2. Complete Contracts I: Static Bilateral Contracting

Huihua NIE
School of Economics
Renmin University of China
www.niehuihua.com
niehuihua(at)vip.163.com

2.1 Moral Hazard I: Single Task

2.1.1 Standard model: symmetric information

Just think about the fable of an abbot and a little monk.

A standard model in contract theory or information economics always starts from a benchmark model where information is symmetric and there is no incentive problem. It always includes the following elements: players with preference, product technology, information structure, and timing.

Suppose all the related information is symmetric and verifiable. The Principal (P)'s problem is to design first best contract to maximize his utility subjected to the Agent (A)'s participation condition or individual rationality (IR) condition. The key variable are A's one-dimension efforts $a \in A$ and wage $w(x_i)$ and core problem is how to share risk. $\Pr(x = x_i \mid a) = p_i(a)$ and $p_i(a) > 0$, $\sum p_i(a) = 1$. A's cost $c(a)$ is convex, and $c' > 0$, $c'' \geq 0$. All parties' utilities are VNM formula, so P's object function is $V(x, w) = v(x - w)$; A's is $U(w, a) = u(w) - c(a)$, his preserve utility is $\underline{U}$. Of course, utility function are concave, so $v' > 0$, $v'' \leq 0$, $u' > 0$, and $u'' \leq 0$. Timing is omitted.

[Note 1] Uncertainty/risk is necessary for risk sharing problem and incentive problem between P and A, so x_i is a stochastic variable, and $p_i(a) > 0$.

[Note 2] Objective function's concavity and cost function's convexity are necessary.

[Note 3] Cost function's separablility is convenient and necessary, for A's risk attitude doesn't change with his effort costs; it ensures objective function's concavity.

[Note 4] P and A's utility functions indicate that A's revenue is P's cost.

[Note 5] There is no "incentive problem".

So, P's mathematic program is

$$\underset{\{a, w(x_i)\}}{Max} \sum_{i=1}^{n} p_i(a) v(x_i - w(x_i))$$

$$s.t. \quad \sum_{i=1}^{n} p_i(a)u(w(x_i)) - c(a) \geq \underline{U}$$

According to concave program's rule, there is unique maximum (a^{FB}, w^{FB}). The Lagrange function is

$$\underset{\{a,w(x_i)\}}{Max} \sum_{i=1}^{n} p_i(a)v(x_i - w(x_i)) + \lambda[\sum_{i=1}^{n} p_i(a)u(w(x_i)) - c(a) - \underline{U}] \quad (2\text{-}1)$$

Differentiating expression (2-1) with respect to w, we get

$$\frac{v'(x_i - w^{FB}(x_i))}{u'(w^{FB}(x_i))} = \lambda \quad \text{for all the } i \in \{1, 2, ..., n\} \quad (2\text{-}2)$$

Because $\lambda > 0$ (otherwise $v' = 0$, or $u' = \infty$), the rate of P's and A's marginal utility is constant.

Optimal wage level

Proposition 1: If P is risk neutral and A is risk averse, P should offer A constant wage and provide full insurance. (Point **n**)

Proof: P is risk neutral, so $v' = cont$. With $\lambda > 0$, $u' = cont$ for all x_i, which means u is irrelevant to output, i.e., $w^{FB}(x_1) = w^{FB}(x_2)$.

Because IR is binding, we get optimal wage $w^{FB}(x_i) = u^{-1}(c(a) + \underline{U})$.

It's Frank Knight (1921)'s insight. How about it?

Proposition 2: If A is risk neutral and P is risk averse, P should sell his firm to A at a certain price. (Point **m**)

The logic is similar to Proposition 1. Since $v'(x_i - w^{FB}(x_i))$ is constant, we let $x_i - w^{FB}(x_i) = k$, i.e. $w^{FB}(x_i) = x_i - k$. Substitute it into IR condition, we get $\sum_{i=1}^{n} p_i(a^{FB})(x_i - k) - c(a) = \underline{U}$, i.e. the sale price $k = \sum_{i=1}^{n} p_i(a^{FB})x_i - \underline{U} - c(a^{FB})$. In this case, the firm is a labor management firm (LMF).

Homework: Read this paper and write a note with/without AI.

Bartlett et al., 1992, "Labor-Managed Cooperatives and Private Firms in North Central Italy: An Empirical Comparison", *Industrial and Labor Relations Review*, 46(1): 103-118.

Proposition 3: If both P and A are risk neutral, P should sell his firm to A.

Proof. Homework.

Proposition 4: If both P and A are risk averse, they share risk according to a certain proportion. (Point **E**)

Proof:

$$\frac{v'(x_i - w^{FB}(x_i))}{u'(w^{FB}(x_i))} = \lambda$$

$$-v'(x_i - w^{FB}(x_i)) + \lambda u'(w^{FB}(x_i)) = 0 \tag{2-3}$$

Differentiating with respect to x,

$$-v''(1 - \frac{dw^{FB}}{dx_i}) + \lambda u'' \frac{dw^{FB}}{dx_i} = 0$$

$$\lambda = \frac{v''(1 - \frac{dw^{FB}}{dx_i})}{u'' \frac{dw^{FB}}{dx_i}} \tag{2-4}$$

Rewriting expression (2-2),

$$\lambda = \frac{v'}{u'} \tag{2-5}$$

Integrating (2-4) and (2-5), we get

$$-\frac{v''(1 - \frac{dw^{FB}}{dx_i})}{v'} + \frac{u'' \frac{dw^{FB}}{dx_i}}{u'} = 0 \tag{2-6}$$

Recall that $r_P = -\frac{v''}{v'}$ and $r_A = -\frac{u''}{u'}$ is P's and A's Arrow-Pratt measure of absolute risk aversion respectively. Hence we have[①]

$$\frac{dw^{FB}}{dx_i} = \frac{r_P}{r_P + r_A} \tag{2-7}$$

Interestingly, we find Proposition 1-3 are the special case of Position 4. If we consider share of revenue as divisible asset, then we can get an argument of common property rights. Also we can summarize these results as follow figure 2-1. θ_i stands for different contingency.

① There are some errors in textbook Y ch.2.

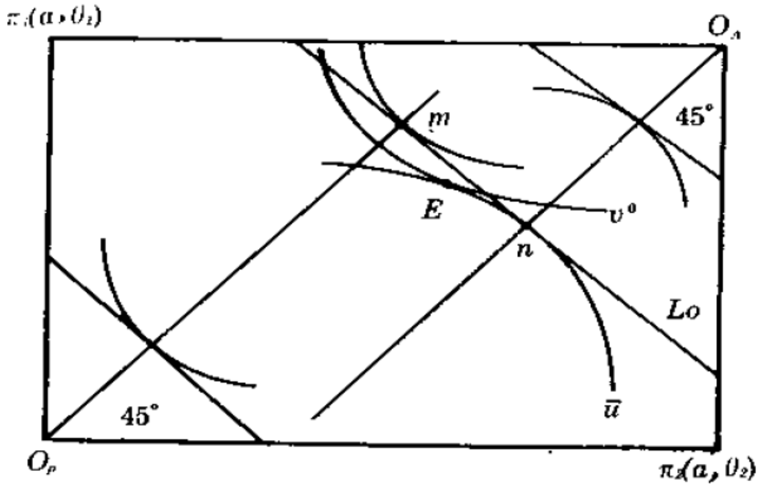


图 5. 1　帕累托最优风险分担合同

Optimal effort level

The analysis of optimal effort level is not intuitive as that of optimal wage effort.

Case I: P is risk neutral and A is risk averse. Where, $w^{FB}(x_i) = u^{-1}(c(a) + \underline{U})$. We can rewrite P's problem as

$$\max_a [\sum_{i=1}^{n} p_i(a)x_i - u^{-1}(c(a) + \underline{U})]$$

FOC:

$$\sum_{i=1}^{n} p_i'(a^{FB})x_i = (u^{-1})'(c(a^{FB}) + \underline{U})c'(a^{FB})$$

Applying inverse function principle, we have

$$\sum_{i=1}^{n} p_i'(a^{FB})x_i = \frac{c'(a^{FB})}{u'(w^{FB})}$$

SOC:

$$\sum_{i=1}^{n} p_i''(a^{FB})x_i + \frac{u''}{(u')^3}(w^{FB})c'(a^{FB})^2 - \frac{c''(a^{FB})}{u'(w^{FB})} \le 0$$

In order to ensure SOC holds and hence the optimal effort is unique, we must have $\sum_{i=1}^{n} p_i''(a^{FB})x_i \le 0$.

Case II: A is risk neutral and P is risk averse, A's action must be efficient. ☞ Why?

Proof:

A’s problem is $\underset{a}{Max}[\sum_{i=1}^{n} p_i(a)x_i - k - c(a)]$,

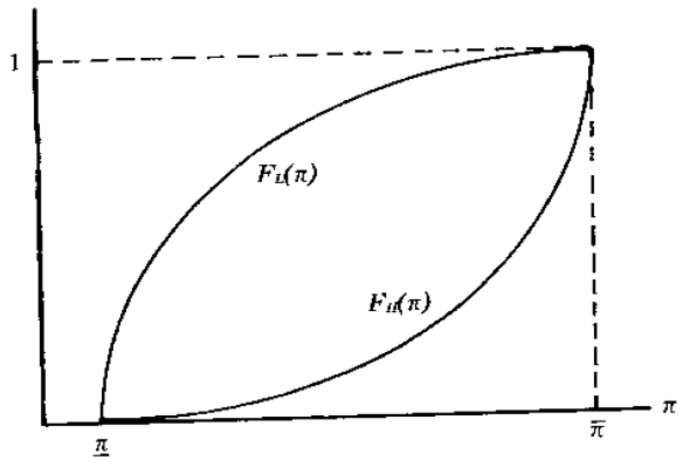
FOC: $p_i'(a^{FB})x_i = c'(a^{FB})$. ***Q.E.D.***

2.1.2 Standard model: asymmetric information

Now information is asymmetric. When A is risk neutral and P is risk neutral or averse, A buyouts the firm and becomes the residual claimer, so he has full incentive to make efforts. The fist-best outcome can be attained.

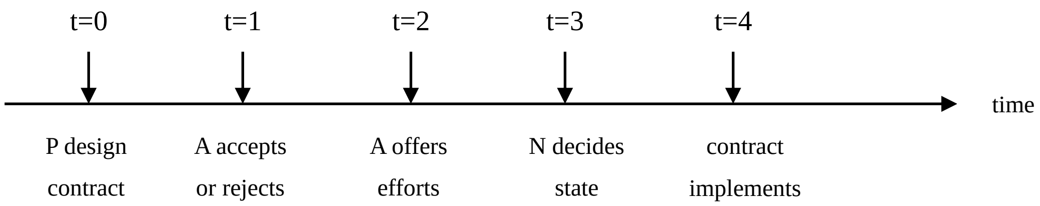
In order to make things interesting, we suppose P is risk neutral and A is risk averse (or bilateral risk aversion). ☞ Is there any difference between two cases?

Suppose that $a \in \{a^H, a^L\}$, $c(a^H) > c(a^L)$, and p^H first-order stochastically dominates p^L. That is $\sum_{i=1}^{k} p_i^H < \sum_{i=1}^{k} p_i^L$ [①] for all the $k = 1, 2, ..., n-1$ and $\sum_{i=1}^{n} p_i^H = \sum_{i=1}^{n} p_i^L = 1$, i.e. $Ep^H > Ep^L$. When x_i is single value, we have figure 2-3.



[Note] The definition of probability.

The timing is as figure 2-4 which indicates it is a dynamic game with complete information.



① It’s not contradict to $p_i^H > p_i^L$, because likelihood ratio is individual but FOSD is local.

Fig. 2-4

Under asymmetrical information, the optimal that P provides A fixed wage and full assurance will lead to moral hazard. Probably P wants high efforts, otherwise he just pays A the constant wage $w^L(x_i)=u^{-1}(c(a^L)+\underline{U})$. So P must design a payment scheme $w(x_i)$ for different efforts provided by A, and A's incentive compatibility (IC) condition constraint must be satisfied (where we use SPE method to solve the dynamic game with complete information):

$$\sum_{i=1}^{n} p_i^H u(w(x_i)) - c(a^H) \geq \sum_{i=1}^{n} p_i^L u(w(x_i)) - c(a^L)$$

$$\Rightarrow \sum_{i=1}^{n} (p_i^H - p_i^L) u(w(x_i)) \geq c(a^H) - c(a^L)$$

So, the P's problem is[①]

$$\underset{w(x_i)}{Max} \sum_{i=1}^{n} p_i^H v(x_i - w(x_i))$$

s.t. (IR) $\sum_{i=1}^{n} p_i^H (a) u(w(x_i)) - c(a^H) \geq \underline{U}$

(IC) $\sum_{i=1}^{n} (p_i^H - p_i^L) u(w(x_i)) \geq c(a^H) - c(a^L)$

$$L(w(x_i),\lambda,\mu) = \sum_{i=1}^{n} p_i^H v(x_i - w(x_i)) + \lambda[\sum_{i=1}^{n} p_i^H (a) u(w(x_i)) - c(a^H) - \underline{U}]$$
$$+\mu[\sum_{i=1}^{n} (p_i^H - p_i^L) u(w(x_i)) - c(a^H) + c(a^L)]$$

Differentiating with respect to w, we have

$$\frac{p_i^H}{u'(w(x_i))} = \lambda p_i^H + \mu(p_i^H - p_i^L)$$

i.e., $\frac{1}{u'(w(x_i))} = \lambda + \mu(1 - \frac{p_i^L}{p_i^H})$ for all the $i = 1, 2, \ldots, n$. (2-8)

Kuhn-Tucker theorem requires $\lambda > 0$ and $\mu > 0$. Under bilateral risk aversion, or generally, we have

$$\frac{v'(x_i - w(x_i))}{u'(w(x_i))} = \lambda + \mu(1 - \frac{p_i^L}{p_i^H}) \quad (2\text{-}9)$$

① Here I adopt different formulation from textbook Y.

(i) Because $\mu \neq 0$, according to Mirrlees-Holmstrom condition (2-9), first best (condition (2-2)) can not be achieved. Note that now A must burden some risk according to (2-8).

(ii) $\frac{p_i^L}{p_i^H}$ is so-called likelihood rate. If $p_i^L \geq p_i^H$, $w \leq w^{FB}$ [①]; if $p_i^L < p_i^H$, $w > w^{FB}$, which means asymmetrical information increase incentive payment to A.

(iii) When $\frac{p_i^L}{p_i^H} < 1$, $\frac{p_i^L}{p_i^H} \downarrow, w \uparrow$; $\frac{p_i^L}{p_i^H} \uparrow, w \downarrow$.When $\frac{p_i^L}{p_i^H} = 1$, there is no new information.

(iv) In order to ensure $w(x_i)$ monotonically increase with x_i, we must let $\frac{p_i^L}{p_i^H}$ monotonically decrease with i, that is monotone likelihood rate property (MLRP). Intuitively, the less performance from lazing, the more incentive for agents to be diligent.

2.1.3 Special Model: Mean-Variance Approach

The specific model is from Holmstrom-Milgrom (1987) involved with linear contracts, normally distributed performance, and exponential utility. Performance is assumed to be equal to $x = a + \varepsilon$, where $\varepsilon \sim (0, \sigma^2)$. P is risk neutral and A has constant absolute risk-averse (CARA) preferences represented by $U = -\exp(-r\hat{w})$, where $w(x) = \alpha + \beta x$, and $\beta \in [0,1]$. For simplicity the cost-of-effort function is quadratic: $c(a) = \frac{ba^2}{2}$.

The P's problem is then to solve

$$\underset{\{a,\alpha,\beta\}}{Max}\ E(x - w)$$

s.t. (IR) $E(-e^{-r\hat{w}}) \geq E(\underline{U})$

(IC) $a \in \arg\max_a E(-e^{-r\hat{w}})$

$$E\hat{w} = Ew - c(a)$$

$$= E[\alpha + \beta(a + \varepsilon)] - \frac{ba^2}{2}$$

$$= \alpha + \beta a - \frac{ba^2}{2}$$

① $\frac{1}{u'(w(x_i))}$ actually is convex.

$$EU = E[-\exp(-r\hat{w})]$$
$$= \int_{-\infty}^{+\infty} -\exp(-r\hat{w}) f(\hat{w}) d\hat{w}$$
$$= \int_{-\infty}^{+\infty} -e^{-r\hat{w}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\hat{w}-\alpha-\beta a+\frac{ba^2}{2}}{\sigma_w}\right)^2} d\hat{w}$$
$$= -e^{-r(E\hat{w}-\frac{1}{2}r\sigma_w^2)}$$

$$\because \sigma_w^2 = Var(\alpha+\beta x) = \beta^2\sigma^2$$

$\therefore$ The *certainty equivalent compensation* of agent is $ACE = \alpha + \beta a - \frac{ba^2}{2} - \frac{1}{2} r\beta^2\sigma^2$.

[Note] The meaning of CE. According to the last term of formula of ACE, actually effort costs have no direct relationship with risk attitude.

So, $a \in \arg\max_a E(-e^{-r\hat{w}}) = \arg\max_a \{\alpha + \beta a - \frac{ba^2}{2} - \frac{1}{2} r\beta^2\sigma^2\}$, FOC:

$$a^* = \beta / b$$

The above solution indicates backward induction applied in the dynamic game with complete information. And The P's problem is equivalent to be

$$\max_{\{a,\alpha,\beta\}} (1-\beta)a - \alpha$$

$$s.t. \quad \alpha + \beta a - \frac{ba^2}{2} - \frac{1}{2} r\beta^2\sigma^2 = \underline{U} \quad \text{(Binding)}$$

$$\therefore \beta^* = \frac{1}{1+rb\sigma^2} \tag{2-10}$$

When $\sigma^2 \to +\infty$, $\beta^* = 0$; $\sigma^2 \to 0$, $\beta^* = 1$. $r\uparrow$, $\beta^*\downarrow$. $b\uparrow$, $\beta^*\downarrow$. Formula (2-10) indicates the tradeoff between incentive and insurance in moral hazard model.

Note that first-best efficient effort is: $a^{FB} \in \arg\max(a - \frac{ba^2}{2})$, i.e., $a^{FB} = 1/b \geq a^*$, so under moral hazard agents provide inefficient effort level.

What's wrong with mean-variance approach?

The mean-variance approach is attractive in that it leads to a simple and intuitive closed-form solution; however, linear contracts are far from optimal. Suppose the support of $\varepsilon \in [-k, +k]$, $0 < k < \infty$. So the principal can provide a "boil-in-oil" contract by punishing the agent very severely for performance outcomes outside of $[a^{FB}-k,a^{FB}+k]$, and pay a constant transfer w^{FB} irrespective of the performance realizations in $[a^{FB}-k,a^{FB}+k]$, and then the principal can attain the first-best. In this setting, linear contracts are suboptimal.

[Note] It's not permitted that there is any cross area of the supports for different actions.

Contrarily, when the support is unbounded, Mirrelees (1975) proved that by extreme punishments the first best can be approximated, which is paradoxical to the assumptions of risk-averse agent.

Another problem: why do we set the optimal wage to be linear? See Holmstrom-Milgrom (1987, *Econometrica*).

2.1.4 General Model: First-Order Approach *

Following Mirrelees (1974, 1975, 1976) and Holmstrom (1979), we now turn to the characterization of general nonlinear incentive schemes. P may be risk averse and has a utility function given by $V(x-w)$; A is risk averse and has a utility function given by $u(w)-c(a)$; where $V'(\cdot)>0$, $V''(\cdot)\leq 0$, $u'(\cdot)>0$, $u''(\cdot)\leq 0$, $c'(\cdot)>0$, $c''(\cdot)\geq 0$. Suppose that performance is $x\in[\underline{x},\overline{x}]$ with CDF $F(x\mid a)$ and conditional density $f(x\mid a)$. P's problem is

$$\underset{\{w(x),a\}}{Max}\int_{\underline{x}}^{\overline{x}} V(x-w(x))f(x\mid a)dx$$

s.t. (IR) $\int_{\underline{x}}^{\overline{x}} u(w(x))f(x\mid a)dx-c(a)\geq \underline{U}$

(IC) $a\in\arg\max_{\tilde{a}\in A}[\int_{\underline{x}}^{\overline{x}} u(w(x))f(x\mid \tilde{a})dx-c(\tilde{a})]$

(ICa) FOC: $\int_{\underline{x}}^{\overline{x}} u(w(x))f_a(x\mid a)dx=c(a)$

(ICb) SOC: $\int_{\underline{x}}^{\overline{x}} u(w(x))f_{aa}(x\mid a)dx-c'(a)<0$

$$L(w(x_i),a,\lambda,\mu)=$$
$$\int_{\underline{x}}^{\overline{x}}\{V(x-w(x))f(x\mid a)+\lambda[u(w(x))f(x\mid a)dx-c(a)-\underline{U}]$$
$$+\mu[u(w(x))f_a(x\mid a)-c'(a)]\}dx$$

Differentiating with respect to $w(x)$, we obtain the FOC:

$$\frac{V'(x-w(x))}{u'(w(x))}=\lambda+\mu\frac{f_a(x\mid a)}{f(x\mid a)}$$

When $\mu=0$, the conditions reduce to Borch's rule (Borch, 1962) for optimal risk sharing.

Using contradiction, we can prove that $\mu > 0$. Unless $f_a(x \mid a) \le 0$, P should pay A more for higher x's than would be optimal for pure risk-sharing reasons.

What's wrong with first-order approach?

Mirrelees (1975) provides an illustration that sometimes one cannot substitute IC condition straight with A's FOC. Specifically, $EU(a) = \sum_{i=1}^{n} p(a_i) u(w(x_i)) - c(a)$ is not necessarily concave on effort for the existence of $\sum_{i=1}^{n} p(a_i)$. We can illustrate it as figure 2-5. C is P's indifference curve and FOC is A's First-order condition for efforts.

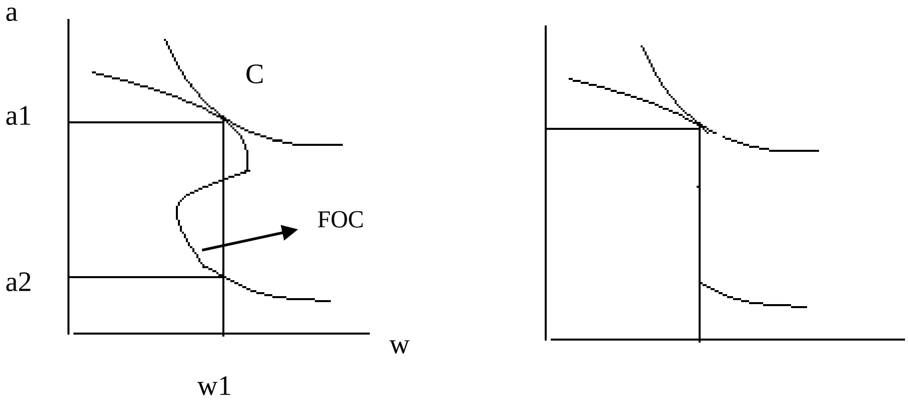


Fig. 2-5

In order to ensure the uniqueness of FOC, Rogerson (1985) gives sufficient conditions including MLRP ($w'(x) \ge 0$) and the convexity of the distribution function condition (CDFC) ($F_{aa}(x \mid a) \ge 0$). Unfortunately, MLRP and CDFC together are very restrictive condition. Even none of the well-know distribution functions satisfy both conditions simultaneously.

Grossman-Hart's approach

The difficulty is that the constraints are nonconvex----they do not rule out a nice convex set of points in the space of wage functions, but rather rule out a very complicated set of possible wage functions. A different approach is developed by Grossman & Hart (1983) to solve the problem.

The only key and special assumption: there are a finite number of possible output outcomes q_i: $0 \le q_1 < q_2 ... < q_N$, and $p_i(a) > 0$. That is to say, $p_i(a) \in (0,1)$. We can get $p_i(a)$ by any linear combination of $p_j(a)$ ($i \ne j$), i.e. $xp_j(a) + (1-x)p_j(a) = p_i(a)$, which satisfies the concavity condition by Rogerson (1985). Otherwise, if $p_i(a) \in [0,1]$, we can't get $p_i(a)$ by any linear combination of $p_j(a)$.

GH's solution is called the three-step procedure by Fudenberg and Tirole (1991), is to focus on contracts that induce the agent to pick a particular action rather than to directly attack the problem of maximizing profits. The first step is to find for each possible effort level the set of wage contracts that induce the agent to choose that effort level. The second step is to find the contract which supports that effort level at the lowest cost to the principal. The third step is to choose the effort level that maximizes profits, given the necessity to support that effort with the costly wage contract from the second step.

Mathematically the three-step procedure is equivalent to the following:

(1) Optimization of A's actions. A's conditions:

(IR) $\sum_{i=1}^{N} p_i(a)[\phi(a)u(w_i) - c(a)] \geq \underline{U}$

(IC) $\sum_{i=1}^{N} p_i(a)[\phi(a)u(w_i) - c(a)] \geq \sum_{i=1}^{N} p_i(\hat{a})[\phi(\hat{a})u(w_i) - c(\hat{a})]$ for all the $\hat{a} \in A$

Note that IR and IC conditions are not necessarily concave for w_i, so we let $u_i \equiv u(w_i)$ and $h = u^{-1}$. Here, since $p_i(a)$ is a linear function, we now have linear constraints.

(2) Implementation: $\underset{(u_1,u_2,...,u_N)}{Min} \sum_{i=1}^{N} p_i(a)h(u_i)$. Obviously, the object is concave, so that the Kuhn-Tucker conditions are necessary and sufficient with previous IR and IC conditions. Condition (2)-(3) implement as envelop theorem where w_i has been optimized in u_i.

(3) Optimization of total payoff: $\underset{a \in A}{Max}[\sum_{i=1}^{N} p_i(a)x_i - c(a)]$.

Breaking the problem into parts makes it easier to solve. Perhaps the most important lesson of the three-step procedure, however, is to reinforce the points that the goal of the contract is to induce the agent to choose a particular effort level and that asymmetric information increases the cost of the inducements.

2.1.5 Application

Sharecropping contracts

The Theory of Share Tenancy is the doctoral dissertation of Steven N. S. Cheung（张五常）, and published by The University of Chicago Pres in 1969. It is argued that the paper is the origin of contract theory, and firstly takes account transaction costs and risk aversion into contract analysis.

Motivation. The output increased dramatically after the Taiwan government put a mandatory regulation on the share percentage from 60% to 37.5% between landlords and tenants in 1949. It seems a paradox in economics, because regulation should decrease output and distort efficiency. The prevailing impression is that share tenancy results in inefficient allocation resources. Why? Firstly, it violates the fundamental rule MR=MC. Secondly, a share contract is less than a fixed price contract which makes tenants as residual claimer. And then Cheung tried to give an explanation from the perspective of contract theory.

Proposition 1: Different contractual arrangements do not imply different efficiencies of resource use as long as these arrangements are themselves aspects of private property rights and there is zero transaction cost. It means that resource allocation (sharecropping, fixed rent and fixed wage) under private property rights is the same whether the landowner cultivates the land himself, hires farm hands to do the billing, leases his holdings on a fixed rent basis, or shares the actual yield with his tenant. However, the rental percentage is not the only variable which the land-owner can adjust to maximize his wealth. The landlord will not allow one tenant to cultivate all the land he owns if parceling his land to several tenants will result in a higher total rent. This is illustrated in Figure 2 (Cheung, 1968). As the number of tenants cultivating the available land increases, the marginal product of land shifts upward relative to the situation where there is only one tenant.

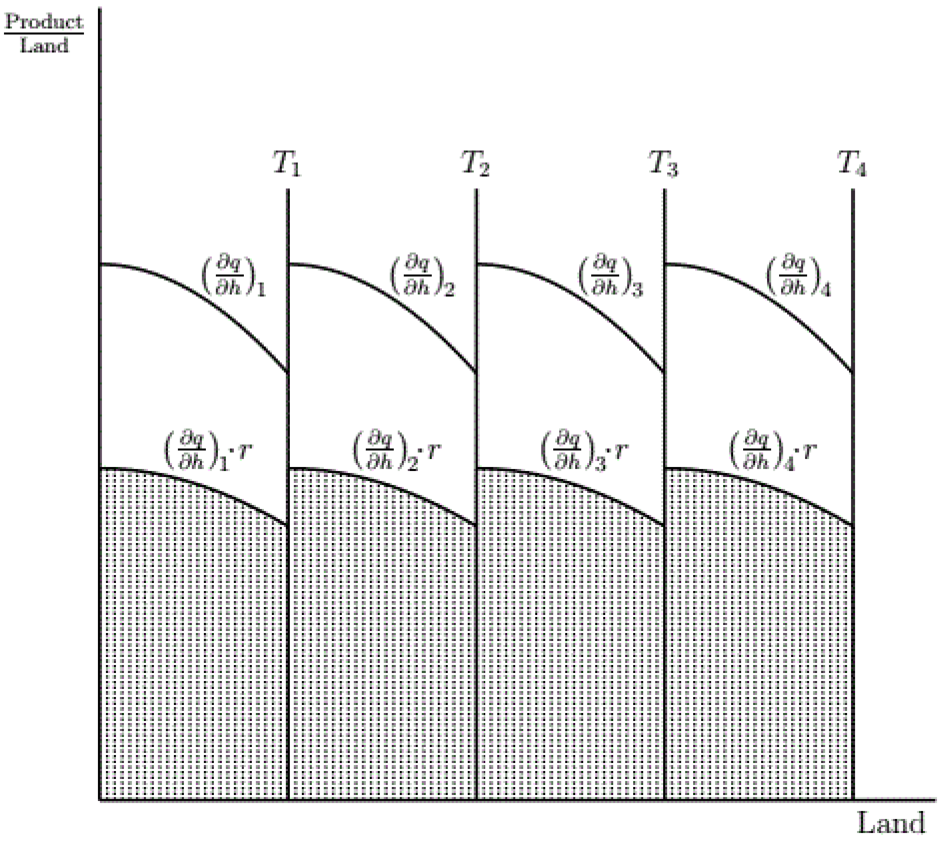


Fig. 2 Sharecropping with multiple tenants

Proposition 2: With positive transaction costs, different contractual arrangements exist because of two kinds of reason: risk and transaction costs. Transaction costs are relevant to the characteristics of land and crops. Two reasons can explain different contracts between landlords and farmers. Notice that there is competition between farmers.

Debt financing

The most incentive-efficient form of outside financing of the entrepreneur's project under limited liability may be some form of debt financing, for the entrepreneur is a "residual claimant"(Jensen-Meckling, 1976; Innes, 1990). For equity contracts, the entrepreneur will be requested to invest correspondingly or sign a Valuation Adjustment Mechanism with shareholders or investors.

[Note] With limited liability, we can model A's utility to be risk neutral.

2.1.6 Summary

- When A is risk neutral and wealthy, to let A be a "residual claimant" is a perfect solution.
- When A is risk averse, then more incentives come at the cost of a risk premium. P must trade off between incentive and assurance.
- When the distribution of output satisfies MLRP, then A 's remuneration is increasing in his performance.