \end{aligned}$$

Note that $F(\theta_l) = 0$, and therefore for the most efficient type, i.e., θ_l, the optimal effort level is the same as the first best level. For other types, the optimal effort levels are lower than the first-best.

Differentiating the last equation, we have

$$\left(D'' + \frac{F(\theta)}{f(\theta)}D'''\right)\dot{e^*} = -D''(e^*)\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right).$$

Assume that

$$\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) \geq 0. \tag{4.2}$$

This assumption is satisfied by popular distributions such as uniform, normal, exponential, and chi-squared. Also assume that

$$D''' \geq 0. \tag{4.3}$$

Under these two reasonable assumptions in (4.2) and (4.3), we have

$$\dot{e^*}(\theta) \leq 0,$$

i.e., effort decreases with θ. Thus the original constraint $\dot{e^*}(\theta) \leq 1$ is not binding.

Since $C^*(\theta) = \theta - e^*(\theta)$, the two assumptions also imply

$$\dot{C^*}(\theta) \geq 0,$$

i.e. the cost increases with θ.

The necessary conditions also shed some light on the relationship between the compensation and the realized cost C. Recall that

$$U^*(\theta) = \int_\theta^{\theta_h} D'(e^*(\hat{\theta}))d\hat{\theta}$$
$$S^*(\theta) = D(e^*(\theta)) + U^*(\theta)$$

$$S^*(\theta) = D(\theta - C^*(\theta)) + \int_\theta^{\theta_h} D'(\hat{\theta} - C^*(\hat{\theta}))d\hat{\theta},$$

or

$$\begin{aligned}
\frac{dS^*}{dC} &= \frac{dS^*}{d\theta}/\frac{dC^*}{d\theta} = \left(D'(1-\dot{C^*}) - D'\right)/\dot{C^*} = -D' \leq 0 \\
\frac{d^2S^*}{dC^2} &= -D''\left(\frac{1}{\dot{C^*}} - 1\right) = -D''\left(\frac{\dot{e^*}}{\dot{C^*}}\right) \geq 0
\end{aligned}$$

The compensation S^* is decreasing and convex in cost.

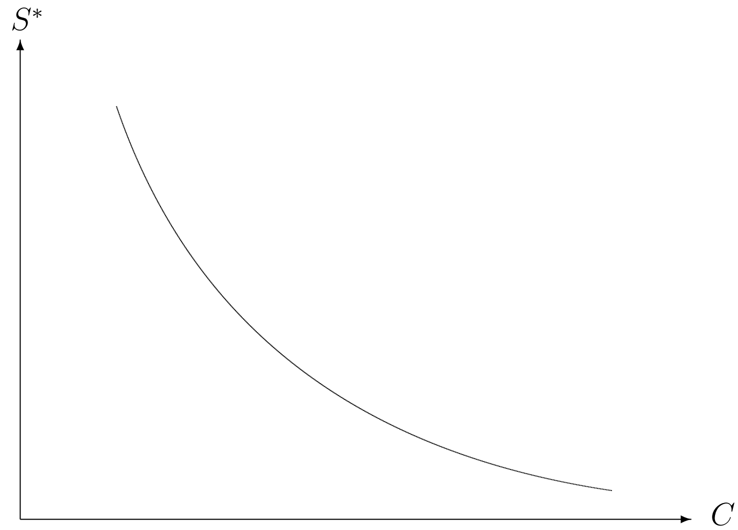


Figure 4.1: Compensation vs. Realized Cost

4.2 Managerial Contracting under Adverse Selection and Moral Hazard

This section is based on Sung [2000].

There are two time periods, 0 and 1. The principal is risk neutral and the agent is risk averse with constant absolute risk aversion r.

The outcome at time 1 is given by $Y = \mu + \sigma B$, where B is a standard normal random variable, and $\mu, \sigma \in \mathcal{R}_+$. The agent exerts costly effort to control μ and σ. His cost is given by $c(\mu, \sigma, \theta)$.

The parameter θ stands for the agent's ability, which is known to the agent himself but not to the principal. The principal knows that θ is a static random variable, independent of B and distributed on a strictly positive compact interval $\Theta := [\theta_L, \theta_H]$, with a probability density function given by $h(\theta)$, where $h(\theta) > 0$ for all $\theta \in \Theta$.

The principal can observe Y and σ. But μ and θ are unobservable. The principal's problem is as follows:

Problem 4.3 *Choose* $\{\mu(\theta); \theta \in \Theta\}$ *and a menu of contracts*

$\{(S(Y,\theta),\sigma(\theta)),\theta\in\Theta\}$ *to maximize* $E_{\theta,B}\left[Y(\theta)-S(Y,\theta)\right]$ *subject to the following constraints.*

(i) $Y(\theta)=\mu^*(\theta,\theta)+\sigma(\theta)B$

(ii) $\forall(\hat{\theta},\theta)$,

$$\mu^*(\hat{\theta},\theta)\in\arg\max_{\mu}\left\{\begin{array}{l} E_B\left[-\exp\left\{-r\left\{S(Y,\hat{\theta})-c(\mu,\sigma(\hat{\theta}),\theta)\right\}\right\}\right]\\ s.t.\ Y(\theta)=\mu+\sigma(\hat{T})B\end{array}\right.$$

(iii) $\forall\theta$,

$$E_B\left[-\exp\left\{-r\left\{S(Y,\theta)-c(\mu^*(\theta,\theta),\sigma(\theta),\theta)\right\}\right\}\right]\geq-\exp\left\{-r\mathcal{W}_0\right\}$$

(iv) $\forall\theta,\ \theta\in\arg\max_{\hat{\theta}}E_B\left[-\exp\left\{-r\left\{S(Y,\hat{\theta})-c(\mu^*(\hat{\theta},\theta),\sigma(\hat{\theta}),\theta)\right\}\right\}\right].$

Assume a menu of linear contracts, i.e., $S(Y,\theta)=\alpha(\theta)+\beta(\theta)Y$, where $\alpha(\theta),\beta(\theta)\in\mathcal{R}$ for all θ. This assumption is without loss of generality, because we will see later the optimal menu of contracts are in fact linear contract in our continuous-time model.

Under the linearity assumption, the manager's expected utility is

$$\begin{aligned}\Phi &= E\left[-e^{-r(\alpha+\beta Y-c(\mu,\sigma,\theta))}\right]\\ &= -e^{-r\left(\alpha-c(\mu,\sigma,\theta)+\beta\mu-\frac{1}{2}r\sigma^2\beta^2\right)}\end{aligned}$$

Let $\Phi^*=\max_{\mu}\Phi$, and define $Q(\theta)$ by $\Phi^*=-e^{-rQ(\theta)}$. Then the salary function S that satisfy the incentive constraint (ii) has to be given in the following form:

$$\begin{aligned}S(Y,T) &= Q(\theta)+c(\mu,\sigma,\theta)-c_{\mu}(\mu,\sigma,\theta)\mu\\ &\quad+\frac{r}{2}\sigma^2c_{\mu}^2(\mu,\sigma,\theta)+c_{\mu}(\mu,\sigma,\theta)Y.\end{aligned}$$

Note that (μ,σ,θ) is short for $(\mu(\theta),\sigma(\theta),\theta)$. If the agent chooses $\mu\in U$ at optimum, any linear salary function should be given in the form of (4.4). Furthermore, one can also show that given any salary function in the form of (4.4) with (μ,σ,θ), the agent of type θ will choose μ at optimum, if c is concave in μ for given σ,T. Therefore, if c is concave in μ for given σ,θ and if the agent chooses $\mu\in U$ at optimum, then the managerial incentive and participation constraints can be equivalently replaced with Eq. (4.4).

Proposition 4.1 *Assume that $\mu(\hat{\theta}, \theta)$, and $Q(\theta)$ are continuously once differentiable in $\hat{\theta}$ and in θ, respectively. Then constraint (ii), conditions (iii) and (iv) in the principal's problem imply that*

$$Q(\theta) = \mathcal{W}_0 + \int_{\theta}^{\theta_H} c_\theta(\mu, \sigma, \theta') d\theta' \tag{4.4}$$

Proof: Let us write $S = S(\mu, \sigma, \theta)$ to emphasize its dependence on those variables. Then at (interior) optimum, the relevant class of managerial salary functions becomes $S(\mu, \sigma, \theta)$ where $(\mu, \sigma, \theta) \in U \times \Sigma \times \Theta$. Suppose that the manager of type θ chooses $S(\hat{\mu}, \hat{\sigma}, \hat{\theta})$. Then the manager's maximum expected utility becomes

$$\begin{aligned} u(\hat{\theta}; \theta) &= \max_{\mu} E\left[-\exp\left\{-r\left(Q(\hat{\theta}) + \hat{c} - \hat{c}_\mu \hat{\mu} + \frac{r}{2}\hat{\sigma}^2 \hat{c}_\mu^2 + \hat{c}_\mu Y_1 \right.\right.\right. \\ &\qquad\qquad \left.\left.\left. - c(\mu, \hat{\sigma}, \theta)\right)\right\}\right]. \\ &= \max_{\mu} -\exp\left\{-r\left(Q(\hat{\theta}) + \hat{c} - \hat{c}_\mu \hat{\mu} + \hat{c}_\mu \mu - c(\mu, \hat{\sigma}, \theta)\right)\right\} \end{aligned}$$

That is, the agent solves

$$\max_{\mu} \Phi(\mu, \hat{\theta}, \theta) := \hat{c} - \hat{c}_\mu \hat{\mu} + \hat{c}_\mu \mu - c(\mu, \hat{\sigma}, \theta)$$

Let $\mu^*(\hat{\theta}, \theta)$ solves the above maximization problem. Then the incentive and truth-telling conditions imply that $\mu(\hat{\theta}) \equiv \mu^*(\hat{\theta}, \hat{\theta})$, or $\mu(\theta) \equiv \mu^*(\theta, \theta)$ for all θ. Then we have

$$\left.\frac{\partial}{\partial \hat{\theta}} \Phi(\mu^*, \hat{\theta}, \theta)\right|_{\hat{\theta}=\theta} = c_\theta(\mu(\theta), \sigma(\theta), \theta), \tag{4.5}$$

where the RHS of the above equation is the partial derivative of c with respect to its third argument. To see this, let us write explicitly the maximum value of Φ when the type-θ agent chooses $S(Y, \hat{\theta})$.

$$\begin{aligned} &\Phi(\mu^*(\hat{\theta}, \theta), \hat{\theta}, \theta) \\ &= c(\mu(\hat{\theta}), \sigma(\hat{\theta}), \hat{\theta}) - c_\mu(\mu(\hat{\theta}), \sigma(\hat{\theta}), \hat{\theta})\mu(\hat{\theta}) \\ &\quad + c_\mu(\mu(\hat{\theta}), \sigma(\hat{\theta}), \hat{\theta})\mu^*(\hat{\theta}, \theta) - c(\mu^*(\hat{\theta}, \theta), \sigma(\hat{\theta}), \theta) \end{aligned}$$

Thus, we have

$$\begin{aligned}
&\frac{\partial \Phi}{\partial \hat{\theta}}(\mu^*(\hat{\theta},\theta),\hat{\theta},\theta) \\
&= c_\theta + c_\mu \frac{d\hat{\mu}}{d\hat{\theta}} + c_\sigma \frac{d\sigma}{d\hat{\theta}} - c_{\mu\mu}\hat{\mu}\frac{d\hat{\mu}}{d\hat{\theta}} - c_{\mu\sigma}\hat{\mu}\frac{d\sigma}{d\hat{\theta}} - c_{\mu\theta}\hat{\mu} - c_\mu \frac{d\hat{\mu}}{d\hat{\theta}} \\
&\qquad + c_{\mu\mu}\mu^* \frac{d\hat{\mu}}{d\hat{\theta}} + c_{\mu\sigma}\mu^* \frac{d\sigma}{d\hat{\theta}} + c_{\mu\theta}\mu^* + c_\mu \frac{\partial \mu^*}{\partial \hat{\theta}} - c_\mu \frac{\partial \mu^*}{\partial \hat{\theta}} - c_\sigma \frac{d\sigma}{d\hat{\theta}}
\end{aligned}$$

Since $\mu^*(\theta,\theta) = \mu(\theta)$, this proves equation (4.5).

On the other hand, constraint (iv) implies that given θ, we have for all $\hat{\theta}$

$$-e^{-rQ(\theta)} \geq -\exp\left\{-r\left(Q(\hat{\theta}) + \Phi(\mu^*,\hat{T},T)\right)\right\} \tag{4.6}$$

(Otherwise the agent of type θ will prefer $S(Y,\hat{\theta})$ to $S(Y,\theta)$.) In particular, since $\Phi^*(\mu^*,\theta,\theta) = 0$,

$$-e^{-rQ(\theta)} = \max_{\hat{\theta}} - \exp\left\{-r\left(Q(\hat{\theta}) + \Phi(\mu^*,\hat{\theta},\theta)\right)\right\}$$

The FOC at $\hat{\theta} = \theta$ is

$$\begin{aligned}
0 &= \frac{d}{d\hat{\theta}}Q(\hat{\theta}) + \frac{\partial}{\partial \hat{\theta}}\Phi(\mu^*(\hat{\theta},\theta),\hat{\theta},\theta) \\
&= \frac{d}{d\hat{\theta}}Q(\hat{\theta}) + c_\theta(\hat{\mu},\hat{\sigma},\theta)
\end{aligned}$$

This proves the assertion. □

Now we can use Eq.(4.4) to simplify the principal's problem under the adverse selection. By Eq.(4.4), the risk-neutral principal's problem is relaxed and restated as

$$\begin{aligned}
&\max_{\mu,\sigma} \int_{\theta_L}^{\theta_H} \left(\mu(\theta) - Q(\theta) - c(\mu,\sigma,\theta) - \frac{r}{2}\sigma^2 c_\mu^2(\mu,\sigma,\theta)\right) h(\theta)d\theta \\
&= \max_{\mu,\sigma} \int_{\theta_L}^{\theta_H} \Big(\mu(\theta) - c(\mu,\sigma,\theta) - \frac{r}{2}\sigma^2 c_\mu^2(\mu,\sigma,\theta) \\
&\qquad\qquad - \mathcal{W}_0 - \int_{\theta}^{\theta_H} c_\theta(\mu,\sigma,\bar{\theta})d\bar{\theta}\Big) h(\theta)d\theta
\end{aligned} \tag{4.7}$$

By changing variables, the problem in (4.7) can be restated as

$$\begin{aligned} \max_{\mu,\sigma} \quad & \int_{\theta_L}^{\theta_H} \left(\mu - c(\mu,\sigma,\theta) - \frac{r}{2}\sigma^2 c_\mu^2(\mu,\sigma,\theta) - x\right) h(\theta) d\theta \\ \text{s.t.} \quad & \frac{dx}{d\theta} = -c_\theta(\mu,\sigma,\theta) \\ & x(\theta_H) = 0 \end{aligned} \tag{4.8}$$

Then, the Hamiltonian is:

$$H = \left(\mu - c(\mu,\sigma,\theta) - \frac{r}{2}\sigma^2 c_\mu^2(\mu,\sigma,\theta) - x\right) h(\theta) - \lambda c_\theta(\mu,\sigma,\theta) \tag{4.9}$$

The necessary conditions are

$$\frac{d\lambda}{d\theta} = -\frac{\partial H}{\partial x} = h(\theta) \tag{4.10}$$

$$\lambda(\theta_L) = 0 \tag{4.11}$$

$$(\mu(\theta), \sigma(\theta)) \in \arg\max_{\mu',\sigma'} H(\mu', \sigma', \theta, x(\theta), \lambda(\theta)) \tag{4.12}$$

Eq.'s (4.10) and (4.11) imply $\lambda(\theta) = \int_{\theta_L}^{\theta} h(\theta')d\theta'$. Note that the effect of the adverse selection is captured by the last term of the Hamiltonian. Let us call $(\mu(\theta), \sigma(\theta))$ in (4.12) the third-best solution to the principal's problem.[1] Since $\lambda(\theta)$ is zero when $\theta = \theta_L$, we can conclude that the choices of (μ, σ) by the manager with the most efficient technology are not affected by the presence of the adverse selection problem. That is, when $\theta = \theta_L$, the third best coincides with the second best.

It is well known that linear contracts may not be optimal in the above discrete-time setup. See Merrlees [1974] and also Baron and Besanko [1987]. Nevertheless. the results in this section can be justified as optimal in a continuous-time setting. See Sung [2000].

[1] Recall that under the pure moral hazard situation as in Sung [1995], the second-best solution to the principal's problem can be found by maximizing over μ and σ

$$\mu - c(\mu,\sigma,\theta) - \frac{r}{2}\sigma^2 c_\mu^2(\mu,\sigma,\theta).$$

4.2.1 Applications

In corporate management, the manager (the agent) can improve the outcome with his effort and/or by choosing a project from a given project opportunity set.

Let us write the outcome process as follows:

$$Y = (\mu + g(\sigma)) + (\sigma + b)B, \tag{4.13}$$

where the function $g(\sigma)$ represents the efficient frontier of (existing) project opportunities. The manager chooses a value for μ with a cost of $c(\mu_t, \theta)$ and a project $(g(\sigma), \sigma)$ without incurring a cost.

The project choice decision is costless but observable to investors (the principal).[2] The effectiveness parameter of his effort, θ, may not be known to investors. (Such a case may more significantly occur in a venture-capitalist contracting.) The costly managerial effort may be directed toward either enhancing existing project opportunities or improving the profitability of ongoing operations.

When project $(g(\sigma), \sigma)$ is chosen, the expected profit or the net present value (NPV) of the firm improves by $g(\sigma)$ and the volatility by σ. Note that our NPV is computed as an amount before the managerial compensation is paid out. Thus it is the NPV in the first-best sense.

After the project choice, the total instantaneous volatility is $(\sigma + b)$ where b is an uncontrollable positive number and $\sigma > -b$. Furthermore, for convenience, we set $g(0) = 0$ to represent the *status quo* without the project choice. The parameter b can be viewed as the instantaneous volatility generated from the ongoing operations of the firm.

When the profit from a chosen project is highly negatively correlated with the profit from ongoing operations, σ can take on a negative number. Thus σ is not the volatility of the income generated by a chosen project, but the contribution of a chosen project to the overall profit of the firm.

The efficient frontier of project opportunities may alternatively be interpreted as a capital budgeting opportunity set, delineating a risk-return tradeoff schedule for different levels of initial investments. Let

[2]The assumption of this costlessness is to highlight the decision making aspect as a part of managerial activities. See Sung [1995] for more explanation.

I be the level of investment at time t that increases both the instantaneous NPV and volatility of the overall profit by $q(I)$ and $s(I)$, respectively. Assume that both q and s are strictly increasing, q is concave and s is convex in I.[3] Let $\sigma \equiv s(I)$ and $g(\sigma) \equiv q(s^{-1}(\sigma))$, where s^{-1} is the inverse of s. Then q is strictly increasing and concave in σ, and the outcome with I is given by $Y = (\mu + q(I)) + (s(I) + b)B$, which is equivalent to $Y = (\mu + g(\sigma)) + (\sigma + b)B$. Thus a decision to invest an amount of $I = s^{-1}(\sigma)$ bears exactly the same interpretations as the choice of a project $(g(\sigma), \sigma)$.

In order to understand why the capital budgeting opportunity set $q(s^{-1}(\sigma))$ may be increasing and concave in σ, consider the following hypothetical case. The manager has found a positive NPV investment opportunity with a required capital investment I. Initially, to increase the NPV on the investment, he may simply double the amount to double the NPV. Of course, doubling I will increase the risk as well. However, as the size of investment becomes larger and the corresponding risk keeps increasing, the incremental NPV may only approach zero because of capital market competitions and transactions costs including price impacts. Thus an investment opportunity requiring I can generate a set of investment opportunities that can be described by an increasing and concave schedule like $q(s^{-1}(\sigma))$.

Whichever interpretation is chosen to be, the problem can be stated as follows:

Problem 4.4 *Choose* $\{\{\mu(\theta)\}; \theta \in \Theta\}$ *and a menu of contracts* $\{(S(Y,\theta), \{\sigma(\theta)\}); \theta \in \Theta\}$ *to maximize* $E\left[Y(\theta) - S(Y,\theta)\right]$ *subject to the following constraints:*

$$(i)\, \forall\theta,\ Y(\theta) = (\mu^*(\theta,\theta) + g(\sigma(\theta))) + (\sigma(\theta) + b)B$$

$$(ii)\, \forall(\hat{\theta},\theta),$$

$$\mu^*(\hat{\theta},\theta) \in \arg\max_{\mu} \left\{ \begin{array}{l} E^B\left[-\exp\left\{-r\left\{S(Y,\hat{\theta}) - c(\mu,\theta)\right\}\right\}\right] \\ s.t.\ Y(\theta) = (\mu + g(\sigma(\theta))) + (\sigma(\hat{\theta}) + b)B \end{array} \right.$$

$$(iii)\, \forall\theta,\ E^B\left[-\exp\left\{-r\left\{S(Y,\theta) - c(\mu^*(\theta,\theta),\theta)\right\}\right\}\right] \geq -\exp\left\{-r\mathcal{W}_0\right\}$$

[3] Suppose that the firm needs to decide the level of investment I in a particular project, the risk of which is independent of the risk of the ongoing operation. Then s is strictly increasing and convex in I.

$$(iv)\,\forall\theta,\,\theta\in\arg\max_{\hat\theta}E^{B}\left[-\exp\left\{-r\left\{S(Y,\hat\theta)-c(\mu^*(\hat\theta,\theta),\theta)\right\}\right\}\right].$$

Problem 4.4 is a slightly modified case of Problem 4.3. Thus similarly to (4.9), the risk-neutral investors' problem is to

$$\max_{m,\sigma}\mu+g(\sigma)-c(\mu,\sigma,\theta)-\frac{r}{2}c_\mu^2(\mu,\sigma,\theta)(\sigma+b)^2-\frac{\int_{\theta_L}^{\theta}h(\theta')\,d\theta'}{h(\theta)}c_\theta(\mu,\sigma,\theta).$$

The FOC necessary conditions are

$$1-c_\mu-r(\sigma+b)^2c_\mu c_{\mu\mu}-\frac{\int_{\theta_L}^{\theta}h(\theta')d\theta'}{h(\theta)}c_{\theta\mu}=0$$
$$g_\sigma-r(\sigma+b)c_\mu^2=0$$

The second condition tells us about the decision rule for project selection under adverse selection. The rule is identical to that of the pure moral hazard model. That is, the project is selected so that marginal NPV from the project before managerial compensation is equal to the marginal managerial compensation-risk premium. Note however that the chosen project under moral hazard and adverse selection can be different from that of pure moral hazard, because the sensitivities of contracts c_μ are in general different between the two cases.

Figure 4.2 conceptually illustrates the optimal project selections for three different cases: A is the first-best (or the NPV-maximizing) project; B is the second-best; and C is the average third-best (with adverse selection).

The horizontal axis represents projects that yield zero NPV in the absence of moral hazard, and the vertical axis NPV of the projects in terms of the increment of the mean of the overall outcome. The origin of the graph for $g(\sigma)$ represents *status quo* or ongoing operations plus some new riskfree NPV-zero projects. The slopes of tangents in the figure are $r(\sigma+b)c_\mu^2$, the marginal compensation-risk premium.

As seen in the figure, the optimal σ increases as the sensitivity of the contract c_μ decreases. The manager with the highest ability chooses the same project as he would without the adverse selection problem. (Since the first order conditions for the manager with the highest ability are the same as in the pure moral hazard, optimal c_μ is the same and thus he chooses the same project.) When $c_{\mu\mu\mu} \geq 0$, managers with lower ability however are more willing to take risk to increase the marginal NPVs or to do less costly hedging activities than they would in the absence of the adverse selection problem. On average, the more severe the adverse selection problem is, the lower the sensitivity of the managerial contract to the outcome.

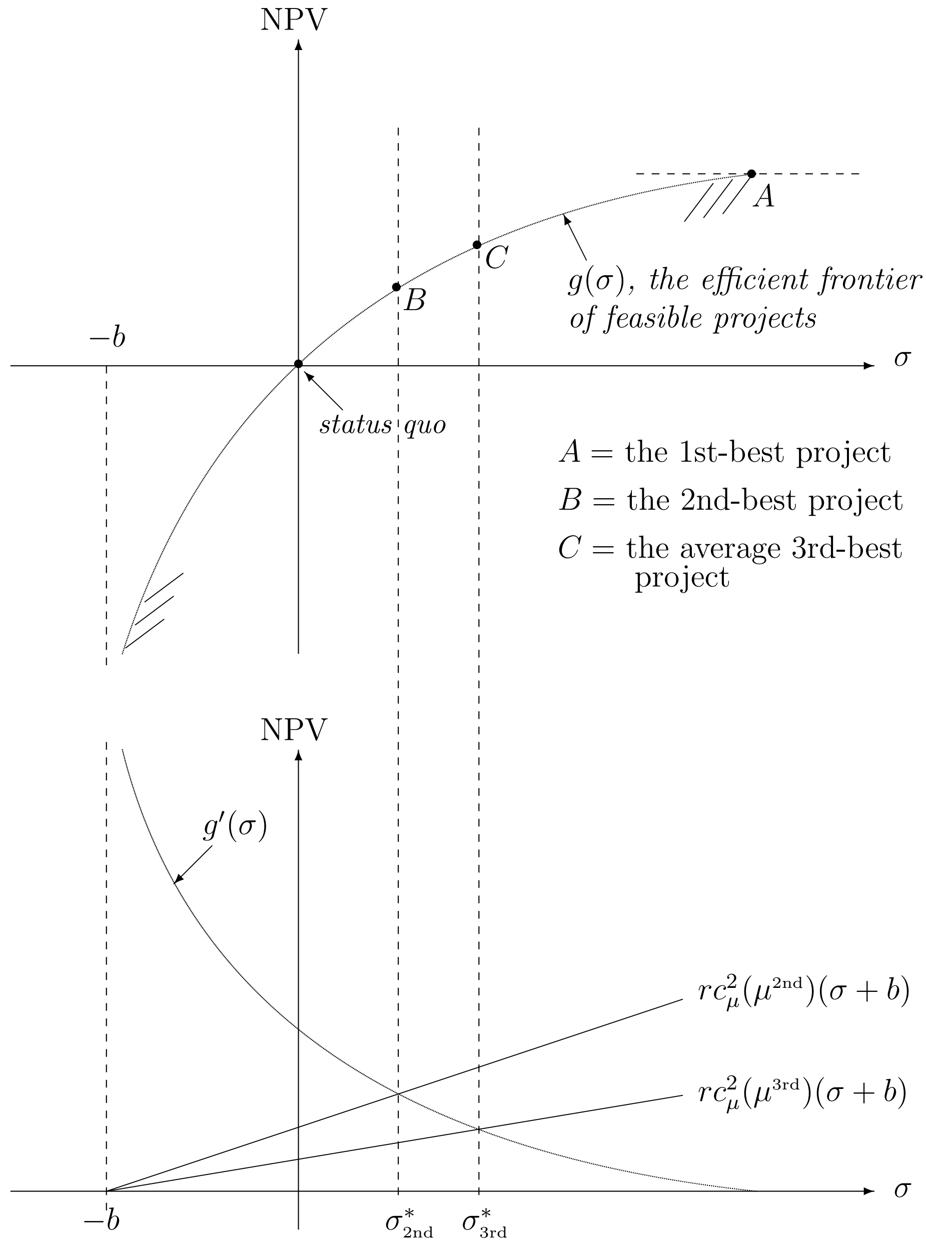


Figure 4.2: Project selection rules under moral hazard and adverse selection

Appendix A

Dynamic Programming Equation with Exponential Utility

Let B_t be n-dimensional standard Wiener processes, and also let

$$\begin{aligned} F &: \mathcal{R}^n \longrightarrow \mathcal{R} \\ \alpha &: [0,1]\times\mathcal{R}^n\times U \longrightarrow \mathcal{R} \\ f,\,\beta &: [0,1]\times\mathcal{R}^n\times U \longrightarrow \mathcal{R}^n \\ \sigma &: [0,1]\times\mathcal{R}^n\times U \longrightarrow \mathcal{R}_+^{n\times n} \end{aligned}$$

Consider the following stochastic control problem:

$$\begin{aligned} \max_u \quad & E\left[-e^{-r\{F(Y_1)-\int_0^1 \alpha(t,Y_t,u_t)dt-\int_0^1 \beta(t,Y_t,u_t)dB_t\}}\right] \\ \text{s.t.} \quad & dY_t = f(t,Y_t,u_t)dt + \sigma(t,Y_t,u_t)dB_t. \end{aligned}$$

The stochastic control problem of the form (A.1) is a general form that will cover our optimization problems in the text as special cases. Define

$$J(u;t,\mathcal{F}_t) = E\left[-e^{-r\{F(Y_1)-\int_t^1 \alpha(s,Y_s,u_s)ds-\int_t^1 \beta^\top(s,Y_s,u_s)dB_s\}} \mid \mathcal{F}_t\right]$$

Assumption A.1 *A function V is continuously differentiable in t and twice continuously differentiable in Y_t, and satisfies for some $p>2$ the following conditions:*

$$\int_0^1 V(t,Y_t)^p dt, \quad \textit{and} \quad \int_0^1 \|\frac{\partial V}{\partial y}(t,Y_t)\|^p dt, \quad \textit{are in } L^1(\Omega,\mathcal{F},P).$$

Lemma A.1 *Assume that* $J(u;t,\mathcal{F}_t) = J(u;t,Y_t)$.[1] *Define the value function* V *as follows:*

$$V(t,Y_t) := \sup_u J(u;t,Y_t).$$

Also assume that α *and* β *are bounded and that there exists an optimal control policy* u^* *such that* $V(t,Y_t) = J(u^*;t,Y_t)$, *and that* V *satisfy the assumption (A.1). Then the value function* V *necessarily satisfy the following dynamic programming equation for any time* $t \in [0,1]$ *a.e.:*

$$0 \equiv \frac{\partial V}{\partial t} + \max_u \left\{ \frac{\partial V}{\partial y}^{\top}(f + r\sigma\beta) + \frac{tr}{2}\frac{\partial^2 V}{\partial y^2}\sigma\sigma^{\top} + rV(\alpha + \frac{r}{2}\|\beta\|^2) \right\}.$$

The terminal condition is $V(1,Y_1) = -e^{-rF(Y_1)}$.

Lemma A.2 *Assume that* α *and* β *are bounded. Consider a function* V *such that* $V : [0,1] \times \mathcal{R}^{m+n} \to \mathcal{R}$, *satisfying the assumption (A.1) and the following partial differential equation*

$$0 \equiv \frac{\partial V}{\partial t} + \sup_u \left\{ \frac{\partial V}{\partial y}^{\top}(f + r\sigma\beta) + \frac{tr}{2}\frac{\partial^2 V}{\partial y^2}\sigma\sigma^{\top} + rV(\alpha + \frac{r}{2}\|\beta\|^2) \right\}$$

with the terminal condition being $V(1,y) = -e^{-rF(y)}$.
Assume that there exists u^* *such that*

$$u^* \in \arg\max_u \left\{ \frac{\partial V}{\partial y}^{\top}(f + r\sigma\beta) + \frac{tr}{2}\frac{\partial^2 V}{\partial y^2}\sigma\sigma^{\top} + rV(\alpha + \frac{r}{2}\|\beta\|^2) \right\}.$$

Then u^* *is an optimal control and* $J(u^*;t,\mathcal{F}_t) = V(t,Y_t)$.

Proof of Lemma A.1: For $0 \le t \le \tau \le 1$, let

$$u = \begin{cases} \hat{u}_s & \text{if } t \le s \le \tau \\ u^*_s & \text{if } \tau < s \le 1 \end{cases}$$

where u^* is an optimal control pair, and $\hat{u}$ is an arbitrary admissible control. Let us define

$$\phi_\tau := e^{r(X_\tau - X_t)}$$

[1]This definition is possible if the controls of interest are Markovian.

and

$$dX_t := \alpha(t, Y_t, u)dt + \beta^\top(t, Y_t, u)dB_t.$$

Then we have by using the Itô's rule,

$$\phi_\tau - 1 = \int_t^\tau r\phi_s dX_s + \frac{1}{2}\int_t^\tau r^2\phi_s\|\beta\|^2 ds.$$

Define $P_\tau := \phi_\tau \cdot V(\tau, Y_\tau)$. Then

$$E\left[P_\tau \,|\, \mathcal{F}_t\right] = E\left[\phi_\tau \cdot V(\tau, Y_\tau) \,|\, \mathcal{F}_t\right].$$

This implies

$$V(t, Y_t) \geq E\left[P_\tau \,|\, \mathcal{F}_t\right]$$

Since $V \in C^{1,2}$ by assumption, an application of the Itô's rule yields

$$\begin{aligned} V(\tau, Y_\tau) &= V(t, Y_t) + \int_t^\tau \frac{\partial V}{\partial s}(s, Y_s)ds + \int_t^\tau \frac{\partial V^\top}{\partial y}(s, Y_s)dY_s^c \\ &\quad + \int_t^\tau \frac{tr}{2}\left(\frac{\partial^2 V}{\partial y^2}(s, Y_s)\sigma\sigma^\top\right)ds \end{aligned} \tag{A.1}$$

Therefore we have

$$\begin{aligned} P_\tau &= V(t, Y_t) + \int_t^\tau \phi_s dV_s + \int_t^\tau V(s, Y_s)d\phi_s + \int_t^\tau r\phi_s\frac{\partial V^\top}{\partial y}(s, Y_s)\sigma\beta ds \\ &= V(t, Y_t) + \int_t^\tau \phi_s\left\{\frac{\partial V}{\partial s}(s, Y_s) + \frac{\partial V^\top}{\partial y}(s, Y_s)(f + r\sigma\beta)\right. \\ &\quad \left. + \frac{tr}{2}\left(\frac{\partial^2 V}{\partial y^2}(s, Y_s)\sigma\sigma^\top\right) + rV(s, Y_s)(\alpha + \frac{1}{2}r\|\beta\|^2)\right\}ds \\ &\quad + \int_t^\tau \phi_s\{\frac{\partial V^\top}{\partial y}(s, Y_s)\sigma + rV(s, Y_s)\beta^\top\}dB_t. \end{aligned} \tag{A.2}$$

Now we want to show that the stochastic integrals are martingales. For this purpose it suffices to show that

$$E[\int_t^\tau \phi_s^2\|\frac{\partial V^\top}{\partial y}(s, Y_s)\sigma + rV(s, Y_s)\beta^\top\|^2 ds] < \infty. \tag{A.3}$$

But ϕ_s can be rewritten as follows:

$$\begin{aligned}
&\phi_s \\
&= e^{r(X_s - X_t)} \\
&= \exp\left\{ r \left(\int_t^s \alpha d\tau + \int_t^s \beta^\top dB_\tau \right) \right\} \\
&= \exp\left\{ \int_t^s r\beta^\top dB_\tau - \frac{1}{2}\int_t^s r^2 \|\beta\|^2 d\tau \right\} \cdot \exp\left\{ r \int_t^s (\alpha + \frac{r}{2}\|\beta\|^2) d\tau \right\}.
\end{aligned}$$

Note that the first exponential is a Girsanov density $G(s)$. Since β is bounded, the L^p norm of $G(s)$, i.e. $\{E[\int_t^1 G^p(s)ds]\}^{\frac{1}{p}}$, is finite for all $p > 0$. By Hölder's inequalities, this implies that (A.3) holds. Therefore the stochastic integral is a martingale and its expectation is zero.

Recall that $0 \geq E[P_\tau|\mathcal{F}_t] - V(t, Y_t)$ for all $\tau \in [t, 1]$. But if $u \equiv u^*$ for all $s \in [t, 1]$, then $0 \equiv E[P_\tau|\mathcal{F}_t] - V(t, Y_t)$ for all $\tau \in [t, 1]$. Hence by dividing $E[P_\tau|\mathcal{F}_t] - V(t, Y_t)$ by $\tau - t$ and letting $\tau \to t+$, we obtain the statement of Lemma A.1. □

Proof of Lemma A.2: Consider a function V that satisfies the conditions stated in the Lemma A.2. Define $P_\tau := \phi_\tau \cdot V(\tau, Y_\tau)$, where ϕ is as defined in the proof of Lemma A.1. Then by the same calculation, we obtain the equation (A.2). But by the definition of V in the statement, the equation (A.2) implies that $E[P_\tau \,|\, \mathcal{F}_t] \leq V(t, Y_t)$ for all τ ($\geq t$). In particular,

$$V(t, Y_t) \geq E[P_1 \,|\, \mathcal{F}_t] = J(u; t, Y_t).$$

for all nonanticipative arbitrary control law u. Furthermore notice that if u^* as defined in the statement of Lemma A.2 is used for the time period between $[t, \tau]$, then for all $\tau \in [t, 1]$, $V(t, Y_t) = E[P_\tau \,|\, \mathcal{F}_t]$. But then $V(t, Y_t) = E[P_1 \,|\, \mathcal{F}_t] = J(u^*; t, Y_t)$. Therefore, u^* is an optimal control, and V is the value function. □

Bibliography

[1] Arora, N., and H. Ou-Yang, "A Continuous-Time Model of Explicit and Implicit Incentives," Working Paper, University of North Carolina, 2000.

[2] Baron, D., and D. Besanko, "Monitoring, Moral Hazard, Asymmetric Information, and Risk Sharing in Procurement Contracting," *Rand Journal of Economics*, 18, 1987, 509-532.

[3] Chung, K.L., and Williams, R.J. *Introduction to Stochastic Integration*, 2nd ed., Boston, Birkhäuser, 1990.

[4] Detemple, J., S. Govindaraj, and M. Loewenstein, "Hidden Actions, Agents with Non-Separable Utility and Wealth-Robust Intertemporal Incentive Contracts," Working Paper, Boston University, 2001.

[5] Elliott R.J. *Stochastic Calculus and Applications*, New York, Springer-Verlag, 1982.

[6] Ehrbar, A., *Stern Stewart's EVA: The Real Key to Creating Wealth*, John Wiley & Sons, Inc., 1998.

[7] Fleming, W. H., and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, New York, Springer-Verlag, 1975.

[8] Fudenberg, D., and J. Tirole, *Game Theory*," MIT Press, Cambridge, MA, 1991.

[9] Grossman, S. and Hart, O. "An Analysis of the Principal-Agent Problem," *Econometrica*, Vol. 51 (1983), 7 - 45.

[10] Hellwig, M., and K. M. Schmidt, "Discrete-Time Approximations of Holmstrom-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision," Working Paper, University of Mannheim, 1998.

[11] Holmstrom, B. "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. 10 (1979), 74 - 91.

[12] Holmstrom, B., and Milgrom, P. "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, Vol. 55 (1987), 303-328.

[13] Jensen, M.C., and Murphy, K.J. "Performance Pay and Top-Management Incentives," *Journal of Political Economy*, Vol. 98 (1990), 225-264.

[14] Jewitt, I. "Justifying the First-order Approach to Principal-agent Problems," *Econometrica*, Vol. 56 (1988), 1177 - 1190.

[15] Laffont, J-J., and Tirole, J., "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, Vol.94 (1986), 614-641.

[16] Milgrom, P., and Roberts, J., *Economics, Organization and Management*, Prentice Hall, New Jersey, 1992.

[17] Mirrlees, J.A. "Notes on Welfare Economics, Information and Uncertainty," in M. Balch, D. McFadden, and S. Wu (Eds.) *Essays in Economic Behavior under Uncertainty*, 1974, 243-258.

[18] Mirrlees, J.A. "The Optimal Structure of Incentives within an Organization," *Bell Journal of Economics*, Vol. 7 (1976), 105 - 131.

[19] Müller, H., "The First-Best Sharing Rule in the Continuous-Time Principal-Agent Problem with Exponential Utility," *Journal of Economic Theory*, 79, 1998, 276-280.

[20] Müller, H., "Asymptotic Efficiency Dynamic Principal-Agent Problems," *Journal of Economic Theory*, 91, 2000, 292-301.

[21] Ou-Yang, H., "Optimal Contracts in a Continuous-Time Delegated Portfolio Management Problem," Working Paper, University of North Carolina, 1999.

[22] Rogerson, W.P. "The First-order Approach to the Principal-Agent Problems," *Econometrica*, Vol. 53 (1985), 1357-1367.

[23] Ross, S. "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, Vol. 63 (1973), 134 - 139.

[24] Salanié, B., *The Economics of Contracts: A Primer*, MIT Press, Cambridge, MA, 1998.

[25] Schättler, H., and Sung, J. "The First-Order Approach to Continuous-Time Principal-Agent Problem with Exponential Utility," *Journal of Economic Theory*, Vol. 61 (1993), 331 - 371.

[26] Schättler, H., and J. Sung, "On Optimal Sharing Rules in Discrete- and Continuous-Times Principal-Agent Problems with Exponential Utility," *Journal of Economic Dynamics and Control*, 21, 1997, 551-574.

[27] Shavell, S. "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, Vol. 10 (1979), 55 - 73.

[28] Stole, L.A., *Lectures on the Theory of Contracts and Organizations*,
http://gsblas.uchicago.edu/papers/lecture.pdf, 1997.

[29] Sung, J. *Managerial Compensation and Corporate Management under Moral Hazard*, unpublished Ph.D. thesis, Washington University in St. Louis, 1991.

[30] Sung, J. "Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems," *RAND Journal of Economics*, Vol. 26 (1995), 720-743.

[31] Sung, J. "Corporate Insurance and Managerial Incentives," *Journal of Economic Theory*, Vol. 74 (1997), 297-332.

[32] Sung, J. "Optimal Contracts under Moral Hazard and Adverse Selection: A Continuous-Time Approach," UIC working paper, 2000.

Index

adverse selection, 45

bond, 7

capital markets, 1
cashflow, 2
Corporate finance, 1

discrete-time, 13
doubly relaxed, 24

EPS, 7
EVA, 8

firm's problems, 2
first order approach, 17
FSLIC, 4

implementable, 37
incentive compatibility, 17

monitoring, 7
moral hazard, 7

ORA, 8

present value, 2
principal-agent problems, 4

Revelation Principle, 46
risk-neutral, 15
riskfree, 2
risky, 5
ROE, 8

S&Ls, 4
safe, 5
stock option, 8