도덕적 해이(moral hazard)와 agent 세분화*

남 익 현**

《目 次》

Ⅰ. 들어가며

정보경제학에서 가장 대표적으로 다루어지는 주제가 도덕적 해이 문제(moral hazard problem)이다. Moral hazard problem은 하나의 주체(대리인, agent)가 어떠한 행동을 하는데, 이러한 행동이 다른 주체(주인, principal)가 거래의 성사로부터 얻는 가치에 영향을 주고 또한 주인이 대리인의 행위를 완벽하게 파악할 수 없는 상황을 말한다.

예를 들어 화재보험의 경우를 살펴보자. 화재보험회사는 주인의 역할을 하며, 보험에 가입한 고객(대리인)이 인화물질을 철저히 관리하고 소화기를 비치하는 등 화재 예방을 위해 노력을 다해주기를 희망할 것이다. 하지만 화재보험회사는 보험가입자가 이러한 바람직한 행동을 실행하는지에 대해 감시하는 것은 불가능하다. 화재보험가입자는 화재예방을 위해 신경을 쓰는 것이 본인에게 비효용을 발생시키므로 이러한 불완전한 감시를 이용하여 주의 의무를 소홀히 하게 된다. 그래서 moral hazard라는 명칭이 유래한 것이다.

Moral hazard problem에서의 핵심주제는 행위자로 하여금 적절한 행동을 취하도록 어떻게 유도할 것인지를 다루는 것이다. 즉 행위자의 이해관계를 이해하고 이를 바탕으로 적절한 incentive mechanism을 구성하여 주인과 대리인 양자의 이익에 기여하도록 하고자 모색한다.

---

* 본 연구는 서울대학교 경영정보연구소의 연구비 지원에 의해 이루어졌습니다.

** 서울대학교 경영대학 교수

Ⅱ. moral hazard problem

앞서 언급한 내용을 보다 구체화하고 그 이론적 배경을 살펴보기로 하자. 하나의 경제주체인 주인(principal)이 다른 경제 주체인 대리인(agent)에게 어떠한 과업을 수행하도록 하는 상황을 다루고자 한다. 이를 principal-agent problem이라고 하며 이 문제에서 발생하는 어려움이 principal과 agent 사이의 moral hazard이다. 즉 principal은 agent가 자신을 위해서 일을 해주기를 바라지만 agent는 자기 자신의 효용함수를 고려하여 행동을 하다 보니 이 둘 사이의 충돌이 발생하는데 이를 moral hazard라고 부르는 것이다.

어떠한 과업을 담당하는 agent는 이 과업을 맡지 않았을 때 다른 대안으로부터 얻을 수 있는 효용수준을 최소한으로 보장받아야 principal의 제안을 수용할 것이다. 이러한 기준이 되는 효용수준을 '최소효용수준(reservation level of utility)'이라고 부른다. agent의 효용함수는 principal로부터 받는 금전적 보상인 상금 w와 자신이 얼마나 노력을 투입하였는지를 나타내는 행동 a로 구성이 된다. 의미 있는 문제가 되기 위해서 principal은 위험중립적(risk-neutral)이고 agent는 위험회피적(risk-averse)이라고 상정한다. agent의 von Neumann-Morgenstern utility function으로 우리는 논문 전개의 편의성을 위해 다음의 특정한 형태를 다루기로 하자. 보다 일반적인 효용함수에 대해서도 동일한 논리가 적용된다.

$$U(w,a) = \sqrt{w} - a$$

이러한 형태의 효용함수에서는 agent가 risk-averse임을 알 수 있다.

agent가 risk-neutral일 경우에는 moral hazard problem을 예방할 수 있다. 기본적인 방향은 agent에게 과업수행에 따른 상금이 기댓값으로 reservation level of utility가 되도록 계약체계을 만들면 된다. 이 경우 principal은 나머지 금액을 확보하게 되는데 이는 항상 일정한 상수가 된다. 이러한 계약체계는 본질적으로 principal이 agent에게 해당 과업의 사업권을 판매하여 모든 책임과 성과를 agent가 부담하도록 하는 것이다. 이와 같이 moral hazard problem이 발생하지 않는 경우, 우리는 first best solution을 얻었다고 한다. 하지만 현실적으로 많은 경우 agent는 본인 수입의 불확실성에 대한 비효용이 있으므로 risk-averse이고, 따라서 앞서 언급한 first-best solution을 포기하고 차선의 방안(second-best solution)을 강구하여야 한다.

Ⅲ. 유한개 행동과 결과에 해당하는 모형의 가정

이번 절에서는 agent가 선택할 수 있는 행동의 개수가 유한하고 agent의 행동에 의해 나오는 결과도 유한한 경우에 대해 모형화하는 것을 다루기로 하자. 먼저 agent가 선택할 수 있는 행동 a는 다음의 행동집합에 한정한다: $A=\{a_1,\cdots,a_N\}$. 여기서 핵심적인 내용은 principal이 agent가 어떤 행동을 선택하였는지 알 수가 없다는 것이다. principal은 다만 agent의 행동에 대한 불완전한 표시로서의 결과물을 관찰할 수 있다. 즉 principal은 결과물 s를 관찰할 수 있으며 이의 정의역은 $S=\{s_1,\cdots,s_M\}$로 표시하자. agent가 행동 a_n을 실행하였을 때 결과물 s_m이 나오는 것과의 관계식에는 불확실성이 내재하며 이러한 사건의 확률은 p_{nm}이라고 표시하자.

principal-agent problem의 핵심은 principal이 agent와 계약을 맺을 때 agent의 행동인 a에 따라 상금을 달리하는 방식을 택할 수가 없다는 것이다. 이는 principal이 agent의 행동을 관찰할 수 없기 때문이다. 다만 agent 행동의 결과물인 s는 관찰할 수 있기 때문에 s에 따른 상금차별화는 가능한 것이다. principal은 a를 직접 관찰할 수 없기 때문에 최종 결과물인 s를 통해 agent가 어떠한 행동을 취했는지를 확률적으로 유추하면서 이에 적용할 계약을 만들어야 한다. 다음으로 일반적인 principal-agent problem에서 다루는 가정을 살펴보자.

〔가정 1〕

$p_{nm}>0\ \forall n,m$.

agent의 효용함수를 $U(w,a)=u(w)-d(a)$로 표시하고 agent의 reservation level of utility를 u_0로 나타내기로 하자. 여기서 $-d(a)$는 agent가 a라는 행동을 취함에 따라 발생하는 비효용을 말한다.

〔가정 2〕

함수 u는 엄정증가(strictly increasing)함수이며, 연속미분가능(continuously differentiable)이고 오목(concave)함수이다.

그리고 $B(a)$는 principal이 agent가 a를 행동하였을 때 얻게 되는 총효용(total benefit)을 표시한다. 따라서 $B(a)$에서 principal이 agent에게 지불하는 상금의 기댓값을 빼면 principal의

순가치가 된다.

〔가정 3: first-order stochastic increasing assumption〕

$n' > n$인 모든 n', n에 대해 그리고 $m = 1, \ldots, M$에 대해 다음이 성립한다.

$$\sum_{i=m}^{M} p_{ni} \leq \sum_{i=m}^{M} p_{n'i}$$

〔가정 4: the monotone-likelihood ratio property〕

$n' > n$인 모든 n', n과 $m' > m$인 모든 m', m에 대해 다음이 성립한다.

$$\frac{p_{n'm}}{p_{nm}} \leq \frac{p_{n'm'}}{p_{nm'}}$$

〔가정 5: concavity of the distribution function condition〕

$a_{n'} = \beta a_n + (1-\beta) a_{n''}$과 $m = 1, \ldots, M$에 대해 다음이 성립한다.

$$\beta \sum_{i=m}^{M} p_{ni} + (1-\beta) \sum_{i=m}^{M} p_{n''i} \leq \sum_{i=m}^{M} p_{n'i}$$

Ⅳ. 기본모형

principal-agent problem을 풀기 위해 우리는 2 단계로 나누어 최적화문제를 푸는데, 이를 살펴보기로 하자.

〈1 단계〉

$a_n \in A$인 모든 a_n에 대해서 개별적으로 다음 문제를 푼다. 이 문제는 agent가 a_n을 선택하도록 상금체계를 만들되, 최소의 비용으로 이를 가능하게 하는 방안을 찾는 것이다. 먼저 $w(s_m)$을 결과가 s_m로 나왔을 때 principal이 agent에게 지불하는 상금으로 표시하자. 또한 편의상

$$x_m = u(w(s_m))$$

로 표시하고, 이는 agent가 받은 상금을 효용으로 환산한 것이다. $v = u^{-1}$, 즉 v를 u의 역함수로

표시하면, agent가 a_n을 선택하였을 때 principal이 주어야 하는 상금의 기댓값은 다음과 같이 된다:

$$\sum_{m=1}^{M} p_{nm} v(x_m).$$

먼저 agent의 입장에서 a_n을 선택하도록 하기 위해 필요한 조건을 살펴보기로 하자. 먼저 agent의 효용이 적어도 reservation level of utility를 충족하여야 한다.

$$\sum_{m=1}^{M} p_{nm} x_m - d(a_n) \geq u_0.$$

이 조건을 참여조건(participation constraint)이라고 부른다. 두 번째로 필요한 조건은, agent의 입장에서 볼 때 행동 a_n을 취했을 때의 기대효용이 다른 행동을 취했을 때보다 불리하여서는 안된다는 것이다.

$$\sum_{m=1}^{M} p_{nm} x_m - d(a_n) \geq \sum_{m=1}^{M} p_{n'm} x_m - d(a_{n'}) \, \forall n'.$$

이를 인센티브조건(incentive-compatibility constraint)이라고 부른다.

이러한 〈1 단계〉의 문제를 다음과 같이 요약할 수 있다.

$$\text{minimize} \sum_{m=1}^{M} p_{nm} v(x_m)$$

subject to

$$\sum_{m=1}^{M} p_{nm} x_m - d(a_n) \geq u_0 \text{ (Participation Constraint)}$$

$$\sum_{m=1}^{M} p_{nm} x_m - d(a_n) \geq \sum_{m=1}^{M} p_{n'm} x_m - d(a_{n'}) \, \forall n' \text{ (Incentive Compatibility Constraints)}$$

여기서 〈1 단계〉 문제를 풀었을 때 나오는 최적 목적함수값을 $C(a_n)$으로 표시하자.

〈2 단계〉

$a \in A$인 모든 a에 대해서 다음 문제를 푼다:

maximize $\{B(a) - C(a)\}$

V. 예 시

예시를 통해 보다 구체적으로 논리를 전개해 보도록 하자. agent의 행동집합은 $A = \{1,2\}$이고 여기서 1은 '적은 노력'을, 2는 '많은 노력'을 나타낸다. agent가 이들 행동 각각에 대해 느끼는 비효용은 d_1, d_2 $(d_1 < d_2)$로 표시하자. agent의 행동에 따른 결과는 $S = \{1,2\}$이며 결과로써 1은 '보통'이며 2는 '양호'라고 하자. 보다 구체적인 상황 설정으로, 공장의 근로자가 작업을 하는데 근로자의 노력 여하, 즉 정신 집중도에 따라 생산물 중 양호품의 비율이 차이가 난다고 한다. 작업자가 적은 노력을 투입할 경우에 비해 많은 노력을 투입할 경우 양호품이 높게 나올 확률이 크게 된다고 한다.

agent의 효용함수를 $U(w,a) = \sqrt{w} - a$라고 할 때 〈1 단계〉 문제를 만들어 보자. 여기서는 〈1 단계〉의 대상이 되는 노력 수준을 2, 즉 '많은 노력'으로 상정하고 작업자로 하여금 2를 선택하도록 유도하기 위해서 어떻게 상금체계를 구성하여야 하는지를 살펴보자. 수식의 표현을 위해 다음과 같이 정의하자. 즉 결과 m에 대해 agent는 w_m를 principal로부터 상금으로 받게 되는데, 이것을 효용으로 환산한 것을 W_m로 표시하기로 하자.

$$W_m = \sqrt{w_m}.$$

그러면 $u_0 = 0$이라고 가정하면 〈1 단계〉의 문제는 다음과 같이 표현된다.

〈〈원본 문제〉〉

minimize $\sum_{m=1}^{2} p_{2m} W_m^2$

subject to

$$\sum_{m=1}^{2} p_{2m} W_m - d_2 \geq 0 \text{ (Participation Constraint)}$$

$$\sum_{m=1}^{2} p_{2m} W_m - d_2 \geq \sum_{m=1}^{2} p_{1m} W_m - d_1 \text{ (Incentive Compatibility Constraints)}$$

Ⅵ. agent 세분화에 따른 확장

본 절에서는 지금까지의 기본 모형을 확장하여 본다. 앞서 고려한 p_{nm}은 어떤 agent가 n이라는 노력 내지는 행동을 하였을 때 m이라는 결과가 나올 확률을 말한다. 그런데 다수의 agent가 동일한 업무를 수행하는 경우를 살펴보자. 앞서 예를 들었듯이 제조업에서 작업을 수행하는 근로자들을 고려해 보자. 근로자들은 각자의 특성에 의해 노력의 투입에 대한 성과, 즉 양호품 비율이 다양할 수 있다. 작업에 대한 적성이 높은 근로자의 경우, 동일한 노력을 투입하면서도 다른 근로자에 비해 보다 높은 비율의 양호품을 생산할 것이다. 그리고 높은 노력을 투입하는데 발생하는 비효용이 상대적으로 작을 것이다. 그런데 우리가 다루었던 principal-agent problem은 기본적으로 1인의 agent가 있을 경우를 상정한 것이다. 따라서 기본적인 principal-agent problem을 다수의 agent 경우로 확대할 경우 p_{nm}을 근로자 전체의 평균으로 이해하게 된다. 즉 근로자들이 n의 노력을 투입할 경우 발생하는 결과에 대한 확률의 평균치를 나타낸다.

보다 구체적으로 2인의 근로자가 있는 경우를 살펴보자. 여기서 구체적인 분석에 앞서 〔가정 3〕을 이용하여 새로운 용어 정의를 하기로 하자.

〔근로자의 적성〕

근로자 'h'는 근로자 'l'에 비해 양호품을 생산하는 적성이 있다고 하며, 이는 다음을 의미한다. 동일한 수준의 노력 투입을 가정할 때 다음이 성립한다.

$p_{12}^h \geq p_{12}^l$ and $p_{22}^h \geq p_{22}^l$.

$d_2^h \leq d_2^l$.

첫 번째 부등식은 두 근로자가 적은 노력을 투입할 경우 적성의 효과(양호품의 증가)를 나타낸 것이고 두 번째 부등식은 두 근로자가 많은 노력을 투입할 경우 적성의 효과를 나타내는 것이다.

따라서 기본적인 principal-agent problem에서는 $p_{nm} = \frac{1}{2}(p_{nm}^h + p_{nm}^l)$으로 두 근로자의 성과 확률의 평균이라고 할 수 있다. 그러므로 기본적인 principal-agent problem에서 모형화한 것은

실제 다음과 같이 보다 정교하게 표현할 수 있다.

$$\text{minimize} \sum_{m=1}^{2} \frac{1}{2}(p_{2m}^{h}+p_{2m}^{l})\, W_m^2$$

subject to

$$\sum_{m=1}^{2} \frac{1}{2}(p_{2m}^{h}+p_{2m}^{l})\, W_m - \frac{1}{2}(d_2^h+d_2^l) \geq 0 \text{ (Participation Constraint)}$$

$$\sum_{m=1}^{2} \frac{1}{2}(p_{2m}^{h}+p_{2m}^{l})\, W_m - \frac{1}{2}(d_2^h+d_2^l) \geq \sum_{m=1}^{2} \frac{1}{2}(p_{1m}^{h}+p_{1m}^{l})\, W_m - \frac{1}{2}(d_1^h+d_1^l)$$

(Incentive Compatibility Constraints)

여기서 다루려고 하는 것은 principal-agent problem을 적성에 따라 구분되는 근로자별로 적용할 수 있으며, 이 경우 보다 principal에게 많은 효용을 제공하는 상금체계가 가능하여진다는 것이다.

만약 모든 작업을 고적성 근로자(h)에게 맡길 수 있을 정도로 생산용량 문제가 없을 경우 principal은 다음의 〈1 단계〉 문제를 풀고자 할 것이다.

〈〈적성 문제〉〉

$$\text{minimize} \sum_{m=1}^{2} p_{2m}^{h} W_m^2$$

subject to

$$\sum_{m=1}^{2} p_{2m}^{h} W_m - d_2^h \geq 0 \text{ (Participation Constraint)}$$

$$\sum_{m=1}^{2} p_{2m}^{h} W_m - d_2^h \geq \sum_{m=1}^{2} p_{1m}^{h} W_m - d_1^h \text{ (Incentive Compatibility Constraints)}$$

이 문제에서의 최적 목적함수 값은 기본 모형에서의 최적 목적함수 값보다 작거나 같으며 principal은 동일한 기대이익을 얻기 때문에 고적성 근로자를 통해 상금체계를 구성하는 것이 보다 유리하게 된다.

보다 구체적으로 살펴보면, 〈〈원본 문제〉〉에서의 최적해를 W_i^* 라고 할 경우 다음과 같이 잉여분을 계산할 수 있다.

$$R_1 = p_{11}^h W_1^* + p_{12}^h W_2^* - d_1^h$$
$$R_2 = p_{21}^h W_1^* + p_{22}^h W_2^* - d_2^h$$
$$R_3 = p_{11}^l W_1^* + p_{12}^l W_2^* - d_1^l$$
$$R_4 = p_{21}^l W_1^* + p_{22}^l W_2^* - d_2^l.$$

먼저 〈〈원본 문제〉〉 최적해에서 participation constraint는 등호로 성립함을 알 수 있다. 만약 $\alpha = p_{21} W_1^* + p_{22} W_2^* - d_2 > 0$이라면, 우리는 새로운 해

$$W_1^{**} = W_1^* - \alpha, W_2^{**} = W_2^* - \alpha$$

가 participation constraint와 incentive compatibility constraints를 만족시킴을 알 수 있으며 (W_1^*, W_2^*) 보다 더 좋은 해임을 알 수 있다. 따라서 $p_{21} W_1^* + p_{22} W_2^* - d_2 = 0$을 만족하여야 한다. 이는 $R_2 + R_4 = 0$을 의미한다. 적성에 의한 내용 $d_2^h \le d_2^l$과 함께 고려해보면, 이는

$$p_{21}^h W_1^* + p_{22}^h W_2^* - R_2 \le p_{21}^l W_1^* + p_{22}^l W_2^* - R_4.$$

이를 정리하면

$$(p_{21}^h - p_{21}^l) W_1^* + (p_{22}^h - p_{22}^l) W_2^* \le 2R_2.$$

여기에 $p_{21}^h - p_{21}^l = -(p_{22}^h - p_{22}^l)$이고 $W_1^* < W_2^*$을 이용하면 $0 \le R_2$. 따라서 $0 \le R_2$를 이용하여 h agent의 문제를 풀게 되면 $0 \le R_2$로 인한 지불 상금의 감액을 얻을 수 있다. 이는 실제 지불금액 기준으로 볼 때, $\sum_{m=1}^{2} p_{2m}^h W_m^{*2} - \sum_{m=1}^{2} p_{2m}^h (W_m^* - R_2)^2 = \sum_{m=1}^{2} p_{2m}^h (2R_2 W_m^* - R_2^2)$만큼 principal의 기대이익은 증가하게 된다.

Ⅶ. 결 어

본 논문에서는 다수의 근로자가 있을 때 principal-agent problem의 변화가 발생하고 이로 인한 principal 입장에서의 기대 이익 증가를 얻을 수 있음을 살펴보았다. 하지만 이러한 수익 증가는 근로자의 입장에서는 상금 기댓값을 감소시킨다는 것을 알 수 있다.

본 논문의 동기는 big data의 출연으로 거대한 자료의 활용이 가능하여짐에 따라 새로운 기회가 발생하게 된다는 것에 있다. 제조업의 경우에도, 전체 양호품 비율뿐만이 아니라 제품별로 언제, 누구에 의해 생산되었음을 확인할 수 있게 되었다. 이처럼 보다 세밀한 자료가 이용 가능해 짐에 따라 각 근로자는 자신의 실수에 의해 불량이 나오는 것이 밝혀지므로 이로 인해 보다 작업에 집중을 하여야 하게 되었다. 따라서 기업의 입장에서는 개별 근로자별로 작업의 적성을 고려하여 보다 세분화된 분석이 가능해질 것이며 이로 인해 비용절감을 얻을 수 있으며 이로 인해 기업이익의 증가를 기대할 수 있게 되었다. 즉 big data를 통해 principal이 agent에 대해 보다 치밀한 monitoring을 할 수 있게 되었으며, 이로 인해 경제적 혜택을 얻을 수 있음을 보여준다.

참 고 문 헌

1. David M. Kreps, A Course in Microeconomic Theory, Princeton University Press, 1990.