Professor Lars Stole
University of Chicago
April 26, 2019, (v1.10)

Part II: Moral Hazard and Hidden Action

Please do not distribute to anyone outside of class.

1 One-dimensional action model

1.1 Basic program

The basic principal-agent game of moral hazard with hidden actions follows a simple timing:

1. Principal offers agent a contract which is a function mapping outputs to payments, $w : \mathcal{X} \to [\underline{w}, \overline{w}] \subseteq \mathbb{R}$ for $x \in \mathcal{X}$ units of output. In the *discrete* setting, we assume $\mathcal{X} = \{x_1, \ldots, x_n\}$ and in the *continuous* setting, $\mathcal{X} = [\underline{x}, \overline{x}]$ (possibly unbounded, $\mathcal{X} = \mathbb{R}$).
2. Agent accepts or rejects the contract. If he accepts, the agent chooses costly effort, $e \in \mathcal{E}$, $\mathcal{E}$ is a compact subset of $\mathbb{R}_+$; e improves the distribution of x;
3. x is realized (according to distribution $F(x|e)$ and density $f(x|e)$) and the agent is paid $w(x)$ as promised.

Both the principal and agent are expected-utility maximizers. We will assume that the principal is risk neutral ($v''(\cdot) = 0$)

$$V = v(x - w(x)) = x - w(x),$$

and the agent is risk averse ($u''(\cdot) < 0$),

$$U = u(w(x)) - \psi(e),$$

where $\psi(e)$ gives the agent's strictly-convex cost of effort ($\psi(0) = \psi'(0) = 0$). Note: all of the basic results in the risk-neutral principal case generalize to the setting where both the principal and agent are risk averse. Where convenient, we will use the conditional expectation operator instead of an explicit integral where it is understood that expectation is with respect to x using the CDF $F(x|e)$ for a given e. For example, the agent's expected utility from choosing e given contract w is

$$\int_{\mathcal{X}} u(w(x)) f(x|e) dx - \psi(e) = E[u(w(x))|e] - \psi(e).$$

We also assume that the agent has an outside reservation utility of $\underline{U}$. Thus, we assume the agent will accept the contract at stage 2 if

$$\max_e \; E[u(w(x))|e] - \psi(e) \geq \underline{U}.$$

Effort improves (in FOSD sense) the distribution of output. Lastly, we need to be precise about the marginal benefit of e on the distribution of output. For now, we will assume that an increase in e raises output in a first-order stochastic sense. I.e., we assume that

$$\tilde{e} > e \iff F(x|\tilde{e}) \underset{\text{FOSD}}{\succsim} F(x|e), \text{ for all } e \in \mathcal{E}.$$

This will not be sufficient to guarantee increasing incentives contracts, so we will strengthen this assumption below.

The principal optimally chooses an effort $e \in \mathcal{E}$ and a wage schedule, $w(\cdot) : \mathcal{X} \to [\underline{w}, \overline{w}] \subset \mathbb{R}+$ to solve the following full, **unrelaxed program**:

Program 1.

$$\max_{\{e \in \mathcal{E}, w(\cdot) \in [\underline{w}, \overline{w}]\}} \int_{\mathcal{X}} (x - w(x)) f(x|e) dx, \tag{1}$$

subject to

$$\int_{\mathcal{X}} u(w(x)) f(x|e) dx - \psi(e) \geq \underline{U}, \quad \textit{(IR)} \tag{2}$$

$$e \in \arg\max_{e} \int_{\mathcal{X}} u(w(x)) f(x|e) dx - \psi(e), \quad \textit{(IC)}. \tag{3}$$

Before embarking on the optimal incentive contract, consider a few benchmark cases where moral hazard does not prevent achieving the first-best outcome.

- **Case 1: e is contractible.** Suppose that the principal can offer a contract of the form $w(x, e)$, where the wage depends upon output and effort. In this case, there is no hidden-action situation. Because the principal is risk neutral and effort is contractible, it is always optimal for her to offer a contract of the form $w(e)$. (Why: because any wage risk will require that the principal pays the agent an additional risk premium.) For any effort the principal wishes to implement, e, the principal will wish to set the wage as low as possible to satisfy the agent's participation (or individual rationality (IR)) constraint. Hence,

  $$w(e) = u^{-1}(\underline{U} + \psi(e)).$$

  The principal will therefore choose e to solve

  $$\max_{e} E[x|e] - u^{-1}(\underline{U} + \psi(e)).$$

  The necessary first-order condition for this program is

  $$\int_{\mathcal{X}} x f_e(x|e) dx - \frac{d}{de}\left[u^{-1}(\underline{U} + \psi(e))\right] = 0$$

  or

  $$\int_{\mathcal{X}} x f_e(x|e) dx = \frac{1}{u'(w)} \psi'(e).$$

We will refer to this as the first-best effort level and denote it e^{fb}. Notice that the marginal utility of money for the agent impacts the value of e^{fb}. If the agent were risk neutral and $u'(w) = 1$, then the first-best effort would correspond to what the principal would choose if she directly chose effort at her own personal cost of $\psi(e)$.

> **Remark:** If the principal is also risk averse, then it is optimal for the wage to depend upon output so as to optimally share risk between the principal and agent. Specifically, given e, the wage will satisfy
>
> $$\frac{v'(x - w(x))}{u'(w(x))} = \lambda, \quad \text{for all } x \in \mathcal{X},$$
>
> where λ is the positive Lagrange multiplier associated with the agent's IR constraint. (This is sometimes referred to as the Borch rule, referencing an influential paper on risk sharing from the 1960's.) The first-best effort, in turn, would satisfy the more general condition
>
> $$\int_{\mathcal{X}} v(x - w(x)) f_e(x|e^{fb}) dx + \lambda \left(\int_{\mathcal{X}} u(w(x)) f_e(x|e^{fb}) dx - \psi'(e^{fb}) \right) = 0.$$

- **Case 2: No uncertainty.** Suppose that the distribution of x is deterministic (i.e., $F(x|e)$ has a single atom of mass 1). To make this concrete, suppose that $x = X(e)$ is a strictly increasing and differentiable function. This case is very similar to contracting on e because a forcing contract can be chosen that exactly implements any e. In particular, suppose the principal wants to implement $\hat{e}$ and $\hat{x} = X(\hat{e})$.

  $$w(x) = \begin{cases} \bar{w} & \text{if } x \geq \hat{x} \\ -\infty & \text{otherwise,} \end{cases}$$

  where $u(\bar{w}) = \underline{U} + \psi(\hat{e})$. This reduces the problem to case 1 above, where e is contractible. Hence, the principal implements e^{fb}.

- **Case 3: Agent is risk neutral,** $u''(\cdot) = 0$**.** For simplicity assume that $u'(x) = v'(x)$. The first-best level of output can then be implemented by essentially selling the enterprise to the agent in the contract w. Specifically, the principal offers the agent the following *sales contract*:

  $$w(x) = x - \pi + \underline{U},$$

  where π represents the value of the enterprise when run efficiently:

  $$\pi = \max_e \; E[x|e] - \psi(e).$$

  The first-best effort which maximizes the value of the enterprise is

  $$\int_{\mathcal{X}} x f_e(x|e^{fb}) = \psi'(e^{fb}).$$

  If the agent accepts and chooses this effort, she receives (in expectation)

  $$\max_e \; E[x|e] - \psi(e) - \pi + \underline{U} = \underline{U}.$$

  Thus, the agent accepts and all surplus is captured by the principal.

- **Case 4: shifting support.** If the support of x shifts as a function of e, then e may be indirectly contractible and the first-best may be implementable. To see this, suppose that x is distributed uniformly on $[e, e+K]$ with $K > 0$. The principal can implement any $\hat{e}$ by a forcing contract similar to case 2:

$$w(x) = \begin{cases} \bar{w} & \text{if } x \geq \hat{e} \\ -\infty & \text{otherwise,} \end{cases}$$

where $u(\bar{w}) = \underline{U} + \psi(\hat{e})$.

1.2 Two-actions

We know return to our baseline assumption that the principal is risk neutral and the agent is strictly risk averse. We will start with a very simple case in which only two effort levels are available to the agent, e_l and $e_h > e_l$. Associated with each action is a distribution of outputs with non-shifting support, $\mathcal{X}$. We will explicitly denote these as $F_L(x) = F(x|e_l)$ and $F_H(x) = F(x|e_h)$. The agent's cost of the low action is $\psi(e_l) = 0$ and the agent's cost of the high action is $\psi(e_h) = \Delta$.

The principal chooses $w(\cdot) : \mathcal{X} \to [\underline{w}, \overline{w}]$ and the associated effort level, $e \in \mathcal{E} = \{e_l, e_h\}$, to solve the following program:

$$\max_{\{e \in \mathcal{E}, w(\cdot) \in [\underline{w}, \overline{w}]\}} E[x - w(x)|e],$$

subject to

$$E[u(w(x))|e] - \psi(e) \geq \underline{U} \text{ (IR)},$$
$$E[u(w(x))|e] - \psi(e) \geq E[u(w(x))|e'] - \psi(e'),\ e \neq e', \text{ (IC)}.$$

There are two possibilities: the principal finds it optimal to implement $e^* = e_l$ or the principal finds it optimal to implement $e^* = e_h$. If the principal desires to implement e_l, the IC constraint can be ignored because the agent cannot choose an effort lower than e_l. Because incentives are unneeded, it is optimal to insure the agent against wage risk. The expected payoff to the principal, given e_l is implemented, is therefore

$$E[x|e_l] - \bar{w} = E[x|e_l] - u^{-1}(\underline{U}),$$

where $\bar{w}$ is implicitly defined by $u(\bar{w}) = \underline{U} + \psi(e_l) = \underline{U}$. We will assume (to make things interesting), that the principal wants to implement e_h instead. In this case, the principal's program reduces to the choice of $w(\cdot)$ with the lowest expected wage bill that satisfies (IR) and (IC):

$$\min_{\{w(\cdot)\}} E[w(x)|e_h],$$

subject to

$$E[u(w(x))|e_h] - \Delta \geq E[u(w(x))|e_l], \text{ (IC)}$$
$$E[u(w(x))|e_h] - \Delta \geq \underline{U} \text{ (IR)}.$$

Remarks:

- Although this is not a convex programming problem, with a change of variables, it is possible to convert it into one. Specifically, replace the control variable $w(x)$ with $z(x) \equiv u(w(x))$; i.e., $w(x) = u^{-1}(z(x))$. Thus, $z : \mathcal{X} \to [u(\underline{w}), u(\overline{w})]$. The above program can now be written as

$$\min_{\{z(\cdot)\in[u(\underline{w}),u(\overline{w})]\}} E[u^{-1}(z(x))|e_h],$$

subject to

$$E[z(x)|e_h] - \Delta \geq E[z(x)|e_l], \quad \text{(IC)}$$
$$E[z(x)|e_h] - \Delta \geq \underline{U} \quad \text{(IR)}.$$

As written, we are minimizing a strictly convex objective ($u^{-1}(\cdot)$ is strictly convex because $u(\cdot)$ is strictly concave), subject to linear IC and IR constraints. For such convex programs, the Karush-Kuhn-Tucker conditions are typically necessary and sufficient.[1]

- If the set of feasible utilities is not bounded, it is possible that a solution does not exist.[2] In this case, the necessary and sufficient conditions are vacuous. We will see this below in an example made famous by Mirrlees (1975/1999).

Let's suppose that a solution exists for now. The Lagrangian is

$$\mathcal{L} = E[x - w(x)|e_h] + \mu\left(E[u(w(x))|e_h] - E[u(w(x))|e_l] - \Delta\right) + \lambda\left(E[u(w(x))|e_h] - \Delta - \underline{U}\right),$$

where $\lambda \geq 0$ is the multiplier on the IR (participation) constraint and $\mu \geq 0$ is the multiplier on the incentive-compatibility (IC) constraint. It is helpful to return to our use of the densities f (rather than using expectations). To simplify notation, let $f_H(x) = f(x|e_h)$ and $f_L(x) = f(x|e_l)$. The Lagrangian can be written as

$$\mathcal{L} \;=\; \int_{\mathcal{X}} \left(x - w(x) + \lambda(u(w(x)) - \Delta - \underline{U})\right) f_H(x) \;+\; \mu\left(u(w(x))(f_H(x) - f_L(x)) - \Delta\right).$$

[1] You need to exercise some care because $w(\cdot)$ is infinite dimensional if $\mathcal{X}$ is not discrete. See Luenberger, *Optimization by Vector Space Methods*, (1969) for the generalization of KKT conditions to infinite-dimensional spaces. For those of you who have not studied functional analysis, notions of compactness and continuity familiar from n-dimensional Euclidean space are often more difficult to establish in infinite-dimensional spaces. The choice of topology, for example, is often critical; it does not matter in $\mathbb{R}^n$. E.g., a weaker topology (more open sets) makes compactness easier to establish, but continuity harder. For the state of the art techniques for proving existence in a large class of moral hazard and adverse selection models, see Kadan, Reny and Swinkels (*Econometrica*, 2017).

[2] What are sufficient conditions for a solution to our infinite-dimensional convex-linear program? Our constraint set is convex in $z(\cdot)$ (i.e., if the functions z and $\tilde{z}$ satisfy the constraints, then the function $z_\lambda = \lambda z + (1-\lambda)\tilde{z}$ also satisfies the constraints) and it is closed. Given our bounds on the range of z, we can conclude that the constraint set is weakly sequentially compact. Because the objective function is continuous and quasi-convex in z, it is weakly lower semicontinuous. A weakly lower semicontinuous function on a weakly sequentially compact set has a minimum. See Jahn, *Introduction to the Theory of Nonlinear Optimization, 3rd Ed*, 2006, for more details. Thus, a key condition for existence is the bounds on the agent's utility. These bounds are violated in the Mirrlees example below. An alternative approach would be to assume $\mathcal{X}$ is finite so that the likelihood ratios are bounded. In this case, appropriately chosen bounds on the agent's utility are slack and one can demonstrate that a solution exists. See Grossman and Hart (1983).

Notice that we can maximize the integrand choosing $w(\cdot)$ pointwise in x. If, for each x, we choose $w(x)$ to maximize the Lagrangian, then the resulting $w(\cdot)$ must maximize the integrand across $\mathcal{X}$. [Note, we have left implicit the constraints on $w(x) \in [\underline{w}, \overline{w}]$ for clarity. If either of those bounds is constraining, then additional positive multipliers will need to be introduced into the analysis. Also note that this trick does not work if derivatives or integrals of w appear in the program. In those cases, we would have to use techniques from optimal control theory.] The necessary first-order condition for wage optimality is, for every $x \in \mathcal{X}$ such that $w(x) \in (\underline{w}, \overline{w})$ is

$$(-1 + \lambda u'(w(x)))f_H(x) + \mu u'(w(x))(f_H(x) - f_L(x)) = 0,$$

or more simply

$$\frac{1}{u'(w(x))} = \lambda + \mu\left(1 - \frac{f_L(x)}{f_H(x)}\right). \tag{4}$$

Remarks:

- If the wage solution to (4) violates $w(x) \in [\underline{w}, \overline{w}]$, then it is understood that $w(x) = \underline{w}$ for lower-bound violations and $w(x) = \overline{w}$ for upper bound violations.
- Note that $\lambda > 0$ and $\mu > 0$. If $\mu = 0$, then the righthand side of the FOC is independent of x, and therefore the agent's wage would be constant. But then the agent would not exert any effort and $e = e_l$, a contradiction. If $\lambda = 0$, then the agent's IR constraint is slack. In this case, the principal could reduce the agent's wage schedule from $w(x)$ to $w_\varepsilon(x)$ where $u(w(x)) - \varepsilon = u(w_\varepsilon(x))$. This variation reduces expected wage payments without impacting the agent's incentives to choose e_h. A contradiction.
- If the principal were also risk averse, the condition in (4) would need to be modified as
$$\frac{v'(x - w(x))}{u'(w(x))} = \lambda + \mu\left(1 - \frac{f_L(x)}{f_H(x)}\right).$$
Note this is akin to the earlier Borch rule except that now the incentives term involving μ emerges on the righthand side.
- The ratio $\frac{f_H(x)}{f_L(x)}$ is often call the likelihood ratio for e_h. Values greater than 1 imply that the observed x is more likely to have been generated by $f_H(x)$ than $f_L(x)$. Because $u'' < 0$, the ratio $\frac{1}{u'(w)}$ is increasing in w. If we define w_λ by
$$\frac{1}{u'(w_\lambda)} = \lambda,$$
then the optimal wage schedule has the property that
$$w(x) > w_\lambda$$
if $f_L(x)/f_H(x) < 1$ (i.e., f_H is more likely than f_L) and the converse
$$w(x) < w_\lambda$$

if $f_L(x)/f_H(x) > 1$ (i.e., f_L is more likely than f_H). We conclude in the two-effort case that to implement high effort, wages should be higher when the output is more indicative of f_H than f_L. Note: This is not the same as saying $w(x)$ is increasing in output. For such a conclusion ,we will need to strengthen our assumption regarding the likelihood ratio.

- **MLRP.**

  Definition 1. *If $f(x|e)$ is differentiable in e, then we say that the distribution satisfies the* **monotone-likelihood ratio property (MLRP)** *iff for any e,*

  $$\frac{f_e(x|e)}{f(x|e)} \text{ is increasing in } x.$$

  If f is not differentiable, (e.g., e can take on only two types), then we modify our definition to accommodate differences in place of derivatives. In the 2-effort case, we say f satisfies MLRP iff

  $$\frac{f_H(x) - f_L(x)}{f_H(x)} \text{ is increasing in } x.$$

  In both settings, a higher value of x is indicative of a higher effort level.

- Milgrom (1981) shows that an alternative (and more general) definition of MLRP is the following. Suppose that we have a prior distribution about the agent's effort, and upon observing x we form a posterior distribution, denoted $G(e|x)$. MLRP is equivalent to

  $$x > \tilde{x} \implies G(e|x) \underset{\text{FOSD}}{\succsim} G(e|\tilde{x}).$$

  In other words, higher x provides good news about the agent's effort.

- Many distributions satisfy MLRP. For example, if x is distributed normally with mean e, then the distribution satisfies MLRP.

  There are two properties worth noting:

  1. The expectation of f_e/f is zero:

     $$\int_{\mathcal{X}} \frac{f_e(x|e)}{f(x|e)} f(x|e)dx = \int_{\mathcal{X}} f_e(x|e)dx = 0.$$

     Notice that this property is true for any likelihood ratio, and not just those that satisfy MLRP.

  2. MLRP implies FOSD, but not the reverse. To see this, note that f_e/f has an expectation of zero, so it must start negative and end positive (because MLRP requires f_e/f is increasing). Thus, for $\mathcal{X} = [\underline{x}, \overline{x}]$ and any $x < \overline{x}$

     $$\int_{\underline{x}}^{x} \frac{f_e(x|e)}{f(x|e)} f(x|e)dx < 0.$$

Simplifying,

$$\int_{\underline{x}}^{x} \frac{f_e(x|e)}{f(x|e)} f(x|e)dx = \int_{\underline{x}}^{x} f_e(x|e)dx = F_e(x|e) < 0.$$

- FOSD does not imply MLRP. It is straightforward to construct a family of FOSD distributions ordered by effort, e, but which violates MLRP. For example, suppose $x = e + \theta + \varepsilon$ and ε is a normally distributed mean-zero random variable. If θ is some background state that can be either good θ_g or bad θ_b with probability 50-50 and is independent of ε, then the resulting distribution (integrating out θ) is FOSD but may fail MLRP if the difference in the good and bad states exceeds the difference in e_h and e_l. Intuitively, the variation in θ may generate a bimodal distribution of outputs. Higher effort shifts the distribution to the right, and so FOSD holds, but there can be a range of lower outputs that are more likely due to bad times and high effort and a range of higher outputs that are more likely due to good times and low effort.

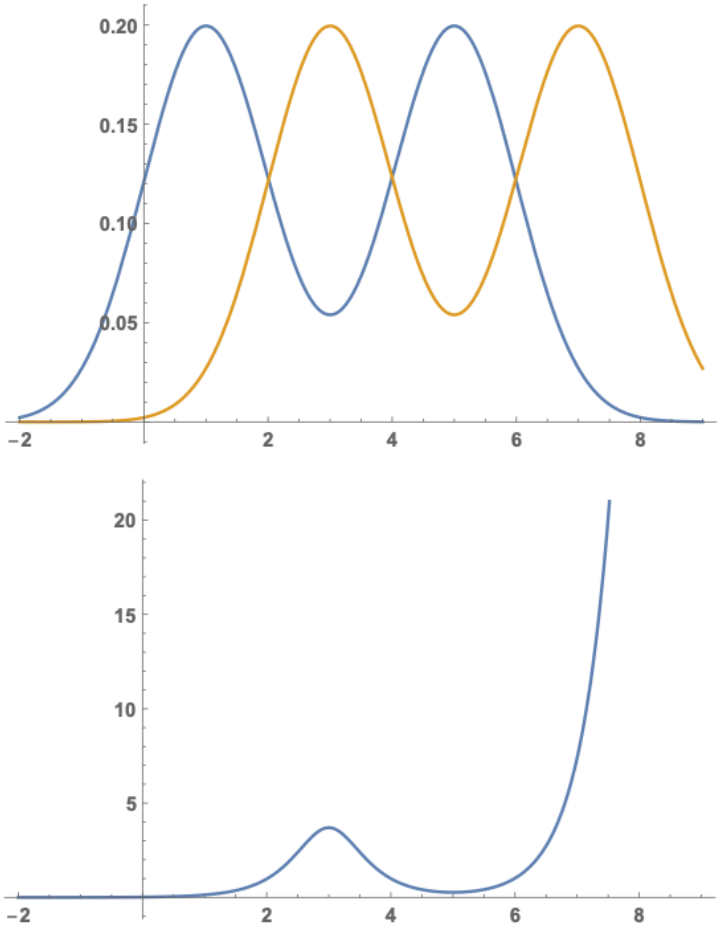

Figure 1: Top panel is a bimodal normal distribution for low effort and for high effort. Bottom panel shows corresponding non-monotone likelihood ratio, f_H/f_L. The low-effort density is the average of two normal distributions with means 1 and 6 (representing the difference in good and bad states equal to 5 and low effort, $e_l = 1$; the high-effort density is the average of two normal distributions with means 4 and 9 (representing again the difference in states of 5 and an effort of $e_h = 4$). All standard deviations are 1.

- If we believe the output distribution conditional on effort satisfies MLRP, then it follows from (4) that the wage schedule is strictly increasing for the two-effort case (and implementing e_h).

Proposition 1. *If $f(x|e)$ satisfies MLRP, then the optimal wage schedule which implements $e^* = e_h$ in the two-effort case (if a solution exists), is increasing in output.*

1.3 Non-existence of an optimal contract.

If (1) the wage function is not required to lie in a compact range (i.e., $w(x) \in \mathbb{R}$), (2) the agent's utility is unbounded from below (e.g., $\lim_{w\to-\infty} u(w) = -\infty$), and (3) if the likelihood ratio is unbounded on $\mathcal{X}$, then an optimal solution to the general program may not exist. Mirrlees (1975/1999) was the first to point out this problem for the continuous-effort

setting. To be clear, the problem is akin to choosing $y \in (0,1)$ to maximize y; a solution does not exist because you can get arbitrarily close to $y = 1$. Mirrlees (1975/1999) shows that one can get arbitrarily close to the first best (although never achieve it) by using a contract that consists of two wage levels and a threshold. Thus, there is an existence problem, and the incentives-insurance conflict is not really an issue in such problems.

em Simple 2-effort example: Here, we construct a simpler example than Mirrlees's that builds on our two-effort model. Before continuing, however, note that the full-information contract in the 2-effort model that implements e_h (assuming $\underline{U} = 0$) forces the action $e^* = e_h$ by setting the wage so that $u(w^{fb}) = \Delta$ (i.e., $w^{fb} == u^{-1}(\Delta)$), and paying this wage only if $e = e_h$.

Now suppose that only x is observable (not e) and $x = e + \varepsilon$ where ε is normally distributed with mean zero and finite variance. In this case, one can show that the likelihood ratio satisfies $\lim_{x\to-\infty} f_L(x)/f_H(x) = +\infty$. As a consequence, it is also the case that $\lim_{x\to-\infty} F_L(x)/F_H(x) = +\infty$ (using L'Hospital's rule). We will use that fact below.

We will now construct a simple contract that can get arbitrarily close to the full-information setting. Specifically, we will construct a sequence of contracts parameterized by k such that along the sequence e_h is incentive compatible and the IR constraint is satisfied with an equality, $E[w(x)|e_h] = \Delta$, but the expected wage converges to $w = u^{-1}(\Delta)$, the full-information wage.

Here is the simple contract parameterized by (w_0, w_1, k):

$$w(k) = \begin{cases} w_1 & \text{if } x \geq k \\ w_0 & \text{if } x < k. \end{cases}$$

The idea of Mirrlees is that one can simultaneously make k and w_0 very low while satisfying IC and IR constraints because the likelihood-ratio property of the normal distribution implies that low output is very, very indicative of low effort. This makes the IC constraint arbitrarily cheap to satisfy, even with a risk averse agent. To see this, consider the wage minimization program for the principal who wants to induce e_h using these simple k-wage contracts (choosing only (w_0, w_1, k)). Incentive compatibility and individual rationality require,

$$(1 - F(k|e_h))u(w_1) + F(k|e_h)u(w_0) - \Delta \geq (1 - F(k|e_l))u(w_1) + F(k|e_l)u(w_0), \quad \text{(IC)},$$

$$(1 - F(k|e_h))u(w_1) + F(k|e_h)u(w_0) - \Delta \geq 0, \quad \text{(IR)},$$

where we have set $\underline{U} = 0$ for simplicity. These expressions can be simplified to

$$(F(k|e_l) - F(k|e_h))(u(w_1) - u(w_0)) \geq \Delta, \quad \text{(IC)},$$

$$u(w_1) - F(k|e_h)(u(w_1) - u(w_0)) \geq \Delta, \quad \text{(IR)}.$$

Suppose that both constraints bind for the principal choosing (w_1, w_0, k). [One can show that this must be true.] Because the principal is free to assign $u(w_0)$ to any value (i.e., w_0 is unconstrained and $u(w_0)$ is unbounded from below), we can substitute $u(w_0)$ out and combine the two constraints into a single constraint which depends only on (w_1, k):

$$\left(\frac{F(k|e_l)}{F(k|e_h)} - 1\right)(u(w_1) - \Delta) = \Delta.$$

Now consider a sequence of k going to $-\infty$. For any k in the sequence, if we choose w_1 according to the above equation (and w_0 is chosen as required in the substitution), we guarantee e_h will be chosen and the contract will be accepted by the agent. Moreover, the expected wage paid to the agent is

$$(1 - F(k|e_h))w_1 + F(k|e_h)w_0 \leq w_1.$$

Notice here, however, that as k is made small (approaches $-\infty$), $\frac{F(k|e_l)}{F(k|e_h)}$ approaches infinity (using the fact about normal distributions above), and thus $u(w_1)$ converges to Δ. Because the expected wage must lie below an upper bound converging to Δ, the expected wage must approach the full-information wage. Hence, we can get arbitrarily close to the full-information solution but never quite get there (i.e., a solution does not exist).

So where did we go wrong in our KT-Lagrangian analysis? Basically, we assumed that a solution existed and applied necessary and sufficient conditions to describe a non-existent solution. In our example, if we try to solve the KKT conditions (i.e., find a wage schedule and multipliers that solve the IR and IC constraints together with (4)) we will fail to find a solution. Practically speaking, if we use numerical methods to solve the problem with an approximation of the normal distribution on a finite grid of outputs (rather than $\mathbb{R}$), we will find a solution for any grid. But as our approximation grid becomes finer and extends further into the left tail, we will see μ becomes smaller and smaller. This is because the likelihood ratio is unbounded for the Normal distribution, but it is bounded on any discrete approximation.

1.4 Informativeness principle

We have thus far assumed that there is a single signal, x, that the principal can use in her contract with the agent. Suppose instead that two outcomes were observed, x and y. When will the principal want to make the wage schedule depend upon both variables?

To answer this question, we can repeat our previous analysis using $f(x, y|e)$ as the joint distribution of (x, y) given effort, e. Suppose the principal wishes to implement e_h rather than e_l. She will solve the following minimization program:

$$\min_{w(\cdot)} \int_{\mathcal{X}} \int_{\mathcal{Y}} w(x, y) f(x, y|e_h) dy dx,$$

subject to

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} u(w(x,y))f(x,y|e_h)dydx - \psi(e_h) \geq \underline{U},$$

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} u(w(x,y))f(x,y|e_h)dydx - \psi(e_h) \geq \int_{\mathcal{X}}\int_{\mathcal{Y}} u(w(x,y))f(x,y|e_l)dydx - \psi(e_l).$$

This is a convex-linear programming problem (after a transformation), and thus the KKT conditions are necessary and sufficient. Assuming a solution exists, for all $w(x) \in (\underline{w}, \overline{w})$ it satisfies the following condition:

$$\frac{1}{u'(w(x,y))} = \lambda + \mu\left(1 - \frac{f(x,y|e_l)}{f(x,y|e_h)}\right).$$

Immediately, we conclude that if

$$\frac{f(x,y|e_l)}{f(x,y|e_h)}$$

is independent of y, then $w(x,y)$ will also be independent of y. This leads us to the following definition.

Definition 2. *The random variable x is a* **sufficient statistic for y with respect to e** *iff there exist functions g and h such that*

$$f(x,y|e) = g(x|e)h(y|x), \quad \textit{for all } x \in \mathcal{X}, y \in Y, \textit{ and } e \in \mathcal{E}.$$

It is immediate from the definition that w will not depend upon y if x is a sufficient statistic for y with respect to e:

$$\frac{f(x,y|e_l)}{f(x,y|e_h)} = \frac{g(x|e_l)h(y|x)}{g(x|e_h)h(y|x)} = \frac{g(x|e_l)}{g(x|e_h)}.$$

This result, due to both Holmström (1979) and Shavell (1979) is sometimes referred to as the informativeness principle:

> **Informativeness principle.** An additional signal y is valuable for incentives if and only if it carries additional information about the agent's effort that is not contained in x (i.e., x is not sufficient for y with respect to e).

Remarks:

- Randomization is not optimal. An implication of the informativeness principle is that introducing randomness into wages is never optimal because the randomizing device does not convey additional information about effort.[3]

[3]This no-randomization result relies on our assumption that the agent's preferences are additively separable in effort and wage. Gjesdal (1982) has an example where randomization improves incentives when effort and risk aversion interact.

- Relative performance evaluation (i.e., making the wage of one agent depend in part on the outcome of another agent) is optimal when the second agent's output is informative about the first agent's effort. For example, if a common shock hits the outputs of both agents, some form of relative performance evaluation will be optimal. If, however, the outcomes of the two agents are statistically independent, there is no gain from relative performance evaluation. See Holmström (1982) for the generalization of the informativeness principle to multi-agent settings.

1.5 General case and validity of the first-order approach

We would like to return to our more general problem in which e is chosen from the interval $[0, \bar{e}]$, the marginal benefit of effort is captured by the marginal effect on the output density function, $f_e(x|e)$, and the marginal cost of effort is characterized by the derivative of the agent's cost function, $\psi'(e)$ (where $\psi'(0) = 0$ and $\lim_{e\to\infty} \psi'(e) = \infty$). Recall the general program (1):

$$\max_{\{e\in[0,\bar{e}],w(\cdot)\in[\underline{w},\overline{w}]\}} \int_{\mathcal{X}} (x - w(x)) f(x|e) dx,$$

subject to

$$\int_{\mathcal{X}} u(w(x)) f(x|e) dx - \psi(e) \geq \underline{U}, \quad \text{(IR)}$$

$$e \in \arg\max_{e} \int_{\mathcal{X}} u(w(x)) f(x|e) dx - \psi(e), \quad \text{(IC)}.$$

1.5.1 The first-order approach

The (IC) constraint is difficult to work with, so we will proceed by using the weaker constraint that the agent's first-order condition is satisfied for e:

$$\int_{\mathcal{X}} u(w(x)) f_e(x|e) dx - \psi'(e), \quad \text{(IC-FOC)}. \tag{5}$$

Thus, the FOA program is to maximize $E[x - w(x)|e]$, subject to (IR) and (IC-FOC).

There is a serious problem with this approach that was noted by Mirrlees (1975). If the optimal incentive contract in the FOA (first-order approach) program does not provide global incentives for the agent's choice of e, then it is possible that the FOC constraint is satisfied at either a minimum or a local (but not global) maximum for the agent's program. Thus, even if the solution to the FOA program satisfies the agent's local SOC, it is still problematic. Imposing a local SOC condition as an additional constraint will not solve all of the potential problems because local maxima can be selected. Without more structure, the solutions to the relaxed FOA program may not contain the true solution to the unrelaxed program. In short, *the solution set to the FOA program may not contain the solution to the general program in (1).*

We will return to the question of when the FOA program is valid, but for now we will proceed as if it is in valid (and we will assume a solution exists) in order to see what aspects of the 2-effort model are general.

We form the Lagrangian using the agent's FOC condition:

$$\mathcal{L} = \int_{\mathcal{X}} (x - w(x)) f(x|e) dx + \lambda \left(\int_{\mathcal{X}} u(w(x)) f(x|e) dx - \psi(e) \right) + \mu \left(\int_{\mathcal{X}} u(w(x)) f_e(x|e) dx - \psi'(e) \right).$$

Maximizing $w(x)$ pointwise, we obtain the analogue of (4) which is valid providing $w(x) \in (\underline{w}, \overline{w})$:

$$\frac{1}{u'(w(x))} = \lambda + \mu \frac{f_e(x|e)}{f(x|e)}, \quad \text{for all } x \in \mathcal{X}. \tag{6}$$

Given that f satisfies MLRP, the optimal wage schedule in the FOA program will be increasing in output, assuming that $\mu > 0$. Holmström (1979) proves that $\mu > 0$ for the general case with a risk-averse principal, *assuming the first-order approach is valid*. Jewitt (1988) noted that if the principal is risk neutral (our present assumption), there is a more direct proof.

Lemma 1. *In the FOA program, if the optimal effort is positive and $w(x) \in (\underline{w}, \overline{w})$, then $\mu > 0$.*

Proof: Given that $e > 0$, the wage cannot be constant and thus $\mu \neq 0$. Rearranging (6) and dividing by $\mu \neq 0$, we can write

$$f_e(x|e) = \left(\frac{1}{u'(w(x))} - \lambda \right) \frac{f(x|e)}{\mu}.$$

Substituting into the agent's FOC, we have

$$\int_{\mathcal{X}} u(w(x)) \left(\frac{1}{u'(w(x))} - \lambda \right) f(x|e) dx = \mu \psi'(e). \tag{7}$$

But because the expectation of $f_e(x|e)/f(x|e)$ is zero, it follows that $E[1/u'(w(x))] = \lambda$. Hence, we can reinterpret (7) as

$$\text{Cov} \left(u(w(x)), \frac{1}{u'(w(x))} \right) = \mu \psi'(e).$$

Because u and $1/u'$ are both increasing in w, the covariance must be positive. Because $\psi'(e) > 0$, it follows that $\mu > 0$. □

Remark: Indeed, there is an even more direct proof than Jewitt's but requires MLRP: if the solution to the FOA program has a positive effort level, $e > 0$, then μ cannot be nonpositive because otherwise (6) and MLRP would imply the wage schedule is either constant or everywhere nonincreasing. The result in either case would be the agent chooses $e = 0$ – a contradiction.

- Wages increase in output. Having established that $\mu > 0$ in the FOA program, we can conclude that the optimal FOA wage schedule is increasing in output if MLRP is satisfied.
- The informativeness principle. Our previous conclusions regarding the optimality of using additional signals also carries over to the general case (assuming the FOA is valid): wages will not depend upon a signal y iff x is sufficient for y with respect to e.

1.5.2 Sufficient conditions for the validity of the first-order approach

This is the disappointing part of the first-order approach. It is difficult to find reasonable, sufficient conditions for the FOA program. The most well known set of conditions require both MLRP and CDFC.

Definition 3. *A distribution $F(x|e)$ satisfies the* **convex distribution function condition** *iff*

$$F(x, \gamma e + (1-\gamma)e') \leq \gamma F(x|e) + (1-\gamma)F(x|e'), \quad \textit{for all } x \in \mathcal{X}, \textit{for all } e, e' \in \mathcal{E}.$$

If F is differentiable, the condition is more simply stated as $F_{ee}(x|e) \geq 0$. It is hard to think of many distributions satisfying MLRP and CDFC. One class of densities that have been singled out by Grossman and Hart (1983) are those that satisfy the **spanning condition**:

Definition 4. *Suppose that $f_1(x)$ and $f_0(x)$ are two densities where $f_1(x)/f_0(x)$ is nondecreasing for all x. Then the density*

$$f(x|e) = \gamma(e)f_1(x) + (1-\gamma(e))f_0(x), \quad e \in [0,1],$$

satisfies the **spanning condition**.

If $f(x|e)$ satisfies the spanning condition and $\gamma(e)$ is increasing, then it satisfies MLRP. If $\gamma(e)$ is also concave in e, then it satisfies CDFC.

Another class of contrived distributions that satisfies MLRP and CDFC is

$$F(x|e) = G(x)^{h(e)},$$

where G is a continuously differentiable cumulative distribution function and $h(e)$ is a continuous differentiable, concave function.[4] As a specific example,

$$F(x|e) = \left(\frac{x - \underline{x}}{\overline{x} - \underline{x}}\right)^{\frac{1}{1-e}}, \quad e \in [0,1).$$

[4] To verify, note that

$$\frac{\partial^2}{\partial e^2} F(x|e) = F(x|e)\log(G(x))\left(\log(G(x))h'(e)^2 + h''(e)\right) > 0,$$

and

$$\frac{f_e(x|e)}{f(x|e)} = \frac{(1 + h(e)\log G(x))h'(e)}{h(e)}$$

is increasing in x.

We can now state the result, originally stated by Mirrlees (1977) and later proven more carefully by Rogerson (1985).[5] Rogerson (1985) proves the sufficiency result for the case in which $\mathcal{X}$ is discrete. Because we want to show the result is true for the case where $\mathcal{X}$ is an interval on the real line, we'll need to add to MLRP and also require that $f_e(x|e)/f(x|e)$ is *continuous* in x.

Proposition 2. *Suppose $\mathcal{X} = [\underline{x}, \overline{x}]$, and $f_e(x|e)/f(x|e)$ is continuous on $\mathcal{X}$ for any $e \in [0, \overline{e}]$. If $f(x|e)$ satisfies MLRP and CDFC, then the solution to the FOA program is a solution to the general program.*

Proof: First, we note that in our setting with a risk neutral principal (6), combined with MLRP and the continuity of f_e/f, implies that the FOA wage schedule is continuous and nondecreasing (and therefore almost everywhere differentiable, allowing us to integrate by parts). As such, we can write the agent's payoff as

$$\begin{aligned}\int_{\mathcal{X}} u(w(x))\ f(x|e)dx - \psi(e) &= u(w(x))F(x|e)|_{\underline{x}}^{\overline{x}} \\ &\quad - \int_{\mathcal{X}} u'(w(x))\frac{dw(x)}{dx}F(x|e)dx - \psi(e) \\ &= u(\overline{w}) - \int_{\mathcal{X}} u'(w(x))\frac{dw(x)}{dx}F(x|e)dx - \psi(e).\end{aligned}$$

Differentiating this with respect to e twice yields

$$-\int_{\mathcal{X}} u'(w(x))\frac{dw(x)}{dx}F_{ee}(x|e)dx - \psi''(e) < 0,$$

for every $e > 0$. Thus, the agent's second-order condition is globally satisfied in the FOA program, hence the solution to the FOA program is a solution to the general program. □

Remarks:

- Because Rogerson (1985) provides a corrected proof to Mirrlees's original statement, the result is often referred to as the Mirrlees-Rogerson sufficient conditions.
- Jewitt (*Econometrica*, 1988) is able to weaken the CDFC condition by placing restrictions on the agent's utility function, but these conditions are restrictive in different ways.

[5] In Mirrlees's unpublished (1977) paper, at one step in the proof MLRP and $\mu > 0$ is used to prove the optimal wage schedule is increasing. In a later step, the increasing wage schedule is used to show the agent's program is concave. Unfortunately, the first step arguing that w is increasing used the result that $\mu > 0$ in (6), but the result that $\mu > 0$ relied on the assumption that the FOA is valid. Hence, the proof is circular. Rogerson (1985) addressed this by considering a doubly-relaxed program where the FOC condition is replaced by the inequality $\frac{d}{de}E[u(w(x))|e] \geq \psi'(e)$. Because it is an inequality constraint, the Kuhn-Tucker multiplier must be nonnegative. Hence, MLRP and the doubly-relaxed optimality condition implies that the doubly-relaxed wage schedule is increasing. Now the second part of Mirrlees's proof can be applied. In our setting in which the principal is risk neutral, we already showed that $\mu > 0$ can be proved directly without relying on the validity of FOA approach.

- Grossman and Hart (1983) find the optimal wage contract in a general setting where $\mathcal{X}$ is discrete and do not use the first-order approach but, rather, directly solve the unrelaxed program. Consistent with Rogerson (1985), they find that MLRP and CDFC imply the optimal incentive contract exists and is increasing in output.
- For recent results related to the first-order approach applied to more general settings including multidimensional signals and efforts, see Kirkegaard (*Theoretical Economics*, 2017).
- The fact that most of these FOA models assume such conditions as spanning or CDFC (in addition to MLRP) in order to get implications about the monotonicity of the wage schedule – something that seems obvious in the real world – is troubling. Some recent research has attempted to find solution methods that do not rely on the typical stronger conditions and instead directly confront the problems of local maxima that were originally noted by Mirrlees (1975/1999). See, for example, Ke and Ryan (*Theoretical Economics*, 2018).
- Given the difficulty of finding assumptions and models that generate monotone wages that resemble what we see in the real world, some believe that perhaps we are focusing on the wrong principal-agent model. This is the primary motivation behind Holmström and Milgrom's (1987) work, which we turn in a later part of these lectures.

1.6 Limited liability constraints

In some circumstances, it makes sense to model the agency problem as one where the agent is risk neutral, but with limited liability. For example, suppose that the agent cannot be given a negative wage (i.e., their liability to the firm is limited and the worst that the firm can do is to pay them zero, $w = 0$). Effectively, this is the same situation as is if the agent were infinitely risk averse at $w = 0$:

$$u(w) = \begin{cases} w & \text{if } w \geq 0 \\ -\infty & \text{if } w < 0. \end{cases}$$

Note that such an agent is risk averse over negative incomes. E.g., $\frac{1}{2}u(1) + \frac{1}{2}u(-1) < u(0)$. With such an agent, the optimal contract will satisfy the constraint $w \geq 0$, so the limited liability is capturing something akin to risk aversion. Mathematically, it is often easier to work with than standard models of risk aversion, so it is a common modeling assumption that still captures many aspects of the standard risk-aversion model. Indeed, in the context of corporate finance, it is quite common to assume that borrowers are risk neutral but also liquidity constrained and unable to pay back their debt. We will turn to a well known paper (Innes (1990)) using this device below.

One critical difference between the standard risk-aversion model and limited-liability should be made clear. In the standard model, the IR constraint binds. Recall that if it didn't, the principal could offer a variation, $w_\varepsilon(x) < w(x)$ defined by $u(w(x)) - \varepsilon = u(w_\varepsilon)$, and still satisfy the IR constraint for ε small. Such a variation lowers costs for the principal

without impacting the IC constraint. Hence, the IR constraint cannot generally be slack. With limited liability, however, this variation is not allowed. For example, if $w(\cdot) \geq 0$ is required and $w^*(x) = 0$ for some x, then it is not possible to lower the wage further for that x. Hence, the previous argument establishing that the IR constraint must bind is no longer valid.

1.7 Application: Optimality of debt contracts

This section is based on Innes, "Limited liability and incentive contracting with ex-ante action choices," *Journal of Economic Theory*, 1990. Among other things, Innes (1990) applies the ideas of optimal incentive design (and the power of MLRP) to demonstrate circumstances in which a debt contract is the optimal security for an entrepreneur to use to raise funds and to simultaneously induce high-powered incentives for effort (by the entrepreneur).

Here is the setting. A risk-neutral entrepreneur has an idea or project that requires both outside investment, I, and her supply of effort, e, to generate positive net value. Specifically, assume that the payoff of the project is $x \in [0, \overline{x}]$ and the payoff is distributed according the density $f(x|e)$ which satisfies MLRP. The entrepreneur may exert effort at personal cost $\psi(e)$.

The entrepreneur designs the capital-market repayment contract, $r(x)$. There are three constraints.

- Limited liability. The entrepreneur offers a repayment contract, $r(x)$, which requires payment $r(x)$ if the project payoff is x. It is assumed that the entrepreneur has no personal funds and must repay out of the project returns. Thus, $r(x) \leq x$. Moreover, the capital market is unwilling to pay more into the project after the initial investment, I. Thus, $x \geq r(x) \geq 0$ for all $x \in \mathcal{X}$.
- The entrepreneur's effort is optimally determined by the entrepreneur, given the repayment contract $r(x)$. Specifically, the entrepreneur chooses e to maximize $E[x - r(x)|e] - \psi(e)$. The market understands this and forms correct expectations for e given $r(\cdot)$.
- The capital market needs to be paid back I. Specifically, given the effort e induced by $r(\cdot)$, the capital market's requirement is $E[r(x)|e] \geq I$. Because there is capital-market competition by investors, this constraint will be met with an equality in equilibrium.

Here's the Entrepreneur's program:

$$\max_{r(\cdot), e \in [0,\overline{x}]} \int_0^{\overline{x}} (x - r(x)) f(x|e) dx - \psi(e),$$

subject to

$$x \geq r(x) \geq 0, \ \forall \ x \in [0, \overline{x}],$$

$$e \in \arg\max_{e \in [0,\overline{e}]} \int_0^{\overline{x}} (x - r(x)) f(x|e) dx - \psi(e),$$

$$\int_0^{\overline{x}} r(x) f(x|e) dx = I.$$

To focus our analysis on the interesting cases, we make a few technical assumptions.

Assumption 1. *In the optimal repayment program, the following three conditions are assumed:*

1. *there exists a solution to the constrained program (i.e., there is some repayment schedule, $r(x)$, and induced effort, e, that generates an expected payment of I) and the value of the program to the entrepreneur is positive (i.e., the entrepreneur prefers to pursue the project rather than not);*
2. *the optimal repayment contract cannot implement the first-best effort;*[6]
3. *the function*

$$\int_z^{\overline{x}} x f(x|e) dx - \psi(e),$$

is strictly concave in e for any $z \in (0, \overline{x})$ (this is weaker than CFDC) and attains a unique maximum for some $e \in (0, \overline{e})$.

We proceed according to the first-order approach and check that the optimal contract derived in the relaxed program is also globally concave. Assumption 1.3 guarantees this. The relaxed program is

$$\max_{r(\cdot), e \in [0,\overline{x}]} \int_0^{\overline{x}} (x - r(x)) f(x|e) dx - \psi(e),$$

subject to

$$x \geq r(x) \geq 0, \ \forall \ x \in [0, \overline{x}],$$

$$\int_0^{\overline{x}} (x - r(x)) f_e(x|e) dx - \psi'(e) \geq 0,$$

$$\int_0^{\overline{x}} r(x) f(x|e) dx \geq I.$$

For simplicity (i.e., to avoid other multipliers), we form the Lagrangian ignoring the constraints on $r(x) \in [0, x]$ which we manually impose later:

$$\mathcal{L} = \int_0^{\overline{x}} (x - r(x)) f(x|e) dx - \psi(e) + \mu \left(\int_0^{\overline{x}} (x - r(x)) f_e(x|e) dx - \psi'(e) \right) + \lambda \left(\int_0^{\overline{x}} r(x) f(x|e) dx - I \right). \quad (8)$$

[6] Innes (1991) considers the first-best case separately and shows it shares the same "live-or-die" contract features as the second-best contract.

Rearranging this Lagrangian, we obtain

$$\mathcal{L} = \int_0^{\overline{x}} r(x) \left[\lambda - \mu \frac{f_e(x|e)}{f(x|e)} - 1 \right] f(x|e) dx + \int_0^{\overline{x}} x \left[1 + \mu \frac{f_e(x|e)}{f(x|e)} \right] f(x|e) dx - \psi(e) - \mu \psi'(e) - \lambda I. \quad (9)$$

Notice that the Lagrangian is linear in $r(x)$, and $r(x)$ only appears in the first term. Now consider the constraint that $r(x) \in [0, x]$. If

$$\left[\lambda - \mu \frac{f_e(x|e)}{f(x|e)} - 1 \right] > 0,$$

then $r(x)$ should be increased as large as possible; hence $r(x) = x$; if the bracketed expression is negative, then $r(x) = 0$ is optimal:

$$r^*(x) = \begin{cases} x & \text{if } \lambda \geq 1 + \mu \frac{f_e(x|e)}{f(x|e)} \\ 0 & \text{otherwise.} \end{cases}$$

Because we used a nonnegative FOC constraint in our Lagrangian, we are assured that $\mu \geq 0$. Innes (1990) shows that if $\mu = 0$ in the solution to the relaxed program, then $e^* = e^{fb}$. By assumption we have ruled out this possibility, and so we assume that $\mu > 0$. Because of MLRP, for any $\mu > 0$ and λ, there will exist a unique project value, $\hat{x}$, such that

$$r^*(x) = \begin{cases} x & \text{if } x \leq \hat{x} \\ 0 & \text{if } x > \hat{x}. \end{cases}$$

Innes refers to such a contract as a "live or die" contract. Such a contract gives maximal incentives to the entrepreneur to exert effort (though still falling short of first best under our assumption 1.2). To determine $\hat{x}$ and e^*, we use the two remaining constraints which, given the structure of r^*, are

$$\int_0^{\hat{x}} x f(x|e^*) dx = I,$$

$$\int_{\hat{x}}^{\overline{x}} x f_e(x|e^*) dx = \psi'(e^*).$$

(Notice our earlier assumption guarantees the FOC is sufficient for effort optimality given our live-or-die repayment contract.)

Remarks:

- This contract is *not* debt. It repays everything for $x \leq \hat{x}$, but for $x > \hat{x}$ *it pays absolutely nothing to the investors*. This is optimal given MLRP because it concentrates the residual income, $x - r(x)$, in the right tail where MLRP gives stronger incentives.

- The intuition for why the "live-or-die" contract is optimal is entirely driven by MLRP. The entrepreneur only wants to give the investors I back in expectation. Removing these funds from the entrepreneur's payoffs reduces her incentive to exert effort. Thus, it is best to repay I using funds that would have given low-powered incentives – the payoffs in the left tail.
- Notice also that this contract is non-monotonic. Indeed, if the entrepreneur privately observed $x = \hat{x} - \varepsilon$ prior to it becomming public, the entrepreneur would be tempted to borrow money from friends and family to secretly boost output just above $\hat{x}$. If the repayment schedule must guard against such opportunism by the entrepreneur, then the repayment schedule should be non-decreasing. His motivates Innes's (1991) separate analysis of monotonic repayment contracts, to which we now turn.

1.7.1 Optimal monotonic repayment contracts.

We now impose the requirement that $r(x)$ is nondecreasing. This could be done by with control-theory techniques using the nonnegative margins of $r(x)$ as the control variables. Instead, Innes (1990) provides an indirect proof regarding the optimality of debt showing that any non-debt contract can be improved with a debt contract.

Here is Innes's (1990) argument for why the optimal monotonic contract must be debt: $r(x) = \min\{x, D\}$. Suppose to the contrary that $r^*(\cdot)$ is the optimal monotonic repayment contract and it implements e^* in equilibrium. Now consider the debt contract that repays investors I under the assumption that effort is unchanged, $e = e^*$. Specifically, we choose the face value of the debt, D, so that

$$\int_0^{\overline{x}} r^*(x) f(x|e^*) dx = I = \int_0^{\overline{x}} r^D(x) f(x|e^*) dx.$$

Innes (1990) shows, through a series of lemmas, that this is a superior contract to r^*, yielding a contradiction to its optimality.

> **Step 1.** *The contract $r^D(\cdot)$ will generate a higher effort level $e^D > e^*$.* (See Innes (1990), Lemma 2, for details). The key is to note that debt has the maximum slope of 1 for $x < D$ and zero slope for $x > D$. Hence, for any monotonic $r^*(x)$ that repays the same amount under $f(x|e^*)$, it must be that there is a maximal $\tilde{x}$ such that $r^D(x) \geq r^*(x)$ for all $x \leq \tilde{x}$ and $r^D(x) < r^*(x)$ for all $x > \tilde{x}$.

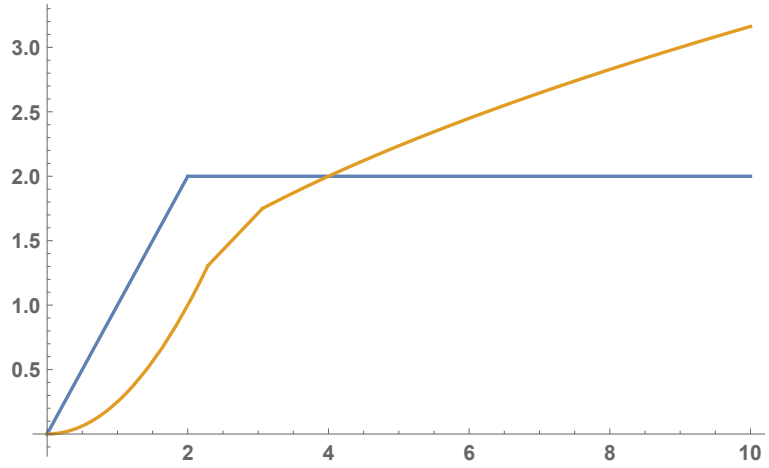

Figure 2: In the figure, $\mathcal{X} = [0, 10]$, $D = 2$, and $\tilde{x} = 4$. Note that the expectations of each curve using $f(x|e^*)$ must be equal, so there must be positive area between r^D and r^* to the left of $\tilde{x}$ and negative area to the right, and the expected values of these areas must be equal.

Because debt shifts repayments to the left, the residual income, $x - r(x)$ is shifted to the right and $MLRP$ guarantees the entrepreneur's effort will increase.

Step 2. Under a debt contract, integration by parts reveals

$$\frac{\partial}{\partial e}\int_0^{\overline{x}} \max\{x, D\} f_e(x|e)dx = -\int_0^D F_e(x|e)dx > 0;$$

consequently, a higher effort level implies

$$\int_0^{\overline{x}} r^D(x) f(x|e^D)dx > \int_0^{\overline{x}} r^D(x) f(x|e^*)dx = I,$$

and investors are better off with the debt contract. Because the entrepreneur could have chosen e^* and obtained the same payoffs as under r^*, but instead prefers to choose $e^D > e^*$, it must be that the entrepreneur's payoff increases:

$$\int_0^{\overline{x}} (x - r^D(x)) f(x|e^D)dx > \int_0^{\overline{x}} (x - r^*(x)) f(x|e^*)dx.$$

Thus, replacing r^D with a debt contract is improving – a contradiction to the optimality of (r^*, e^*).

Innes (1990) concludes that the optimal repayment schedule for entrepreneurial incentives, given the monotonicity constraint, is debt.

Remarks:

- Innes (1990) shows that if one is only worried about the moral hazard problem, and if effort impacts the distribution of profit in a MLRP manner, then debt is optimal in the class of nondecreasing repayment contracts.

- In an argument that goes back to at least Jensen and Meckling (1976), debt has bad consequences for risk shifting. Specifically, if the entrepreneur can choose a riskier distribution of returns (even with constant mean), she will generally choose riskier projects because debt eliminates her downside risk. As a consequence, problems of risk shifting seem to argue for the use of equity rather than debt financing.
- Herbert (*REStud*, 2018) has a general model of moral hazard that includes both kinds of moral hazard (shirking in effort and shifting risk). His model allows the entrepreneur to choose a probability distribution at some convex cost which is modeled as the divergence from an initial discrete probability distribution. He shows that when the entrepreneur can change the distribution in this general way (not just MLRP or variance), then debt is again the optimal contract. The argument is less intuitive than Innes (1990) or Jensen and Meckling (1976), as it uses variational techniques applied to a family of divergence measures. Given the generality of the model, the result is quite remarkable.

2 Dynamic agency and foundations for linear contracts

The optimal solution in the basic principal-agent model is arguably complex because there is an imbalance between the agent's one-dimensional effort space and the infinite-dimensional control space of the principal. The goal of Holmström and Milgrom (1987) is to construct a dynamic model of agency that allows the agent to respond to the history of production over an interval of time in such a way that the principal's optimal contract becomes linear in some sense.

Dynamic setting: What follows is a simplified version of Holmström and Milgrom (1987) where we restrict attention to the case of a risk-neutral principal.

- there are T periods without discounting; in each period the agent chooses an action which is a probability distribution over $n+1$ outcomes, $\phi = (\phi_0, \phi_1, \ldots, \phi_n) \in \Phi \subsetneq \Delta^n$; Φ is a compact full-dimensional subset, interior to the n-dimensional simplex. The associated cost to the agent is $c(\phi)$ (differentiable, strictly convex). The resulting outcome in each period t is denoted $x_t \in \{x_0, x_1, \ldots, x_n\}$ and arises with probability $\phi = (\phi_0, \ldots, \phi_n)$.
- the principal offers a wage that is paid in date $t = T$ that depends upon the sequence of outputs observed: $w(x_1, \ldots, x_T)$;
- the risk-neutral principal maximizes the expected value of

$$\left(\sum_{t=1}^{T} x_t\right) - w(x_1, \ldots, x_T);$$

- the agent is risk averse and has CARA utility, and maximizes the expected value of

$$u\left(w - \sum_{t=1}^{T} e_t\right) = -e^{-r\left(w - \sum_t c(\phi_t)\right)}.$$

Remarks:

- In each period, the agent has observed the previous history of outputs, $X^{t-1} = (x_0, \ldots, x_{t-1})$ and can base the choice of effort, ϕ_t, on this information. Thus, for a given incentive contract, $w(\cdot)$, an effort strategy will be a mapping from X^{t-1} to ϕ_t that maximizes the agent's expected utility going forward.
- The agent's preferences are very special and have been chosen so that the the principal's and the agent's programs are stationary from one period to the next. Because the cost of effort is treated like money, the agent's CARA utility implies that outcome realizations that change the agent's expected wage do not generate wealth effects. Instead, the marginal rate of substitution between money and effort is constant, and the degree of risk aversion is constant. Here's the critical implication: if the wage function were separable $w(x_1, \ldots, x_T) = \sum_{t=1}^{T} w_t(x_t)$, the agent's current period effort would be independent of history. This can best be seen by looking at the one-period model.

2.1 One-period model

The principal solves

Program 2.

$$\max_{w(\cdot), \phi \in \Phi} \sum_{i=0}^{N} \phi_i (x_i - w(x_i)) \quad \textit{subject to}$$

$$e \in \arg\max_{e} \sum_{i=0}^{N} \phi_i u(w(x_i) - e),$$

$$\sum_{i=0}^{N} \phi_i u(w(x_i) - c(\phi)) \geq u(\underline{w}),$$

where $\underline{w}$ is the certainty equivalent of the agent's outside opportunity.

Given our assumption of CARA utility, we have the following result.

Lemma 2. Suppose that $(w^*(\cdot), \phi^*)$ solves the principal's one-period program for some $\underline{w}$. Then $(w^*(\cdot) + \underline{w}' - \underline{w}, \phi^*)$ solves the program for an outside certainty equivalent of $\underline{w}'$.

Proof: Because utility is exponential,

$$\sum_{i=0}^{N} \phi_i u(w_i^* - c(\phi^*)) = -u(\underline{w} - \underline{w}') \sum_{i=1}^{N} \phi_i u(w_i^* + \underline{w}' - \underline{w}).$$

Thus, ϕ^* is still incentive compatible and the IR constraint is satisfied for $u(\underline{w}')$. Similarly, given the principal's utility is exponential, the optimal choice of ϕ^* is unchanged. □

A second result (Theorem 3 in Holmstrom, Milgrom (1987)) is that, under some minor assumptions on $c(\phi)$, if ϕ is implementable in the one-period model, then there is a unique wage schedule that satisfies IC and IR with certainty equivalent wage $\underline{w} = 0$.

Lemma 3. For any ϕ interior to Φ, either (1) ϕ cannot be implements, or (2) the implementing scheme with any certain equivalent $\underline{w}$ is unique.

Consult Holmstrom and Milgrom (1987) for the proof, but the idea is that any solution to the n first-order conditions fully determines the wage schedule, up to the certainty equivalent. We will denote this optimal wage function for certainty equivalent of $\underline{w} = 0$ as $\tilde{w}(\cdot|\phi)$.

2.2 T-period model

Now consider the multi-period problem where the agent chooses an effort each period after having observed the history of outputs up until that time. Let superscripts denote histories of variables; i.e., $X^t = \{x_1, \ldots, x_t\}$. The agent gets paid at the end of the period, $w(X^T)$ and has a combined cost of effort equal to $\sum_t c(\phi_t)$. Thus, the agent's final utility given wage w and effort sequence $\phi^T = (\phi_1, \ldots, \phi_T)$ is

$$u(w, e^T) = -e^{-r\left(w - \sum_t c(\phi_t)\right)}.$$

Because the agent observes X^{t-1} before deciding upon ϕ_t, for a given wage schedule we can write the agent's strategy as $\phi_t(X^{t-1})$. We want to first characterize the wage schedule which implements an arbitrary sequence of efforts, $\{\phi_t(X^{t-1})\}_{t=1,\ldots,T}$. We use dynamic programming to this end.

Let U_t be the agent's expected utility going forward (ignoring past effort costs $\phi_1, \ldots, \phi_t$) after observing X^t (and therefore x_t). Thus,

$$U_t(X^t) \equiv E\left[u\left(w(X^T) - \sum_{\tau=t+1}^{T} c(\phi_\tau)\right)|X^t\right].$$

Note here that U_t differs from a standard value function by the constant $u(-\sum_{\tau=1}^{t} c(\phi_\tau))$. Let $\underline{w}_t(X^t)$ be the certain equivalent of income of U_t. That is, $u(\underline{w}_t(X^t)) \equiv U_t(X^t)$. Note that $\underline{w}_t(X^{t-1}, x_{it})$ is the certain equivalent for obtaining output x_i in period t following a history of X^{t-1}.

To implement $\phi_t(X^{t-1})$, it must be the case that

$$\phi_t(X^{t-1}) \in \arg\max_{\phi \in \Phi} \sum_{i=0}^{N} \phi_i u(\underline{w}_t(X^{t-1}, x_{it}) - c(\phi)),$$

where we have dropped the irrelevant multiplicative constant. Applying our previous Lemma 2 and 3, if $\phi_t(X^{t-1})$ is implementable with certainty equivalent $\underline{w}_{t-1}(X^{t-1})$ then $\phi_t(X^{t-1})$ is implementable using wage

$$\tilde{w}(x_{it}|\phi_t(X^{t-1})) \equiv \underline{w}_t(X^{t-1}, x_{it}) - \underline{w}_{t-1}(X^{t-1})$$

with a certainty equivalent of $\underline{w} = 0$. Rearranging the above relationship,

$$\underline{w}_t(X^{t-1}, x_{it}) = \tilde{w}_t(x_{it}|\phi_t(X^{t-1})) + \underline{w}_{t-1}(X^{t-1}).$$

Integrating this difference equation from $t = 1$ to T yields

$$w(X^T) \equiv \underline{w}_T(X^T) = \sum_{t=1}^{T} \tilde{w}_t(x_{it}|\phi_t(X^{t-1})) + \underline{w}_0,$$

or in other words, the final wage is the sum of the individual single-period wage schedules for implementing $\phi_t(X^{t-1})$.

Define the row vector over outputs corresponding to $\tilde{w}$ as

$$\tilde{\mathbf{w}}_t(\phi_t(X^{t-1})) \equiv (\tilde{w}_t(x_{0t}|\phi_t(X^{t-1})), \ldots, \tilde{w}_t(x_{nt}|\phi_T(X^{T-1})).$$

Define A_i^t as an "account" that gives the number of times outcome i has occurred up to date t and $A^t = (A_0^t, \ldots, A_N^t)'$. We can thus re-write our implementing wage as

$$w(X^T) = \sum_{t=1}^{T} \tilde{\mathbf{w}}_t(\phi_t(X^{t-1})) \cdot (A^t - A^{t-1}) + \underline{w}_0. \tag{10}$$

We thus have characterized a wage schedule, $w(X^T)$, for implementing $\phi_t(X^{t-1})$. Notice at this point that the agent will only care about the history X^{t-1} when choosing ϕ_t if the principal chooses to implement a history-dependent sequence of efforts. We now show that with a risk neutral principal (or more generally in Holmström and Milgrom (1987), with a CARA-utility principal), the optimal contract will be history independent.

Theorem 1. The optimal contract is to implement $\phi_t(X^{t-1}) = \phi^* \quad \forall\ t$ and offer the wage schedule

$$w(X^T) = \sum_{t=1}^{T} w^* \cdot A^T.$$

Proof: Proof by induction. The theorem is true by definition for $T = 1$. Suppose that it holds for $T = \tau$ and consider $T = \tau + 1$. Let V_T^* be the principal's value function for the

T-period problem where the static optimum, (w^*, ϕ^*) is repeated each period. The value of the contract to the principal is

$$E\left[\sum_{t=1}^{\tau}\sum_{i=1}^{n}(x_{it} - \tilde{w}_t(x_{it}|\phi_t(X^{t-1})) + \sum_{i=0}^{n}(x_{i,\tau+1} - \tilde{w}_{\tau+1}(x_{i,\tau+1}|\phi_{\tau+1}(X^{\tau}))\right]$$
$$\leq V_\tau^* + E\left[\sum_{i=0}^{n}(x_{i,\tau+1} - \tilde{w}_{\tau+1}(x_{i,\tau+1}|\phi_{\tau+1}(X^{\tau}))\right] \leq V_\tau^* + V_1^* = V_{\tau+1}^*.$$

The upper bound is met by repeating the one-period contract T times. □

Remarks:

1. Note importantly that what we have shown is that the optimal contract is *linear in accounts*. Specifically, if $\mathbf{w}^* = (w^*(x_0), \ldots, w^*(x_n))$ in the one-period optimal contract, then
$$w(X^T) = \sum_{t=1}^{T} \mathbf{w}^* \cdot A^T,$$
or alternatively, letting $\alpha_i \equiv w^*(x_i) - w^*(x_0)$ and $\beta \equiv T \cdot w^*(x_0)$,
$$w(X^T) = \sum_{i=1}^{N} \alpha_i A_i^T + \beta.$$
This is not generally linear in profits.

2. If there are only two accounts, however, such as success ($x = x_1$) or failure ($x = x_0$), then wages are linear in "profits" (i.e., successes). From above we have
$$w(X^T) = \alpha A_1^T + \beta.$$

3. With more than two accounts, we do not get linearity. Getting an output of 50 three times is not the same as getting the output of 150 once and 0 twice.

4. Note that the history of accounts is irrelevant. Only total instances of outputs are important. This is also true in the continuous case. Thus, A^T is "sufficient" with respect to X^T. This is not inconsistent with Holmström (1979) and Shavell (1979). Sufficiency notions should be thought of as sufficient information regarding the binding constraints. Here, the binding constraint is shifting to another constant action, for which A^T is sufficient.

5. The key to our results are stationarity which in turn is due exclusively to time-separable CARA utility and an i.i.d. stochastic process.

Continuous Model:

We now consider the limit in the binary version of the model ($n = 2$) as the time periods become infinitesimal and ask what happens if the agent controls the drift of a Brownian-motion process.

Results:

1. In the limit, we obtain a linearity in accounts result, where the accounts are movements in the stochastic process. With unidimensional Brownian motion, (i.e., the agent controls the drift rate on a one-dimensional Brownian motion process), we obtain linearity in profits.
2. Additionally, in the limit, if only a subset of accounts can be contracted upon (specifically, a linear aggregate), then the optimal contract will be linear in those accounts. Thus, if only profits are contractible, we will obtain the linearity in profits result in the limit – even when the underlying process is multinomial Brownian motion.
3. If the agent must take all of his actions simultaneously at $t = 0$, then our results do not hold. Instead, we are in the world of static nonlinear contracts. In a continuum, Mirrlees's example would apply, and we could obtain an outcome arbitrarily close to the first best.

2.3 Simple economics of linear contracts

To see the usefulness of Holmström and Milgrom's (1987) setting for simple comparative statics, consider the following model. The agent has exponential utility with a CARA parameter of r; the principal is risk neutral. Profits (excluding wages) are $x = \mu + \varepsilon$, where μ is the agent's action choice (the drift rate of a unidimensional Brownian process) and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Because we made the returns to effort linear, we will embed a convexity in the agent's cost of effort function,[7] and use a simple, quadratic function: $c(\mu) = \frac{k}{2}\mu^2$.

In the full information world where e is contractible, the principal would offer the agent a fixed wage equal to $w = \underline{w} + c(\mu)$, and would maximize $E[x|\mu] - \underline{w} - c(\mu)$ which is equivalent to maximizing $\mu - c(\mu)$. Assuming $\underline{w} = 0$ for simplicity, the full information first-best contract, $\mu^{FB} = \frac{1}{k}$, the agent is paid a constant wage to cover the cost of effort, $w^{FB} = \frac{1}{2k}$, and the principal receives net profits of $\pi = \frac{1}{2k}$.

When effort is not contractible, Holmström and Milgrom's linearity result tells us that we can restrict attention to wage schedules of the form $w(x) = \alpha x + \beta$. With this contract, the

[7] This is without loss of generality. We could define $\tilde{e} = \frac{k}{2}\mu^2$ so that the agent's cost is linear in "effort", $\tilde{e}$, as before; in this case the mean of the normal distribution would be determined by the strictly concave function $\sqrt{\frac{2\tilde{e}}{k}}$.

agent's certainty equivalent[8] upon choosing an action μ is

$$\alpha\mu + \beta - \frac{k}{2}\mu^2 - \frac{r}{2}\alpha^2\sigma^2.$$

The first-order condition is $\alpha = \mu k$ which is necessary and sufficient because the agent's certainty equivalent function is globally concave in μ.

It is very important to note that the utilities-possibility frontier for the principal and agent is *linear* for a given (α, μ) and independent of β. The independence of β is an artifact of CARA utility (that's our result from Theorem 2 above), and the linearity is due to the combination of CARA utility and normally distributed errors (the latter of which is due to the central limit theorem).

As a consequence, the principal's optimal choice of (α, μ) is independent of β; β is chosen solely to satisfy the agent's IR constraint. Thus, the principal solves

$$\max_{\alpha,\mu} \ \mu - \frac{k}{2}\mu^2 - \frac{r}{2}\alpha^2\sigma^2,$$

subject to $\alpha = \mu k$. The solution gives us (α^*, μ^*, π^*):

$$\alpha^* = (1 + rk\sigma^2)^{-1},$$

$$\mu^* = (1 + rk\sigma^2)^{-1}k^{-1} = \alpha^*\mu^{FB} < \mu^{FB},$$
$$\pi^* = (1 + rk\sigma^2)^{-1}(2k)^{-1} = \alpha^*\pi^{FB} < \pi^{FB}.$$

The simple comparative statics are immediate. As either r, k, or σ^2 decrease, the power of the optimal incentive scheme increases (i.e., α^* increases). Because α^* increases, effort and profits also increase closer toward the first best. Thus when risk aversion, the uncertainty in measuring effort, or the curvature of the agent's effort function decrease, we move toward the first best. The intuition for why the curvature of the agent's cost function matters can be seen by totally differentiating the agent's first-order condition for effort. Doing so, we find that $\frac{d\mu}{d\alpha} = \frac{1}{C''(\mu)} = \frac{1}{k}$. Thus, lowering k makes the agent's effort choice more responsive to a change in α.

Remarks:

1. Consider the case of additional information. The principal observes an additional signal, y, which is correlated with ε. Specifically, $E[y] = 0$, $V[y] = \sigma_y^2$, and $Cov[\varepsilon, y] =$

[8]Note that the moment generating function for a normal distribution is $M_x(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ and the defining property of the m.g.f. is that $E_x[e^{tx}] = M_x(t)$. Thus,

$$E_\varepsilon[-e^{-r(\alpha\mu+\alpha\varepsilon+\beta-C(\mu)))}] = -e^{-r(\alpha\mu+\beta-C(\mu))+\frac{1}{2}\alpha^2 r^2\sigma^2}.$$

Thus, the agent's certainty equivalent is $\alpha\mu + \beta - C(\mu) - \frac{r}{2}\alpha^2\sigma^2$.

$\rho\sigma_y\sigma_e$. The optimal wage contract is linear in both aggregates: $w(x,y) = \alpha_1 x + \alpha_2 y + \beta$. Solving for the optimal schemes, we have

$$\alpha_1^* = (1 + rk\sigma_\varepsilon^2(1-\rho^2))^{-1},$$

$$\alpha_2^* = -\alpha_1 \frac{\sigma_\varepsilon}{\sigma_y}\rho.$$

As before, $\mu^* = \alpha_1^* \mu^{FB}$ and $\pi^* = \alpha_1^* \pi^{FB}$. It is as if the outside signal reduces the variance on ε from σ_ε^2 to $\sigma_\varepsilon^2(1-\rho^2)$. When either $\rho = 1$ or $\rho = -1$, the first-best is obtainable.

2. The allocation of effort across tasks may be greatly influenced by the nature of information. To see this, consider a symmetric formulation with two tasks: $x_1 = \mu_1 + \varepsilon_1$ and $x_2 = \mu_2 + \varepsilon_2$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ and are independently distributed across i. Suppose also that $C(\mu) = \frac{1}{2}\mu_1^2 + \frac{1}{2}\mu_2^2$ and the principal's net profits are $\pi = x_1 + x_2 - w$. If only $x = x_1 + x_2$ were observed, then the optimal contract has $w(x) = \alpha x + \beta$, and the agent would equally devote his attention across tasks. Additionally, if $\sigma_1 = \sigma_2$ and the principal can contract on both x_1 and x_2, the optimal contract has $\alpha_1 = \alpha_2$ and so again the agent equally allocates effort across tasks.

   Now suppose that $\sigma_1 < \sigma_2$ and the principal can contract on x_1 and x_2 separately. The resulting first-order conditions imply that $\alpha_1^* > \alpha_2^*$. Thus, optimal effort allocation may be entirely determined by the information structure of the contracting environment. The intuition here is that the "price" of inducing effort on task 1 is lower for the principal because information is more informative. Thus, the principal will "buy" more effort from the agent on task 1 than task 2.

2.4 Extensions: Multi-task Incentive Contracts

We now consider more explicitly the implications of multiple tasks within a firm using the linear contracting model of Holmström and Milgrom (1987). This analysis closely follows Holmström and Milgrom (1991).

The Basic Linear Model with Multiple Tasks:

The principal can contract on the following k vector of aggregates:

$$x = \mu + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \Sigma)$. The agent chooses a vector of efforts, μ, at a cost of $C(\mu)$.[9] The agent's utility is exponential with CARA parameter of r. The principal is risk neutral, offers wage schedule $w(x) = \alpha' x + \beta$, and obtains profits of $B(\mu) - w$. [Note, α and μ are vectors; $B(\mu)$, β and $w(x)$ are scalars.]

[9] Note that Holmström and Milgrom (1991) take the action vector to be t where $\mu(t)$ is determined by the action. We'll concentrate on the choice of μ as the primitive.

As before, CARA utility and normal errors implies that the optimal contract solves

$$\max_{\alpha,\mu} \quad B(\mu) - C(\mu) - \frac{r}{2}\alpha'\Sigma\alpha,$$

such that

$$\mu \in \arg\max_{\tilde{\mu}} \quad \alpha'\tilde{\mu} - C(\tilde{\mu}).$$

Given the optimal (α, μ), β is determined so as to meet the agent's IR constraint:

$$\beta = \underline{w} - \alpha'\mu + C(\mu) + \frac{r}{2}\alpha'\Sigma\alpha.$$

The agent's first-order condition (which is both necessary and sufficient) satisfies

$$\alpha_i = C_i(\mu), \quad \forall\ i,$$

where subscripts on C denote partial derivatives with respect to the indicated element of μ. Comparative statics on this equation reveal that

$$\left[\frac{\partial \mu}{\partial \alpha}\right] = \left[C_{ij}(\mu)\right]^{-1}.$$

This implies that in the simple setting where $C_{ij} = 0 \quad \forall\ i \neq j$, that $\frac{d\mu_i}{d\alpha_i} = \frac{1}{C_{ii}(\mu)}$. Thus, the marginal affect of a change in α on effort is inversely related to the curvature of the agent's cost of effort function. We have the following theorem immediately.

Theorem 2. The optimal contract satisfies

$$\alpha^* = \left(I + r[C_{ij}(\mu^*)]\Sigma\right)^{-1} B'(\mu^*).$$

Proof: Form the Lagrangian:

$$\mathcal{L} \equiv B(\mu) - C(\mu) - \frac{r}{2}\alpha'\Sigma\alpha + \lambda'(\alpha - C'(\mu)),$$

where $C'(\mu) = [C_i(\mu)]$. The $2k$ first-order conditions are

$$B'(\mu^*) - C'(\mu^*) - \lambda[C_{ij}(\mu^*)] = 0,$$

$$-r\Sigma\alpha^* + \lambda = 0.$$

Substituting out λ and solving for α^* produces the desired result. □

Remarks:

1. If ε_i are independent and $C_{ij} = 0$ for $i \neq j$, then

   $$\alpha_i^* = B_i(\mu^*)(1 + rC_{ii}(\mu^*)\sigma_i^2)^{-1}.$$

   As r, σ_i, or C_{ii} decrease, α_i^* increases. This result was found above in our simple setting of one task.

2. Given μ^*, the cross partial derivatives of B are unimportant for the determination of α^*. Only cross partials in the agent's utility function are important (i.e., C_{ij}).

Simple Interactions of Multiple Tasks:

Consider the setting where there are two tasks, but where the effort of only the first task can be measured: $\sigma_2 = \infty$ and $\sigma_{12} = 0$. A motivating example is a teacher who teaches basic skills (task 1) which is measurable via student testing and higher-order skills such as creativity, etc. (task 2) which is inherently unmeasureable. The question is how we want to reward the teacher on the basis of basic skill test scores.

Suppose that under the optimal contract $\mu^* > 0$; that is, both tasks will be provided at the optimum.[10] Then the optimal contract satisfies $\alpha_2^* = 0$ and

$$\alpha_1^* = \left(B_1(\mu^*) - B_2(\mu^*) \frac{C_{12}(\mu^*)}{C_{22}(\mu^*)} \right) \left(1 + r\sigma_1^2 \left(C_{11}(\mu^*) - \frac{C_{12}(\mu^*)^2}{C_{22}(\mu^*)} \right) \right)^{-1}.$$

Some interesting conclusions emerge.

1. If effort levels across tasks are complements (i.e., $C_{12} < 0$), the larger in magnitude the cross-effort effect (i.e., more complementary the effort levels), the higher is α_1^*. If effort levels are substitutes, the reverse is generally true and α_1^* is decreased.[11] In our example, if a teacher only has 8 hours a day to teach, the optimal scheme will put less emphasis on basic skills the more likely the teacher is to substitute away from higher-order teaching.
2. The above result has a flavor of the public finance results that when the government can only tax a subset of goods, it should tax them more or less depending upon whether the taxable goods are substitutes or complements with the un-taxable goods. See, for example, Atkinson and Stiglitz [1980 Ch. 12] for a discussion concerning taxation on consumption goods when leisure is not directly taxable.
3. There are several reasons why the optimal contract may have $\alpha_1^* = 0$.
   - Note that in our example, $\alpha_1^* < 0$ if $B_1 < B_2 C_{12}/C_{22}$. Thus, if the agent can freely dispose of x_1, the optimal constrained contract has $\alpha_1^* = 0$. No incentives are provided.
   - Suppose that technologies are otherwise symmetric: $C(\mu) = c(\mu_1 + \mu_2)$ and $B(\mu_1, \mu_2) \equiv B(\mu_2, \mu_1)$. Then $\alpha_1^* = \alpha_2^* = 0$. Again, no incentives are provided.

[10] Here we need to assume something like $C_2(\mu_1, \mu_2) < 0$ for $\mu_2 \leq 0$ so that without any incentives on task 2, the agent still allocates some effort on task 2. In the teaching example, absent incentives, a teacher will still teach some higher-order skills.

[11] Note that there are two effects with substitutes operating in opposite directions. For efforts that are not too substitutable, the effect will hold. For example, suppose $C(\mu_1, \mu_2) = \frac{1}{2}\mu_1^2 + \frac{1}{2}\mu_2^2 + \lambda\mu_1\mu_2$; then for $\lambda > 0$ but not too large, an increase in λ leads to a decrease in α_i^*.

(Here, we need to be careful because C_{ij} is not invertible with perfect substitutes. If one parameterizes the degree of substitution, e.g., $C(\mu_1, \mu_2) = \frac{1}{2}\mu_1^2 + \frac{1}{2}\mu_2^2 + \lambda\mu_1\mu_2$, then the limiting case where $\lambda = 0$ corresponds to $\alpha_1^* = \alpha_2^* = 0$.

- Note that if $C_i(0) > 0$, there is a fixed cost to effort. This implies that a corner solution may emerge where $\alpha_i^* = 0$. A final reason for no incentives.

2.5 Application: Limits on Outside Activities.

Consider the principal's problem when an additional control variable is added: the set of allowable activities. Suppose that the principal cares only about effort devoted to task 0: $\pi = \mu_0 - w$. In addition, there are N potential tasks which the agent could spend effort on and which increase the agent's personal utility. We will denote the set of these tasks by $K = \{1, \ldots, N\}$. The principal has the ability to exclude the agent from any subset of these activities, allowing only tasks or activities in the subset set $A \subset K$. Unfortunately, the principal can only contract over x_0, and so $w(x) = \alpha x_0 + \beta$.

It is not always profitable, even in the full-information setting, to exclude these tasks from the agent, because they may be efficient and therefore reduce the principal's wage bill. As a motivating example, allowing an employee to use the company's internet service for personal use may be a cheap perk for the firm to provide and additionally lowers the necessary wage which the firm must pay. Unfortunately, the agent may then spend all of the day surfing the internet rather than working.

Suppose that the agent's cost of effort is

$$C(\mu) = c\left(\mu_0 + \sum_{i=1}^{N} \mu_i\right) - \sum_{i=1}^{N} v_i(\mu_i).$$

The v_i functions represent the agent's personal utility from allocating effort to task i; v_i is assumed to be strictly concave and $v_i(0) = 0$. The principal's expected returns are simply $B(\mu) = p\mu_0$, where p is the price or shadow value of output.

We first determine the principal's optimal choice of $A^*(\alpha)$ for a given α, and then we solve for the optimal α^*. The first-order condition which characterizes the agent's optimal μ_0 is

$$\alpha = c'\left(\sum_{i=0}^{N} \mu_i\right),$$

and (substituting)

$$\alpha = v_i'(\mu_i), \quad \forall\, i.$$

Note that the choice of μ_i depends only upon α. Thus, if the agent is allowed an additional personal task, k, the agent will allocate time away from task 0 by an amount equal to $v_k^{-1\prime}(\alpha)$. The benefit of allowing the agent to spend time on task k is $v_k(\mu_k(\alpha))$ (via a

reduced wage) and the (opportunity) cost is $p\mu_k(\alpha)$. Therefore, the optimal set of tasks for a given α is

$$A^*(\alpha) = \{k \in K | v_k(\mu_k(\alpha)) > p\mu_k(\alpha)\}.$$

We have the following results for a given α.

Theorem 3. Assume that α is such that $\mu(\alpha) > 0$ and $\alpha < p$. Then the optimal set of allowed tasks is given by $A^*(\alpha)$ which is monotonically expanding in α (i.e., $\alpha \leq \alpha'$, then $A^*(\alpha) \subset A^*(\alpha')$).

Proof: That the optimal set of allowed tasks is given by $A^*(\alpha)$ is true by construction. The set $A^*(\alpha)$ is monotonically expanding in α iff $v_k(\mu_k(\alpha)) - p\mu_k(\alpha)$ is increasing in α. I.e.,

$$[v_k'(\mu_k(\alpha)) - p]\frac{d\mu_k(\alpha)}{d\alpha} = [\alpha - p]\frac{1}{v_k''(\mu_k(\alpha))} > 0.$$

□

Remarks:

1. The fundamental premise of exclusion is that incentives can be given by either increasing α on the relevant activity or decreasing the opportunity cost of effort (i.e, by reducing the benefits of substitutable activities).
2. The theorem indicates a basic proposition with direct empirical content: *responsibility (large α) and authority (large $A^*(\alpha)$) should go hand in hand*. An agent with high-powered incentives should be allowed the opportunity to expend effort on more personal activities than someone with low-powered incentives. In the limit when $\sigma \to 0$ or $r \to 0$, the agent is residual claimant $\alpha^* = 1$, and so $A^*(1) = K$. Exclusion will be more frequently used the more costly it is to supply incentives.
3. Note that for α small enough, $\mu(\alpha) = 0$, and the agent is not hired.
4. The set $A^*(\alpha)$ is independent of r, σ, C, etc. These variables only influence $A^*(\alpha)$ via α. Therefore, an econometrician can regress $||A^*(\alpha)||$ on α, and α on $(r, \sigma, \ldots)$ to test the multi-task theory. See Holmström and Milgrom (1994).

Now, consider the choice of α^* given the function $A^*(\alpha)$.

Theorem 4. Providing that $\mu(\alpha^*) > 0$ at the optimum,

$$\alpha^* = p\left(1 + r\sigma^2\left(\frac{1}{c''\left(\sum_i \mu_i(\alpha^*)\right)} + \sum_{k \in A^*(\alpha^*)} \frac{1}{v_k''(\mu_k(\alpha^*))}\right)\right)^{-1}.$$

The proof of this theorem is an immediate application of our first multi-task characterization theorem. Additionally, we have the following implications.

Remarks:

1. The theorem indicates that when either r or σ decreases, α^* increases. (Note that this implication is not immediate because α^* appears on both sides of the equation; some manipulation is required. With quadratic cost and benefit functions, this is trivial.) By our previous result on $A^*(\alpha)$, the set of allowable activities also increases as α^* increases.
2. Any personal task excluded in the first-best arrangement (i.e., $v_k'(0) < p$) will be excluded in the second-best optimal contract given our construction of $A^*(\alpha)$ and the fact that v_k is concave. This implies that there will be more constraints on agent's activities when performance rewards are weak due to a noisy environment.
3. Following the previous remark, one can motivate rigid rules which limit an agent's activities (seemingly inefficiently) as a way of dealing with substitution possibilities. Additionally, when the "personal" activity is something such as rent-seeking (e.g., inefficiently spending resources on your boss to increase your chance of promotion), a firm may wish to restrict an agent's access to such an activity or withdraw the bosses discretion to promote employees so as to reduce this inefficient activity. This idea was formalized by Milgrom (1988) and Milgrom and Roberts (1988).
4. This activity exclusion idea can also explain why firms may not want to allow their employees to "moonlight". Or more importantly, why a firm may wish to use an internal sales force which is not allowed to sell other firms' products rather than an external sales force whose activities vis-a-vis other firms cannot be controlled.

2.6 Application: Task Allocation Between Two Agents.

Now consider two agents, $i = 1, 2$, who are needed to perform a continuum of tasks indexed by $t \in [0, 1]$. Each agent i expends effort $\mu_i(t)$ on task t; total cost of effort is $C\left(\int \mu_i(t)dt\right)$. The principal observes $x(t) = \mu(t) + \varepsilon(t)$ for each task, where $\sigma^2(t) > 0$ and $\mu(t) \equiv \mu_1(t) + \mu_2(t)$. The wages paid to the agents are given by:

$$w_i(x) = \int_0^1 \alpha_i(t)x(t)dt + \beta_i.$$

By choosing $\alpha_i(t)$, the principal allocates agents to the various tasks. For example, when $\alpha_1(.4) > 0$ but $\alpha_2(.4) = 0$, only agent 1 will work on task .4.

Two results emerge.

1. For any required effort function $\mu(t)$ defined on $[0, 1]$, it is never optimal to assign

two agents to the same task: $\alpha_1^*(t)\alpha_2^*(t) \equiv 0$. This is quite natural given the free-riding problem which would otherwise emerge.

2. More surprisingly, suppose that the principal must obtain a uniform level of effort $\mu(t) = 1$ across all tasks. At the optimum, if $\int \mu_i(t)dt < \int \mu_j(t)dt$, then the hardest to measure tasks go to agent i (i.e., all tasks t such that $\sigma(t) \geq \overline{\sigma}$.) This results because you want to avoid the multi-task problems which occur when the various tasks have vastly different measurement errors. Thus, the principal wants information homogeneity. Additionally, the agent with the hard to measure tasks exerts lower effort and receives a lower "normalized commission" because the information structure is so noisy.

2.7 Application: Common Agency.

Bernheim and Whinston (1986) were to first to undertake a detailed study of the phenomena of common agency with moral hazard. "Common agency" refers to the situation in which several principals contract with the same agent in common. The interesting economics of this setting arise when one principal's contract imposes an externality on the contracts of the others.

Here, we follow Dixit (1996) restricting attention to linear contracts in the simplest setting of n independent principals who simultaneously offer incentive contracts to a single agent who controls the m-dimensional vector t which in turn effects the output vector $x \in \mathcal{R}^{\Updownarrow}$. Let $x = t + \varepsilon$, where $x, t \in \mathcal{R}^{\Updownarrow}$ and $\varepsilon \in \mathcal{R}^{\Updownarrow}$ is distributed normally with mean vector 0 and covariance matrix Σ. Cost of effort is a quadratic form with a positive definite matrix, C.

1. The first-best contract (assuming t can be contracted upon) is simply $t = C^{-1}b$.

2. The second-best *cooperative* contract. The combined return to the principals from effort vector t is $b't$, and so the total expected surplus is

$$b't - \frac{1}{2}t'Ct - \frac{r}{2}\alpha'\Sigma\alpha.$$

   This is maximized subject to $t = C^{-1}\alpha$. The first-order condition for the slope of the incentive contract, α, is

$$C^{-1}b - [C^{-1} + r\Sigma]\alpha = 0,$$

   or $b = [I + rC\Sigma]\alpha$ or $b - \alpha = rC\Sigma\alpha > 0$.

3. The second-best *non-cooperative* un-restricted contract. Each principal's return is given by the vector b^j, where $b = \sum b^j$. Suppose each principal j is unrestricted in choosing its wage contract; i.e., $w^j = \alpha^{j'} + \beta^j$, where α^j is a full m-dimensional vector. Define $A^{-j} \equiv \sum_{i \neq j} \alpha^i$ and $B^{-j} \equiv \sum_{i \neq j} \beta^i$. From principal j's point of view, absent any contract from himself $t = C^{-1}A^{-j}$ and the certainty equivalent is $\frac{1}{2}A^{-j'}[C^{-1} - r\Sigma]A^{-j} + B^{-j}$. The aggregate incentive scheme facing the agent is

$\alpha = A^{-j} + \alpha^j$ and $\beta = B^{-j} + \beta^j$. Thus the agent's certainty equivalent *with* principal j's contract is

$$\frac{1}{2}(A^{-j} + \alpha^j)'[C^{-1} - r\Sigma](A^{-j} + \alpha^j) + B^{-j} + \beta^j.$$

The incremental surplus to the agent from the contract is therefore

$$A^{-j'}(C^{-1} - r\Sigma)\alpha^j + \frac{1}{2}\alpha^{j'}[C^{-1} - r\Sigma]\alpha^j + \beta^j.$$

As such, principal j maximizes

$$b^{j'}C^{-1}A^{-j} - rA^{-j'}\Sigma\alpha^j + b^{j'}C^{-1}\alpha^j - \frac{1}{2}\alpha^{j'}[C^{-1} + r\Sigma]\alpha^j.$$

The first-order condition is

$$C^{-1}b^j - [C^{-1} + r\Sigma]\alpha^j - r\Sigma A^{-j} = 0.$$

Simplifying, $b^j = [I + rC\Sigma]\alpha^j + rC\Sigma A^{-j}$. Summing across all principals,

$$b = [I + rC\Sigma]\alpha + rC\Sigma(n-1)\alpha = [I + nrC\Sigma]\alpha,$$

or $b - \alpha = nrC\Sigma\alpha > 0$. Thus, the distortion has increased by a factor of n. Intuitively, it is as if the agent's risk has increased by a factor of n, and so therefore incentives will be reduced on every margin. *Hence, unrestricted common agency leads to more effort distortions.*

Note that $b^j = \alpha^j - rC\Sigma\alpha$, so substitution provides

$$\alpha^j = b^j - rC\Sigma[I + nrC\Sigma]^{-1}b.$$

To get some intuition for the increased-distortion result, suppose that $n = m$ and that each principal cares only about output j; i.e., $b_i^j = 0$ for $i \neq j$, and $b_j^j > 0$. In such a case,

$$\alpha_i^j = -rC\Sigma[I + nrC\Sigma]^{-1}b < 0,$$

so each principal finds it optimal to pay the agent not to produce on the other dimensions!

4. The second-best *non-cooperative restricted* contract. We now consider the case in which each principal is restricted in its contract offerings so as to not pay the agent for for output on the other principals' dimensions. Specifically, let's again assume that $n = m$ and that each principal only cares about x_j: $b_i^j = 0$ for $i \neq j$, and $b_j^j > 0$. The restriction is that $\alpha_i^j = 0$ for $j \neq i$. In such a setting, Dixit demonstrates that the equilibrium incentives are higher than in the un-restricted case. Moreover, if efforts are perfect substitutes across agents, $\alpha^j = b^j$ and first-best efforts are implemented.