Suggested Solutions to:
Regular Exam, Fall 2019
Contract Theory
January 17, 2020

This version: January 27, 2020

Question 1: Insurance and adverse selection

Part (a)

At the first-best optimum (i.e., the optimum when A's type is observable), both types are offered a contract with full insurance (so that $\overline{u}_N = \overline{u}_A$ and $\underline{u}_N = \underline{u}_A$). Explain, in words, the economic logic behind this result.

- Two crucial assumptions that lead to this result are that (i) A is risk averse and (ii) P is risk neutral. The objective of P is to maximize its (expected) payoff. Under first best, the only constraints are the individual rationality constraints. Therefore, it is in the interest of P to choose A's level of insurance (for any given price A must pay for this insurance) in a way that makes A's payoff as large as possible, at least as long as this can be done at no cost for P. For if A's payoff from the insurance is higher, then P can charge more for the insurance without making A prefer his outside option. Given that A is risk averse and P is risk neutral, providing A with more insurance leads to a higher payoff for A at no cost for P. Hence the first-best optimum involves P providing full insurance to A and then choosing the effective price for this insurance so high that each type of A is indifferent between the outside option and the insurance contract.
  - The reason why the logic above does not apply under second best is that then P has a smaller number of instruments available: P cannot observe A's type, which means that the level of A's insurance must also be such that A voluntarily chooses the right contract.

Part (b)

Show that the constraints (IC-high) and (IC-low) jointly imply that $\underline{u}_N - \underline{u}_A \geq \overline{u}_N - \overline{u}_A$.

- Add up the ICs:

$$\left(1-\overline{\theta}\right)\overline{u}_N + \overline{\theta}\overline{u}_A + \left(1-\underline{\theta}\right)\underline{u}_N + \underline{\theta}\underline{u}_A \geq \left(1-\overline{\theta}\right)\underline{u}_N + \overline{\theta}\underline{u}_A + \left(1-\underline{\theta}\right)\overline{u}_N + \underline{\theta}\overline{u}_A.$$

Re-arranging and noticing that some terms cancel out, we obtain

$$-\left(\overline{\theta}-\underline{\theta}\right)\overline{u}_N + \left(\overline{\theta}-\underline{\theta}\right)\overline{u}_A + \left(\overline{\theta}-\underline{\theta}\right)\underline{u}_N - \left(\overline{\theta}-\underline{\theta}\right)\underline{u}_A \geq 0.$$

Since $\overline{\theta} > \underline{\theta}$, the inequality simplifies to

$$-\overline{u}_N + \overline{u}_A + \underline{u}_N - \underline{u}_A \geq 0$$

or

$$\underline{u}_N - \underline{u}_A \geq \overline{u}_N - \overline{u}_A,$$

which we were asked to show.

Part (c)

Assume that the constraints (IR-high) and (IC-low) are lax at the second-best optimum (so that they can be disregarded). Show that, at the second-best optimum, the high type is fully insured ($\overline{u}_N = \overline{u}_A$) whereas the low-type is underinsured ($\underline{u}_N > \underline{u}_A$).

- The Lagrangian:

$$\begin{aligned}\mathcal{L} &= v\left[w - \underline{\theta}d - (1-\underline{\theta})\,h(\underline{u}_N) - \underline{\theta}h(\underline{u}_A)\right] + (1-v)\left[w - \overline{\theta}d - \left(1-\overline{\theta}\right)h(\overline{u}_N) - \overline{\theta}h(\overline{u}_A)\right] \\ &\quad + \lambda\left[(1-\underline{\theta})\,\underline{u}_N + \underline{\theta}\underline{u}_A - \underline{U}^*\right] + \mu\left[\left(1-\overline{\theta}\right)\overline{u}_N + \overline{\theta}\overline{u}_A - \left(1-\overline{\theta}\right)\underline{u}_N - \overline{\theta}\underline{u}_A\right],\end{aligned}$$

  where λ is the shadow price associated with IR-low and μ is the shadow price associated with IC-high.

- FOC w.r.t. $\overline{u}_N$:

$$\frac{\partial \mathcal{L}}{\partial \overline{u}_N} = -(1-v)\left(1-\overline{\theta}\right)h'(\overline{u}_N) + \mu\left(1-\overline{\theta}\right) = 0$$

  or

$$\boxed{(1-v)\,h'(\overline{u}_N) = \mu.} \tag{1}$$

  - This implies that $\boxed{\mu > 0}$; i.e., $\boxed{\text{IC-high binds at the optimum}}$.

- FOC w.r.t. $\underline{u}_N$:

$$\frac{\partial \mathcal{L}}{\partial \underline{u}_N} = -v(1-\underline{\theta})\,h'(\underline{u}_N) + \lambda(1-\underline{\theta}) - \mu\left(1-\overline{\theta}\right) = 0$$

  or

$$\boxed{v(1-\underline{\theta})\,h'(\underline{u}_N) = \lambda(1-\underline{\theta}) - \mu\left(1-\overline{\theta}\right).} \tag{2}$$

  - This implies that $\boxed{\lambda > 0}$ (spell out the arguments!); i.e., $\boxed{\text{IR-low binds at the optimum}}$.

- FOC w.r.t. $\overline{u}_A$:

$$\frac{\partial \mathcal{L}}{\partial \overline{u}_A} = -(1-v)\,\overline{\theta}h'(\overline{u}_A) + \mu\overline{\theta} = 0$$

  or

$$\boxed{(1-v)\,h'(\overline{u}_A) = \mu.} \tag{3}$$

- FOC w.r.t. $\underline{u}_A$:

$$\frac{\partial \mathcal{L}}{\partial \underline{u}_A} = -v\underline{\theta}h'(\underline{u}_A) + \lambda\underline{\theta} - \mu\overline{\theta} = 0$$

  or

$$\boxed{v\underline{\theta}h'(\underline{u}_A) = \lambda\underline{\theta} - \mu\overline{\theta}.} \tag{4}$$

- Combining (1) and (3) immediately yields (here we use $h'' > 0$)

$$\overline{u}_N = \overline{u}_A \equiv \overline{u}.$$

  - That is, full insurance for the high type, which was one of the results we were asked to show.

- Multiply (2) by $\underline{\theta}$:

$$v\underline{\theta}\left(1-\underline{\theta}\right)h'\left(\underline{u}_N\right)=\lambda\underline{\theta}\left(1-\underline{\theta}\right)-\mu\underline{\theta}\left(1-\overline{\theta}\right).$$

- Multiply (4) by $\left(1-\underline{\theta}\right)$:

$$v\underline{\theta}\left(1-\underline{\theta}\right)h'\left(\underline{u}_A\right)=\lambda\underline{\theta}\left(1-\underline{\theta}\right)-\mu\overline{\theta}\left(1-\underline{\theta}\right).$$

- Subtract the latter from the former:

$$\begin{aligned}&v\underline{\theta}\left(1-\underline{\theta}\right)h'\left(\underline{u}_N\right)-v\underline{\theta}\left(1-\underline{\theta}\right)h'\left(\underline{u}_A\right)\\=\;&\left[\lambda\underline{\theta}\left(1-\underline{\theta}\right)-\mu\underline{\theta}\left(1-\overline{\theta}\right)\right]-\left[\lambda\underline{\theta}\left(1-\underline{\theta}\right)-\mu\overline{\theta}\left(1-\underline{\theta}\right)\right]\end{aligned}$$

  or

$$\begin{aligned}&v\underline{\theta}\left(1-\underline{\theta}\right)\left[h'\left(\underline{u}_N\right)-h'\left(\underline{u}_A\right)\right]\\=\;&\mu\left[\overline{\theta}\left(1-\underline{\theta}\right)-\underline{\theta}\left(1-\overline{\theta}\right)\right]=\mu\left(\overline{\theta}-\underline{\theta}\right).\end{aligned}$$

  Since $v\underline{\theta}\left(1-\underline{\theta}\right)>0$, $\overline{\theta}>\underline{\theta}$, $\mu>0$, and $h''>0$, the above inequality implies that

$$\underline{u}_N>\underline{u}_A.$$

  - That is, the low type is underinsured, which is the second one of the results we were asked to show.

Part (d)

In some other adverse selection models that we studied, the outside option for the "good" type was (sufficiently much) more attractive than the "bad" type's outside option. This gave rise to a phenomenon called "countervailing incentives." Answer, in words, the following questions: (i) What is by meant by "countervailing incentives"? (ii) What are the possible consequences of this phenomenon in terms of efficiency and rent extraction at the second-best optimum? (iii) What is the intuition for the results under (ii)?

- **(i)** In the standard adverse selection model, with two types who have equally attractive outside options, it is the good type that has an incentive to pass himself off as the bad type. If we instead assume that the good type has a more attractive outside option than the bad type has, then this will create an incentive in the opposite direction (a "countervailing incentive")—that model feature will add a positive value of being perceived as a good type as opposed to a bad type. If the difference in outside options is only moderate, the incentive to be perceived as a bad type is still the strongest; but for a sufficiently large difference in outside options (with the one of the good type being the more attractive), the net effect is that the agent has an incentive to pass himself off as the good type. We refer to the incentives to be perceived as the good type (created by the difference in outside options) as "countervailing incentives", regardless of whether the net effect is such that the agent wants to be perceived as a good or a bad type.

- **(ii)** As we gradually increase the extent to which the good type's outside option is more attractive than the bad type's, we obtain, in turn, the following outcomes:

- The same outcome as in the standard model with equally attractive outside options: Efficiency for the good type but distortion downwards for the bad type; rents to the good type but no rents for the bad type.
- Efficiency for the good type but distortion downwards for the bad type; rents to neither type.
- Efficiency for both types; rents to neither type.
- Efficiency for the bad type but distortion upwards for the bad type; rents to neither type.
- The following case can arise at least if we ignore the possibility that P shuts down one of the types: Efficiency for the bad type but distortion upwards for the bad type; rents to the bad type but no rents for the bad type.

- **(iii)** One of the above results is that, for intermediate levels of difference in outside options, there is no inefficiency. The intuition for this is that the incentives to be perceived as a bad type and the countervailing incentives to be perceived as a good type are roughly equally strong and hence they cancel each other out: no type has an incentive to be perceived as another type. Therefore there are no (binding) incentive compatibility constraints, so we are effectively back in the first-best situation.

- Another one of the above results is that, for a large difference in outside options, it is the good type's quantity that is distorted (and it is distorted upwards). The intuition for this is that now the countervailing incentives are so strong that (a) it is the IC-bad constraint that binds and (b) the good type is not anymore the "money machine"–his outside option is so high that it is easier for P to earn money on the bad (and less able) type. Because of (a) P must distort at least one type's quantity and because of (b) the most profitable option for P is to distort the good type's quantity.

Question 2: A patent race among n firms, with moral hazard

As the first hint in the question suggests, we can make use of the first-order approach. This amounts to replacing the (infinitely many) incentive compatibility constraints in (IC) with the single requirement that the first-order condition associated with agent's optimization problem is satisfied. The agent's pay-off as a function of x_i, for given values of $\mathbf{x_{-i}}$, $\underline{t}$, and $\overline{t}$, equals

$$U_i = p_i(\mathbf{x})\,\overline{t}_i + [1 - p_i(\mathbf{x})]\,\underline{t}_i - x_i. \tag{5}$$

The first-order condition therefore becomes

$$\frac{\partial U_i}{\partial x_i} = \frac{\partial p_i(\mathbf{x})}{\partial x_i}\left(\overline{t}_i - \underline{t}_i\right) - 1 = 0 \Leftrightarrow \overline{t}_i - \underline{t}_i = \left[\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right]^{-1}. \tag{6}$$

Plugging this into the principal's objective and into the remaining constraints, the principal's new problem can be written as

$$\max_{x_i} p_i(\mathbf{x})\left[v - \left(\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right)^{-1}\right] - \underline{t}_i,$$

subject to

$$p_i(\mathbf{x})\left(\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right)^{-1} + \underline{t}_i - x_i \geq 0 \tag{IR}$$

and

$$\underline{t}_i \geq 0, \qquad \overline{t}_i \geq 0. \tag{LL}$$

The second hint in the question suggests that we should look for an equilibrium in which the IR constraint does not bind (and verify that this guess is correct at the end of the exercise). If the IR constraint does not bind, then we must have $\underline{t}_i = 0$ and $\bar{t}_i > 0$ at the optimum (by inspection of the above optimization problem). This means that the optimization problem simplifies further, as we can ignore the constraints and plug $\underline{t}_i = 0$ into the objective. Thus, the principal maximizes

$$p_i(\mathbf{x})\left[v - \left(\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right)^{-1}\right]$$

with respect to x_i, without having to take any constraints into account (except $x_i \geq 0$, which will not be binding). The first-order condition of this problem can be written as

$$\frac{\partial p_i(\mathbf{x})}{\partial x_i}\left[v - \left(\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right)^{-1}\right] + p_i(\mathbf{x})\left(\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right)^{-2}\frac{\partial^2 p_i(\mathbf{x})}{\partial x_i^2} = 0. \tag{7}$$

In the question it is stated that the probability-of-winning function $p_i(\mathbf{x})$ has the following functional form:

$$p_i(\mathbf{x}) = \begin{cases} \frac{x_i}{\sum_{j=1}^n x_j} & \text{if } \sum_{j=1}^n x_j > 0 \\ \frac{1}{n} & \text{if } \sum_{j=1}^n x_j = 0. \end{cases}$$

For $\sum_{j=1}^n x_j > 0$, we therefore have

$$\frac{\partial p_i(\mathbf{x})}{\partial x_i} = \frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^2}, \qquad \frac{\partial^2 p_i(\mathbf{x})}{\partial x_i^2} = -\frac{2\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^3}, \qquad \text{and} \qquad \frac{\partial^3 p_i(\mathbf{x})}{\partial x_i^3} = \frac{6\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^4}. \tag{8}$$

Plugging (8) into (7) , we have

$$\frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^2}\left[v - \left(\frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^2}\right)^{-1}\right] - 2p_i(\mathbf{x})\left(\frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^2}\right)^{-2}\frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^3} = 0. \tag{9}$$

Imposing symmetry, this simplifies to

$$\frac{(n-1)x}{(nx)^2}\left[v - \left(\frac{(n-1)x}{(nx)^2}\right)^{-1}\right] - \frac{2}{n}\left(\frac{(n-1)x}{(nx)^2}\right)^{-2}\frac{(n-1)x}{(nx)^3} = 0$$

or

$$\frac{n-1}{n^2x}\left(v - \frac{n^2x}{n-1}\right) = \frac{2}{n-1} \Leftrightarrow x^* = \frac{(n-1)^2 v}{(n+1)n^2}. \tag{10}$$

Moreover, the previous analysis tells us that the symmetric equilibrium values of $\underline{t}_i$ and $\bar{t}_i$ are given by $\underline{t}^* = 0$ and (using (10))

$$\bar{t}^* = \left[\frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^n x_j\right]^2}\right]^{-1} = \frac{n^2 x^*}{n-1} = \frac{(n-1)v}{n+1}. \tag{11}$$

Next, let us verify that our guess that the IR constraint is satisfied at the optimum indeed is correct.

The IR constraint at the candidate optimum can be written as:

$$p_i(\mathbf{x})\left(\frac{\partial p_i(\mathbf{x})}{\partial x_i}\right)^{-1} - x_i \geq 0$$

or

$$\frac{1}{n}\left(\frac{\sum_{j\neq i} x_j}{\left[\sum_{j=1}^{n} x_j\right]^2}\right)^{-1} - x^* \geq 0$$

or

$$\frac{1}{n}\left(\frac{(n-1)x^*}{(nx^*)^2}\right)^{-1} \geq x^* \Leftrightarrow \frac{n}{n-1} \geq 1,$$

which always holds.

Part (b)

Compare the equilibrium outcomes of the model without moral hazard, described immediately above, and the moral hazard model that you solved in part (a). In particular, at the symmetric equilibrium,

(i) which model yields the highest payoff for the owners, and

(ii) which model yields the highest total surplus (defined as the sum of the owner's and the manager's payoffs)?

In order to answer these questions, you do not need to do any math (and if you nevertheless do that, you will not get any credit for this). Instead, you should explain verbally how we should expect the logic of the two models to work and why we should expect your particular answers to the two questions to be correct. If you think there are different effects that work in opposite directions and that the answer to one or both questions therefore is ambiguous (without looking at the specific math results), then you should answer that—but you should also explain your reasoning and what these different effects are.

- Hint: Note that, in the model without moral hazard, if the owners could enter a binding agreement with each other that required them all to choose a zero effort, then each owner's payoff would equal $\pi^{coop} = \frac{v}{n}$, which is strictly larger than the non-cooperative payoff $\pi^* = \frac{v}{n^2}$. In this sense, the fact that there is competition between the owners hurts them.

First note that we should expect the second model to yield higher effort levels at the equilibrium. The reason is that there is no moral hazard problem in the second model—and we know that moral hazard typically leads to underprovision of effort.

In this context, however, "underprovision" is good for total surplus, because the competition between the firms leads to an "arms-race effect": The firms (and society overall) would save on effort costs if all firms chose a lower effort than in equilibrium, while the number of produced patents would not be affected. In other words, in this situation, underprovision of effort is beneficial for total surplus, as it saves on resources. The owners of the firms would also, for the same reason, benefit from the lower effort levels in the first model.

(Whether the agents would prefer the first or the second model would depend on in which model their rents are largest, and this appears to be less straightforward to see.)