

The O-ring theory of the firm

Michael T. Rauh 

Indiana University, Bloomington, IN, USA
(Email: mtrauh@indiana.edu)

Abstract

We develop an O-ring production function characterized by specialization and division of labor and where shirking or negative shocks can have major adverse consequences. We show that when the principal can monitor individual output, the firm tends to be large (potentially larger than first best), with a high degree of specialization and division of labor, weak incentives, and low pay as in traditional nonunion manufacturing. Moral hazard can only limit the size of the firm relative to the first best when the principal can only monitor team output, in which case the firm has the opposite characteristics.

One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on, is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations.

Smith (1904, Book I, Chapter I)

1 | INTRODUCTION

The above production process described by Smith has several features in common with many modern production technologies, particularly in manufacturing. The first is that production can be divided into a number of distinct tasks: drawing the wire, straightening it, etc. It is this aspect of the production process that allows for a *division of labor* (an allocation of tasks across agents) and *specialization* (investments in task-specific human capital). Becker and Murphy (1992) develop a model along these lines where an increase in employment leads to a more extensive division of labor (fewer tasks per agent), greater specialization, and therefore higher productivity.

A second feature is that a breakdown at one stage of production due to shirking, poor decision-making, or a negative shock can have serious adverse consequences. A batch of bent pins or an automobile with defective brakes is at best useless. Automobile recalls can be extremely costly in terms of damage to the firm's reputation even when the corrective fix is relatively inexpensive. This feature is the central element in Kremer's (1993) *O-ring* theory of the firm coined after the source of the space shuttle *Challenger* disaster.

Finally, the O-ring nature of the production technology implies something about the nature of the moral hazard problem. If the principal can directly monitor individual effort (the *first best* case) or individual output (the *second best*), then shirkers

I thank Editor Casadesus-Masanell and an anonymous Co-Editor and referee for comments that substantially improved the paper. I also thank conference participants at the International Industrial Organization Conference (IIOC) in 2015 and 2016, especially Dongsoo Shin. This paper was supported by a Kelley School of Business summer grant. It is a revised and retitled version of the previous working paper "Moral hazard, firm size, and the size-wage differential."

will be discovered and punished with probability 1. In the second best case, agents who experience negative shocks will also be punished. For example, if pins lack sharp points, then under a clear division of labor, the principal will presumably know which agent is responsible. In other settings, such as restaurants, things may not be so clear. Was it the food? The service? The principal may not know unless the customer can communicate the source of the dissatisfaction. If the principal can only monitor team output (the *third best*), the entire team will be punished for sure if any worker shirks. In each of these cases, there is no free-rider problem because shirkers cannot hide behind the efforts of other workers.

In this paper, we develop an O-ring theory of the firm that combines the main elements in Becker and Murphy (1992) and Kremer (1993) and extends those two papers to the case of moral hazard. In the contract theory literature, the production function is usually nondescript and generic. In contrast, we show that the O-ring production technology developed here has powerful implications for the nature of the moral hazard problem, incentive contracting, the size of the firm in terms of employment, and the extent of specialization and division of labor.

We consider a production process where the set of tasks is the unit interval. The number of agents is endogenous and chosen by the principal, who divides the set of tasks equally across all employed agents. Each agent chooses his production effort and investment in task-specific human capital for each of his assigned tasks. A unit of output requires one unit of output in each task (e.g., one automobile requires one headlight assembly, one steering column, etc.), so output is zero if any agent shirks or experiences a negative shock in any of his assigned tasks. As in Becker and Murphy (1992), an increase in employment implies fewer tasks per agent, which allows each agent to make greater investments in human capital for each of his smaller set of assigned tasks. The result is an increase in productivity, which leads to increasing returns to employment.

We motivate the stochastic component of the production technology as follows. In addition to production effort and investments in human capital, each agent monitors his assigned tasks and makes decisions about whether or not a problem has arisen, whether or not to halt production to fix it, whether he can fix it himself, and which potential solution is appropriate. When there is only one agent, there is a high probability that at least some of these decisions will be faulty because he has limited cognitive resources and performs all the tasks himself. When there are two agents, the probability that either one will make a mistake should be lower because each performs only half the set of tasks and can therefore devote more care and attention to each of them. On the other hand, we now have two probabilities instead of one, so the effect of an increase in employment is ambiguous.¹

Formally, we assume that the probability that the agent experiences a negative shock in at least one of his assigned tasks is an increasing function of the proportion of tasks he performs. Assuming independence, the probability of a product defect is the product of the individual probabilities. An increase in the number of agents therefore has two effects: it reduces the probability that each individual will experience a negative shock but it also increases the number of stages of production where a negative shock can occur. We say that the production process exhibits the *O-ring property* if eventually the probability of a product defect is increasing in the number of agents and converges to one as the number of agents increases without bound.²

A central question of the paper is: what limits the size of the firm? The seminal answer, due to Coase (1937), is that there exist certain *transaction costs* associated with conducting economic activity within firms. In this paper, we focus on the transaction costs associated with moral hazard; that is, agency costs. In our model, there is a one-to-one correspondence between employment and the extent of the division of labor, so the same question can also be posed as: what limits the extent of the division of labor? Becker and Murphy (1992, p. 1138) take issue with Smith's contention that the division of labor is limited by the extent of the market and argue that "a variable of great importance is the cost of combining specialized workers." But in their model, the relevant cost function is *exogenous* and only loosely justified in terms of "principal-agent conflicts, free-riding, and the difficulties of communication." In contrast, in this paper, the cost function is constructed from the optimization problem of the principal whose objective is to implement effort and investments in human capital at minimum cost.

To determine the effect of moral hazard on the size of the firm and the extent of the division of labor, we start with the first best benchmark where the principal can directly monitor effort so there is no moral hazard. Regardless of the monitoring technology, the optimal employment level balances the following considerations: (i) the increasing returns to employment due to specialization and division of labor, (ii) the O-ring property of the production technology, where the probability of team failure increases with the size of the team, and (iii) the marginal cost of employment (the cost of hiring another agent). We obtain the standard result that the first best contract provides zero incentives and full insurance. Since each agent receives a constant transfer, the first best cost of employment (the number of agents times the expected transfer to each agent) is *linear* in employment and the marginal cost of hiring another worker is constant. But a linear cost function cannot contain the increasing returns to employment due to specialization and division of labor so that the size of the first best firm can only be limited either by the extent of the labor market (the total number of available workers) or the O-ring property.

We then consider the second best monitoring technology, where the principal cannot monitor effort but she can monitor individual output. We show that the optimal second best contract is to award a bonus when individual output is high (when the agent's actions are first best and the shock is positive) and to fire the agent otherwise. We then show that the bonus (what we call the second best *incentive*) is decreasing in employment. This is because an increase in employment reduces the proportion of tasks performed by each agent, which increases the individual probability of a positive shock. This increases the expected transfer when the agent chooses the first best actions, which allows the principal to reduce the second best incentive. We also show that the reduction in the second best bonus reduces the second best expected transfer, which is therefore also decreasing with employment. When this effect is sufficiently large, the marginal cost of hiring another worker can decline with employment and can even be *negative*. In that case, the second best marginal cost of employment is less than the (constant) first best marginal cost of employment and the second best firm can be larger than first best. When the second best firm is inefficiently large, it resembles a traditional nonunionized manufacturing firm with weak incentives, low expected pay, and an excessive degree of specialization and division of labor. Motivation is achieved not through strong incentives but rather because a worker who shirks will be fired for sure.

As employment increases, agents perform fewer and fewer tasks, the probability that the agent experiences a positive shock converges to 1, and the second best expected transfer converges to the first best transfer. The second best cost function therefore asymptotes to the linear first best cost function. As before, the increasing returns to employment due to specialization and division of labor cannot be contained by an asymptotically linear cost of employment. We conclude that when the principal can monitor individual output, the size of the O-ring firm under moral hazard is limited either by the extent of the market or the O-ring property but not the marginal cost of employment as suggested by Becker and Murphy.

Finally, we consider the third best monitoring technology, where the principal can only observe team output. In this case, the third best incentive depends on the team probability that all agents experience positive shocks rather than the individual probabilities. Under the O-ring property, an increase in employment increases the individual probability of a positive shock but reduces the team probability that all shocks are positive. The third best results are therefore the exact opposite of the second best case. An increase in employment reduces the team probability of success, which reduces the expected transfer to each agent when he chooses the first best actions. This induces the principal to increase the third best incentive, which also increases the third best expected transfer and marginal cost of hiring another worker. It follows that the third best incentive, expected transfer, and marginal cost of employment are all increasing in employment, which is the opposite of the second best case.

As employment increases still further, the third best incentive, expected transfer, and marginal cost of employment all explode rather than asymptoting to their first best levels as in the second best case. We show that the third best marginal cost of employment always exceeds the first and second best marginal costs of employment; therefore, the third best firm is generally smaller than the other two. It is only in this context, where the principal can only observe team output, that the size of the firm and the extent of the division of labor can be limited by moral hazard alone.

The relevant literature consists of only a handful of papers. This is because employment is endogenous in our model and an integral feature of our results, whereas in most of the contract theory literature, the number of workers is fixed and exogenous. Ziv (1993) considers a version of Holmström (1982) where team size is endogenous under specific functional forms for the production function, the agents' utility function, and the distribution of signals. When the signals of individual performance are contractible and exponentially distributed, the second best firm can be larger than first best. This is because the variance of the signal is increasing in effort under the exponential distribution. The principal therefore chooses less effort and more agents. In this paper, we obtain the same result under the assumption that the probability of a positive shock is decreasing in the number of tasks performed by the agent. It is therefore an intrinsic and robust feature of the O-ring firm rather than a consequence of specific distributional assumptions.

The remaining literature is set within the linear framework of Holmström and Milgrom (1987). Liang, Rajan, and Ray (2008) assume that only team performance is contractible and normally distributed, where the variance is assumed to be exogenously increasing in team size. Their paper is primarily interested in results where the optimal incentive is independent of the parameters of the model. Auriol, Friebel, and Pechlivanos (1999) consider the case where greater employment reduces the effectiveness of helping efforts (effort by one agent that increases the outputs of other agents). In their model, the second best firm is always smaller than first best when agents are risk averse. Their paper is primarily focused on commitment issues. Rauh (2014) provides conditions under which incentives and employment are substitutes or complements when only team performance is contractible. Like Becker and Murphy (1992), the model includes an *ad hoc* employment cost function to keep the optimal team size finite.

The plan for the rest of the paper is as follows. In Section 2, we present the model primitives. We consider the first best in Section 3, the second best in Section 4, and the third best in Section 5. Section 6 concludes. All proofs are in the Appendix.

2 | MODEL PRIMITIVES

There is one principal (she) and a large number N of identical agents (he). Let n be the number of agents employed by the principal, where n is an integer such that $0 \leq n \leq N$. We call N the *extent* of the (labor) market.

Let $S = [0, 1]$ be the set of tasks and q_s the level of output in task $s \in S$. Each unit of output Q requires one unit of output q_s in each task $s \in S$, so output is essentially the minimum of the task output levels

$$Q = \inf_{s \in S} q_s. \quad (1)$$

If $q_s = 0$ for any task $s \in S$, then $Q = 0$. This captures the idea that a mistake at one stage of production can have serious consequences for the firm.

After choosing the employment level n , the principal allocates the set S of tasks across the n employed agents. Let $S_i \subseteq S$ be the set of tasks assigned to agent i . We assume an equal division of labor in the sense that the principal partitions S into n intervals with equal measure $\rho = 1/n$. For example, if $n = 3$, then

$$S_1 = [0, 1/3), \quad S_2 = [1/3, 2/3), \quad \text{and} \quad S_3 = [2/3, 1], \quad (2)$$

and $\rho = 1/3$. An increase in employment n implies a finer partition of tasks and fewer tasks per worker and therefore a more extensive division of labor.

As discussed in Section 1, Becker and Murphy (1992) focus on Smith's contention that the division of labor is limited by the extent of the market, but it can also be limited by purely technological considerations. For example, Smith suggests that pin-making can be divided into at most 18 distinct operations. We can incorporate this technological constraint into the model as follows. Abusing notation, let N_L be the extent of the labor market and N_D the technological upper bound on the division of labor. The employment constraint is therefore $n \leq N$, where $N = \min\{N_L, N_D\}$. For simplicity, we continue to refer to N as "the extent of the labor market" because of our focus on the question "why is there not one big firm?" but for many small firms, it is the technological constraint that binds. In the case of pins, Smith states that in some manufactories, the 18 distinct operations "are all performed by distinct hands" but also that "I have seen a small manufactory of this kind where ten men only were employed..." In the former case, the effective constraint was presumably technological, and this alternative interpretation should be borne in mind in cases where the employment constraint binds.³

Given his task assignment S_i , agent i makes the following choices. First, he chooses his total effort level $T_i \in \{0, 1\}$. Let q_{is} denote output in task $s \in S_i$ and

$$q_i = \inf_{s \in S_i} q_{is}, \quad (3)$$

the *individual output* of agent i . If the agent shirks $T_i = 0$, then $q_{is} = 0$ for all $s \in S_i$, $q_i = 0$, and $Q = 0$. If he chooses $T_i = 1$, he must decide how to allocate it across his assigned tasks. Let τ_{is} be the proportion of $T_i = 1$ devoted to task $s \in S_i$, where⁴

$$T_i = \int_{s \in S_i} \tau_{is} ds. \quad (4)$$

Finally, the agent must allocate the total time τ_{is} devoted to task $s \in S_i$ between production effort e_{is} and investments in task-specific human capital l_{is} (the time spent learning about the task). This allocation must satisfy the constraints $e_{is} \geq 0$, $l_{is} \geq 0$, and

$$e_{is} + l_{is} \leq \tau_{is}. \quad (5)$$

We refer to l_{is} as the level of *specialization* of agent i in task s .

After making these decisions, each agent experiences an idiosyncratic shock. With probability $\pi(\rho)$, the shock is positive and output in task $s \in S_i$ is given by

$$q_{is}^+ = A l_{is}^\gamma e_{is}, \quad (6)$$

where $A > 0$ and $\gamma > 0$ are productivity parameters. Note that specialization l_{is} improves the productivity of production effort e_{is} . In this case, individual output is given by

$$q_i^+ = \inf_{s \in S_i} q_{is}^+. \quad (7)$$

With probability $1 - \pi(\rho)$, the shock is negative and his output is zero in at least one of his tasks. It follows that individual output q_i and team output Q are zero.

We assume that shocks are independent across agents so that $\pi(\rho)^n$ is the probability that all agents experience positive shocks, whereas $1 - \pi(\rho)^n$ is the probability that at least one agent experiences a negative shock. Let

$$Q^+ = \min_{1 \leq i \leq n} q_i^+. \quad (8)$$

The O-ring production function is therefore

$$Q = \begin{cases} Q^+ & \text{with probability } \pi(\rho)^n \\ 0 & \text{with probability } 1 - \pi(\rho)^n. \end{cases} \quad (9)$$

We make the following assumptions throughout the paper.

Assumption 1. *The probability $\pi(\rho)$ that an agent experiences a positive shock is*

- (i) *a continuous function of ρ on $[0, 1]$ and continuously differentiable on $(0, 1]$ with $\pi' < 0$,*
- (ii) *$\pi(1) > 0$ and $\pi(0) = 1$, and*
- (iii) *$\pi(\rho)^n \rightarrow 0$ as $n \rightarrow \infty$ (the first O-ring property).*

Since $\pi' < 0$, the probability that an agent experiences a positive shock is decreasing in the proportion of tasks ρ assigned to him as discussed in Section 1. The condition $\pi(1) > 0$ means that the probability of a positive shock remains positive even when one agent performs all the tasks. We do not want to rule out a single agent firm *ex ante*. At the opposite end of the spectrum, $\pi(0) = 1$ means that the probability of a positive shock approaches one as the proportion of assigned tasks approaches zero.⁵ An increase in employment n reduces the proportion of tasks $\rho = 1/n$ performed by each agent and raises the probability $\pi(\rho)$ of a positive shock. But it also increases the number of shocks, so the effect of employment on the probability $\pi(\rho)^n$ that the team is successful is ambiguous. We assume that eventually, the latter effect dominates in the sense of the *first O-ring property*: $\pi(\rho)^n \rightarrow 0$ as $n \rightarrow \infty$ (there will be a second O-ring property later). This will act as one force that limits the size of the firm. In the Appendix, we show that the following example satisfies all of the above assumptions.

Example. *Let $\pi(\rho) = (\rho/D)^\rho$ and $\pi(0) \equiv 1$, where D is a constant such that $D > e$, where $e \approx 2.718$ is Euler's constant.*

Each agent has the utility function

$$U(t_i, T_i) = u(t_i) - c(T_i), \quad (10)$$

where t_i is the transfer from the principal to agent i , $c(0) = 0$, and $c(1) = \psi > 0$. We assume that agents cannot be forced to pay a fine, so the transfer must be nonnegative $t_i \geq 0$. We make the following standard assumptions.

Assumption 2. *The utility of money $u : [0, \infty) \rightarrow \mathbb{R}$ is*

- (i) *continuous,*
- (ii) *twice continuously differentiable on $(0, \infty)$ with $u' > 0$ and $u'' < 0$, and*
- (iii) *$u(0) = 0$.*

Since u is increasing, we can define its inverse $t = h(u)$, where $h = u^{-1}$. Note that $u' > 0$ and $u'' < 0$ imply $h' > 0$ and $h'' > 0$. All agents have an outside option of zero. The following example will be used repeatedly throughout the paper.

Main Example. *Let $u(t_i) = \sqrt{t_i}$ and $\pi(\rho) = (\rho/D)^\rho$ as defined above.*

The production function (9) combines the main elements in Becker and Murphy (1992) and Kremer (1993). Becker and Murphy develop a model of specialization and division of labor that leads to the production function defined by (1) and (6). Their model is deterministic in the sense that $\pi(\rho)$ is identically equal to 1. Kremer considers an O-ring production function

similar to that in (9) but without specialization and division of labor and where $\pi(\rho)$ is a constant that varies across workers but does not depend on ρ . Neither paper explicitly considers moral hazard.

3 | THE FIRST BEST

We begin with the first best benchmark where the principal can monitor all the agents' choices. In this case, the principal can effectively make all the agents' choices for them because she can specify the actions that she prefers as part of the contract.⁶ The principal therefore chooses employment $n \geq 0$, total efforts $\{T_i\}$, task efforts $\{\tau_{is}\}$, production efforts $\{e_{is}\}$, and task-specific human capital $\{l_{is}\}$ to maximize her expected profit. Given these choices, the individual output q_i of agent i is

$$q_i = \begin{cases} q_i^+ & \text{with probability } \pi(\rho) \\ 0 & \text{with probability } 1 - \pi(\rho), \end{cases} \quad (11)$$

where q_i^+ is defined in (7) above. Let \bar{t}_i be the transfer to agent i when $q_i = q_i^+$ and \underline{t}_i the transfer when $q_i = 0$. The principal chooses all of the above to maximize

$$\Pi^F = \pi(\rho)^n Q^+ - \sum_{i=1}^n \{ \pi(\rho) \bar{t}_i + [1 - \pi(\rho)] \underline{t}_i \} \quad (12)$$

subject to the constraints (4) and (5), $T_i \in \{0, 1\}$ for all i , and the relevant participation constraints discussed in a moment. The first term in (12) is expected output from the O-ring production function (9) and the second is the sum of the expected transfers across employed agents.⁷ If $n = 0$, then $\Pi^F \equiv 0$.

Given that agent i does not shirk, so that $T_i = 1$, our first result shows how the total time $T_i = 1$ should be allocated across tasks and between production efforts and investments in human capital within tasks.

Lemma 1. *The allocation of $T_i = 1$ that maximizes q_i^+ is given by*

$$e_{is} = \frac{1}{\rho(1 + \gamma)}, \quad l_{is} = \frac{\gamma}{\rho(1 + \gamma)}, \quad \text{and} \quad \tau_{is} = \frac{1}{\rho} \quad (13)$$

for all i and $s \in S_i$. In that case, $q_i^+ = Bn^{1+\gamma}$ for all i , where

$$B = \frac{A}{\gamma} \left(\frac{\gamma}{1 + \gamma} \right)^{1+\gamma}. \quad (14)$$

When agent i experiences a positive shock, his individual output q_i^+ in (7) is essentially the minimum of his task outputs q_{is}^+ in (6). It is therefore clearly optimal to allocate the total time $T_i = 1$ equally across all assigned tasks: $\tau_{is} = T_i/\rho = 1/\rho$ for all $s \in S_i$. One can then easily verify that e_{is} and l_{is} in (13) maximize q_{is}^+ subject to the constraints $e_{is} \geq 0$, $l_{is} \geq 0$, and (5). Substituting the efficient allocation (13) into (6), individual output is $q_i = Bn^{1+\gamma}$ when the shock is positive. When all shocks are positive, team output is the minimum of the individual outputs $Q = Bn^{1+\gamma}$, which exhibits increasing returns to employment n . This is because an increase in employment implies fewer tasks per agent, who can then devote more time τ_{ij} to each of their assigned tasks. This improves productivity through greater investments l_{ij} in task-specific human capital.

Definition 1. *We say the principal implements efficient effort if all agents i choose $T_i = 1$ and allocate it as in Lemma 1. We refer to $q^h = Bn^{1+\gamma}$ as high individual output and $Q^h = Bn^{1+\gamma}$ as high team output. We say the principal implements inefficient effort if all agents i choose $T_i = 1$ but at least one of them allocates it differently.*

We now list the options available to the principal:

- (1) zero employment $n = 0$, which gives zero profit $\Pi^F = 0$,
- (2) $n \geq 1$, at least one agent shirks $T_i = 0$ (so team output is zero for sure), and $\Pi^F \leq 0$,
- (3) $n \geq 1$ and inefficient effort, and
- (4) $n \geq 1$ and efficient effort.

We assume that the principal prefers $n = 0$ over $n \geq 1$ and $T_i = 0$ for all i ,⁸ so the principal's choice boils down to options (1) and (4). For the rest of the paper, we assume that in all contexts, the principal chooses the latter.⁹ Under this assumption, the principal chooses $n \geq 1$ and the transfers $\bar{t}_i \geq 0$ and $\underline{t}_i \geq 0$ to maximize

$$\Pi^F = \pi(\rho)^n Q^h - \sum_i \{ \pi(\rho) \bar{t}_i + [1 - \pi(\rho)] \underline{t}_i \} \quad (15)$$

subject to the participation constraints

$$\pi(\rho)u(\bar{t}_i) + [1 - \pi(\rho)]u(\underline{t}_i) - \psi \geq 0. \quad (16)$$

Lemma 2. *The first best solution entails a constant transfer $\bar{t}_i = \underline{t}_i = h(\psi)$ for all i .*

As usual, the first best transfers provide full insurance $\bar{t}_i = \underline{t}_i = h(\psi)$, which is the same constant transfer as in the textbook model with one agent.¹⁰ In our model, this standard result is important because it implies that the first best cost of employment

$$C^F(n, \psi) = nh(\psi) \quad (17)$$

is linear in employment n . Substituting these transfers into (15),

$$\Pi^F = \pi(\rho)^n Q^h - C^F(n, \psi). \quad (18)$$

Let $n^F(B, \gamma, \psi)$ denote the first best employment correspondence.¹¹

The first best optimal employment level(s) is determined by the interaction of four factors. The first is the extent N of the labor market. The second is the increasing returns to employment $Q^h = Bn^{1+\gamma}$ due to specialization and division of labor. The third is the first O-ring property, which states that the team probability $\pi(\rho)^n$ gets arbitrarily small as employment increases. Finally, there is the cost of employment C^F . But since C^F is linear in employment, it cannot limit the size of the firm or the extent of the division of labor. The first best O-ring firm will therefore hire the entire labor pool unless the first O-ring property is sufficiently strong. In that case, it is efficient for all production to take place in one big firm.

Proposition 1. *Let $L = \lim_{n \rightarrow \infty} \pi(\rho)^n n^\gamma$.*

- (i) *If $0 \leq L < \frac{h(\psi)}{B}$ and N is sufficiently large, then $n < N$ for all $n \in n^F(B, \gamma, \psi)$.*
- (ii) *If $\frac{h(\psi)}{B} < L \leq \infty$ and N is sufficiently large, then $n^F = N$.*

The limit L is determined by two factors that work in opposite directions: the first O-ring property $\pi(\rho)^n \rightarrow 0$ and the increasing output per head $\frac{Bn^{1+\gamma}}{n} = Bn^\gamma \rightarrow \infty$. In (i), the limit L is relatively small, which requires $\pi(\rho)^n \rightarrow 0$ to occur sufficiently rapidly. In this case, the first O-ring property is stronger than the increasing output per head effect and the size of the first best firm and the extent of the division of labor are limited by the O-ring property and not the extent N of the labor market. Indeed, $n < N$ for all optimal employment levels $n \in n^F(B, \gamma, \psi)$. In (ii), L is relatively large, so the divergence in output per head is stronger and the first best firm is limited only by the size N of available labor pool.

Main Example. *In our Main Example, $\frac{h(\psi)}{B} = \frac{\psi^2}{B}$ and $\pi(\rho)^n n^\gamma = (1/D)n^{\gamma-1}$, which converges to*

$$L = \begin{cases} 0 & \text{when } 0 \leq \gamma < 1 \\ 1/D & \text{when } \gamma = 1 \\ \infty & \text{when } \gamma > 1. \end{cases} \quad (19)$$

The first best expected profit is

$$\Pi^F = \frac{B}{D}n^\gamma - n\psi^2 \quad (20)$$

with critical point

$$n_F^* = \left(\frac{B\gamma}{D\psi^2} \right)^{\frac{1}{1-\gamma}} \quad (21)$$

for all $0 \leq \gamma < 1$. If $L = 0$, we are in case (i) of the above result where the first best firm is limited by the first O-ring property. In fact, (21) is the first best employment level if it happens to be an integer. If not, then the first best employment level is one or both of the two closest integers. If $L = 1/D$, then

$$n^F = \begin{cases} 0 & \text{if } \frac{1}{D} < \frac{\psi^2}{B} \\ \text{indeterminate} & \text{if } \frac{1}{D} = \frac{\psi^2}{B} \\ N & \text{if } \frac{1}{D} > \frac{\psi^2}{B}. \end{cases} \quad (22)$$

The first case $1/D < \psi^2/B$ falls under (i) of the proposition, while the third $1/D > \psi^2/B$ falls under (ii). Note that the second case $1/D = \psi^2/B$, where optimal employment is indeterminate, is not covered by the proposition. Finally, if $L = \infty$, then (ii) applies, where $n^F = N$.

4 | THE SECOND BEST

We now introduce moral hazard. In this section, we consider the second best case where the principal can monitor individual output q_i but not the agents' choices $\{e_{is}\}$, $\{l_{is}\}$, $\{\tau_{is}\}$, or $\{T_i\}$. Let $t_i(q_i)$ be the transfer to agent i conditional on q_i . Note that the principal does not condition transfers on team output Q because that would impose additional risk without adding any new information.

As before, the principal effectively chooses between zero employment and positive employment with efficient effort. When agent i exerts efficient effort, his individual output is given by

$$q_i = \begin{cases} q^h & \text{with probability } \pi(\rho) \\ 0 & \text{with probability } 1 - \pi(\rho). \end{cases} \quad (23)$$

A principal who implements positive employment with efficient effort chooses employment $n \geq 1$ and transfers $t_i(q_i)$ to maximize expected profit

$$\Pi^I = \pi(\rho)^n Q^h - \sum_{i=1}^n \{ \pi(\rho)t_i(q^h) + [1 - \pi(\rho)]t_i(0) \} \quad (24)$$

subject to the limited liability constraints $t_i(q_i) \geq 0$ for all $0 \leq q_i \leq q^h$, the participation constraints

$$\pi(\rho)u(t_i(q^h)) + [1 - \pi(\rho)]u(t_i(0)) - \psi \geq 0, \quad (25)$$

and the relevant incentive compatibility constraints. To determine the latter, we list the alternatives available to agent i . If he shirks $T_i = 0$, then $q_i = 0$ for sure. If he chooses inefficient effort, then

$$q_i = \begin{cases} \hat{q}_i & \text{with probability } \pi(\rho) \\ 0 & \text{with probability } 1 - \pi(\rho), \end{cases} \quad (26)$$

where $0 \leq \hat{q}_i < q^h$. The incentive compatibility constraints are therefore

$$\pi(\rho)u(t_i(q^h)) + [1 - \pi(\rho)]u(t_i(0)) - \psi \geq u(t_i(0)) \quad (27)$$

$$\pi(\rho)u(t_i(q^h)) + [1 - \pi(\rho)]u(t_i(0)) - \psi \geq \pi(\rho)u(t_i(\hat{q}_i)) + [1 - \pi(\rho)]u(t_i(0)) - \psi \quad (28)$$

for all $0 \leq \hat{q}_i < q^h$. Constraint (27) states that agent i prefers efficient effort over shirking, while (28) states that i prefers efficient effort over inefficient effort.

Lemma 3.

(i) For any employment level $n \geq 1$, the transfers

$$t_i(q_i) = \begin{cases} h \left[\frac{\psi}{\pi(\rho)} \right] & \text{if } q_i = q^h \\ 0 & \text{if } 0 \leq q_i < q^h \end{cases} \quad (29)$$

minimize the expected cost of implementing efficient effort.

- (ii) The second best incentive $h(\psi/\pi)$ is increasing in the proportion ρ of tasks assigned to agent i and decreasing in employment n .
- (iii) $h(\psi/\pi) \downarrow h(\psi)$ as $n \rightarrow \infty$.

When the agent chooses efficient effort and the shock is positive, his individual output is $q_i = q^h$ and the principal rewards him with the bonus $h(\psi/\pi)$. We call this the *second best incentive*. In all other cases (when the shock is negative or if he chooses inefficient effort or shirks), then $0 \leq q_i < q^h$ and the transfer is zero. Since his outside option is zero, this is effectively the same as firing the agent.¹² An increase in employment n implies fewer tasks per agent ρ and a higher probability $\pi(\rho)$ of a positive shock. An increase in $\pi(\rho)$ raises the expected payoff of the agent when he chooses efficient effort (the left-hand side of the incentive compatibility constraint (28)), which allows the principal to continue implementing efficient effort but with weaker incentives. It follows that $h(\psi/\pi)$ is increasing in ρ and decreasing in employment n . As $n \rightarrow \infty$, $\pi(\rho) \rightarrow 1$, and the second best incentive $h(\psi/\pi)$ converges to the first best transfer $h(\psi)$. This must be the case because in the limit, when the probability of a positive shock is one, the principal can perfectly monitor the agents: if $q_i = q^h$, she can deduce that the agent chose efficient effort and if $0 \leq q_i < q^h$, she can deduce that the agent either shirked or chose inefficient effort.

It is sometimes stated that workers on an assembly line should not receive any incentives, even under moral hazard, because they cannot vary the intensity of their effort. For example, consider the case with only one agent who chooses a continuous effort level e at cost $C(e)$. Assume the agent has a negative exponential utility function with constant coefficient of risk aversion $r > 0$. Output is given by $q = e + \epsilon$, where ϵ is normally distributed with mean zero and variance V . Let $P(e)$ be the expected benefit of effort to the principal who is restricted to the class of linear contracts $\alpha + \beta q$. As is well known, under these assumptions, the optimal incentive is

$$\beta = \frac{P'(e)}{1 + rVC''(e)}, \quad (30)$$

where primes indicate derivatives. Milgrom and Roberts (1992, p. 222) define $1/C''$ as the agent's *responsiveness to incentives* and argue that "an employee working on a fixed rate production line cannot increase his or her own output in response to piece rate incentives" so that responsiveness and the optimal incentive should both zero.

There are two issues with this argument. First of all, it neglects the fact that under moral hazard, the agent still has the option of withholding effort altogether. In our model, the optimal incentive contract in Lemma 3 is a *Mirrlees* contract where the agent is rewarded when output is q^h and punished otherwise. The optimal contract is therefore designed and dedicated to deterring shirking (zero effort) and inefficient effort. Furthermore, it is well known that linear contracts are not optimal under the assumptions in the previous paragraph and that the principal can approximate the first best outcome using a *Mirrlees* contract.¹³ In our model, the decision to exert efficient effort depends on the reward $t_i(q^h)$ as well as the probability $\pi(\rho)$ of a positive shock. Since $\pi(\rho)$ increases with employment n , the incentive $h(\psi/\pi)$ declines. In our model, it is only at high employment levels that the principal offers weak incentives (i.e., a bonus close to the first best transfer) and incentives can be quite strong when the team size is small.

We now consider the second best expected transfer

$$H^I(\rho, \psi) = \pi(\rho)h \left[\frac{\psi}{\pi(\rho)} \right]. \quad (31)$$

Lemma 4.

- (i) The second best expected transfer exceeds the first best transfer $H^I(\rho, \psi) > h(\psi)$ for all $0 < \rho \leq 1$ and $\psi > 0$.
- (ii) $H^I(\rho, \psi)$ is increasing in ρ and decreasing in n for all $\psi > 0$.
- (iii) $H^I(\rho, \psi) \downarrow h(\psi)$ as $n \rightarrow \infty$.

The second best expected transfer exceeds the first best transfer for the usual reason that it includes agency costs. In particular, under moral hazard, the principal can only implement efficient effort by offering incentives. This imposes risk on the agents that obliges the principal to pay a risk premium. An increase in employment n reduces the proportion ρ of tasks performed by each worker, which has two conflicting effects on the expected transfer H^I : it increases the probability $\pi(\rho)$ of a positive shock but reduces the incentive $h(\psi/\pi)$. Since agents are risk averse, h is strictly convex, and the latter effect dominates. The second best expected transfer is therefore decreasing in employment n . As $n \rightarrow \infty$, $\rho \rightarrow 0$, and $\pi(\rho) \rightarrow 1$, the second best incentive $h(\psi/\pi)$ and expected transfer $H^I(\rho, \psi)$ decline monotonically to their first best levels $h(\psi)$. In the present context, an increase in employment reduces agency costs that disappear in the limit.

The model therefore predicts that incentives and expected pay are decreasing in the level of specialization and division of labor when individual performance can be monitored. This seemingly contradicts the fact that in many occupations such as medicine, we observe that specialists are better paid than nonspecialists. But it is important to note that in our model, specialization makes the worker more productive within the O-ring firm but does not affect the agents' outside option. Our model therefore focuses on *firm-specific* specialization and is more appropriate for manufacturing, for example, than physicians whose specialization increases their *general* human capital.

Substituting the second best transfers into (24), we obtain

$$\Pi^I = \pi(\rho)^n Q^h - C^I(n, \psi), \tag{32}$$

where

$$C^I(n, \psi) = nH^I(\rho, \psi) \tag{33}$$

is the second best cost of employment. The second best expected transfer H^I exceeds the first best transfer $h(\psi)$, so $C^I > C^F$ for all $n \geq 1$ because C^I includes the cost of moral hazard. This suggests two questions. First, can moral hazard limit the size of the second best firm? And second, how does the size of the second best firm compare to the first best?

In the first best benchmark, the size of the firm and the division of labor are limited either by the extent N of the labor market or the first O-ring property $\pi(\rho)^n \rightarrow 0$ but not the first best cost of employment C^F because it is linear. In the present context, the second best expected transfer $H^I(\rho, \psi)$ converges to the first best transfer $h(\psi)$ as $n \rightarrow \infty$, so the second best cost C^I of employment is approximately the same as C^F at high employment levels. It follows that C^I is *asymptotically linear* and therefore cannot compete with the increasing returns to employment due to specialization and division of labor. As before, the size of the firm is limited by the extent of the labor market or the first O-ring property and not by agency costs when individual output can be monitored. The following result is almost exactly the same as Proposition 1 in the first best case. Let $n^I(B, \gamma, \psi)$ be the second best employment correspondence.

Proposition 2. Let $L = \lim_{n \rightarrow \infty} \pi(\rho)^n n^\gamma$.

- (i) If $0 \leq L < \frac{h(\psi)}{B}$ and N is sufficiently large, then $n < N$ for all $n \in n^I(B, \gamma, \psi)$.
- (ii) If $\frac{h(\psi)}{B} < L \leq \infty$ and N is sufficiently large, then $n^I = N$.

We now turn to the second question, on the size of the second best firm relative to the first best. The answer depends on the second best marginal cost of employment C_n^I , where the subscript indicates partial differentiation.¹⁴ Let

$$\epsilon_\pi = \frac{\rho \pi'}{\pi} \tag{34}$$

be the elasticity of π with respect to ρ and

$$\epsilon_h = \frac{\psi h'(\psi/\pi)}{\pi h(\psi/\pi)}, \tag{35}$$

the elasticity of h with respect to $u = \psi/\pi$.

Lemma 5. The second best marginal cost of employment is given by

$$C_n^I = H^I(\rho, \psi) \Delta^I(n, \psi), \tag{36}$$

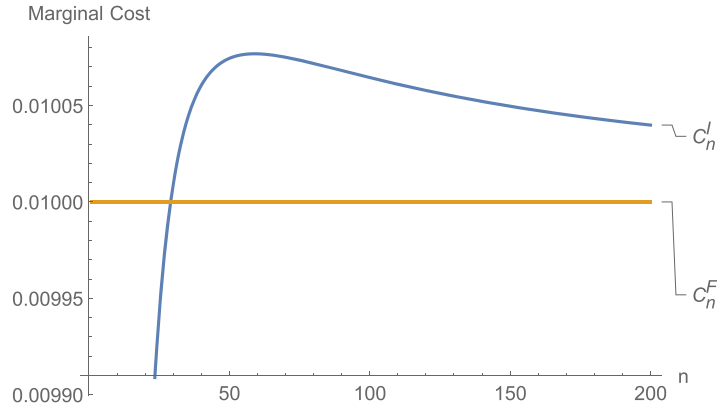


FIGURE 1 First and second best marginal costs of employment [Color figure can be viewed at wileyonlinelibrary.com]

where

$$\Delta^I(n, \psi) = 1 - \epsilon_\pi(1 - \epsilon_h) < 1. \tag{37}$$

The marginal cost of employment C_n^I can be decomposed into two separate components. The first component H^I reflects the fact that each additional hire must be paid the expected transfer, while the second Δ^I reflects the reduction in H^I due to an increase in employment n . Since $H^I = \pi h$, the magnitude of Δ^I depends on the elasticities ϵ_π and ϵ_h , where $\epsilon_\pi < 0$ because $\pi' < 0$ and $\epsilon_h > 1$ (we prove this in the Appendix) because agents are risk averse and h is therefore strictly convex. It follows that $\Delta^I < 1$, so the marginal cost of employment is less than the expected transfer. In the textbook monopsony model, the marginal cost of employment exceeds the wage because the firm must increase the wage to hire more labor and this wage increase must be applied to its current workers as well as the marginal ones. In contrast, the second best marginal cost of employment is less than the expected transfer because the latter is decreasing in employment.

The larger the elasticities ϵ_π and ϵ_h are in absolute value, the smaller the marginal cost of employment. When ϵ_π is large in absolute value, an increase in employment n leads to a large increase in the probability π of a positive shock. When ϵ_h is large, an increase in π leads to a large reduction in the incentive $h(\psi/\pi)$. Combining these two effects, an increase in employment will lead to a large reduction in the expected transfer H^I , which implies small values for Δ^I and the marginal cost of employment C_n^I . In fact, our next example shows that the marginal cost of employment can even be negative.

Example. Consider our Main Example with $B = 7$, $\gamma = 1/2$, $D = 100$, and $\psi = 1/10$. In Figure 1 above, we plot the first $C_n^F = \psi^2$ and second

$$C_n^I = \psi^2 \left(\frac{D}{\rho}\right)^\rho \left[1 + \rho \left(1 + \log \frac{\rho}{D}\right)\right] \tag{38}$$

best marginal costs of employment as functions of n .

Although not shown in the figure, $\Delta^I < 0$ and therefore $C_n^I < 0$ from $n = 1$ until $n \approx 5.27$. For example, one can verify by hand that $C^I = 1$ when $n = 1$ and $C^I \approx 0.28$ when $n = 2$. In this range, an increase in employment reduces the expected transfer so much that the marginal cost of employment is actually negative. The two marginal cost curves cross at $n \approx 29$. After that, the second best marginal cost C_n^I gradually asymptotes to the first best C_n^F as the second best expected transfer $H^I(\rho, \psi)$ converges to the first best transfer $h(\psi)$. In Figure 2 below, we plot first best Π^F and second best Π^I expected profit as functions of employment.

Not surprisingly, first best expected profit exceeds second best expected profit at every employment level. From (18) and (32), we observe that expected revenue $\pi(\rho)^n Q^h$ is the same for the first and second best O-ring firms, so any difference between the first and second best employment levels is explained entirely by the difference in the marginal cost of employment. Note that $\Pi^I < 0$ for all $n \geq 28$ so $C_n^I < C_n^F$ for all employment levels where the second best expected profit is nonnegative. We would therefore expect that $n^I > n^F$ and indeed, $n^I = 14$ and $n^F = 12$ assuming that $N \geq 14$ and imposing the integer constraint.

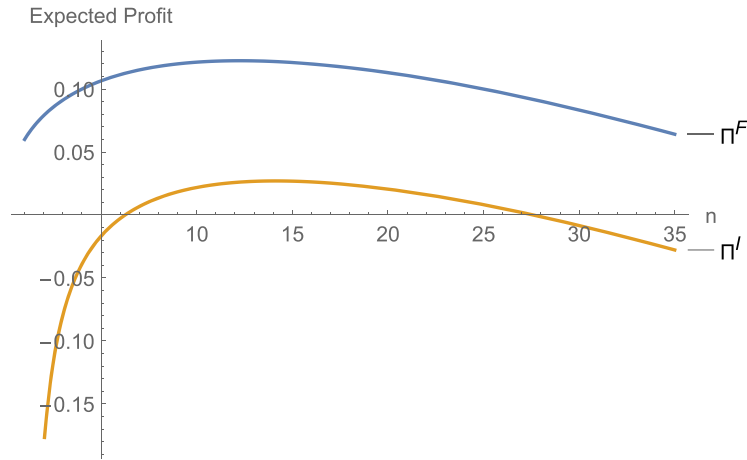


FIGURE 2 First and second best expected profit [Color figure can be viewed at wileyonlinelibrary.com]

Our final result in this section provides conditions under which the second best firm is smaller than the first best firm. Since both problems can have multiple solutions, we consider the largest first best employment level defined by

$$\bar{n}^F(B, \gamma, \psi) = \max_n n^F(B, \gamma, \psi). \tag{39}$$

Proposition 3. Let $\mathbb{N}_+^I(B, \gamma, \psi)$ be the set of integers $0 \leq n \leq N$ such that $\Pi^I \geq 0$. If $C_n^I(n, \psi) \geq C_n^F(\psi) = h(\psi)$ for all $n \in \mathbb{N}_+^I(B, \gamma, \psi)$, then $n \leq \bar{n}^F(B, \gamma, \psi)$ for all $n \in n^I(B, \gamma, \psi)$.

From Lemma 5, the size of the second best firm relative to the first best depends on the absolute values of the elasticities ϵ_π and ϵ_h . When these elasticities are small in absolute value, Δ^I is close to one and the second best marginal cost of employment C_n^I is close to the second best expected transfer H^I , which exceeds the first best marginal cost of employment $C_n^F = h(\psi)$. This is the situation described in the above result. In contrast, when the elasticities ϵ_π and ϵ_h are large, the second best marginal cost of employment is small and can even be negative as in the above example. In that case, the second best firm will be inefficiently large with an excessive degree of specialization and division of labor, weak incentives, and low expected pay relative to the outside option. Motivation is achieved through a combination of weak incentives and the fact that the agent knows that he will be fired for sure if he shirks. At high employment levels, the individual probability of a negative shock and the probability that a given individual agent will be fired are both relatively low. But the team probability π^n that all shocks are positive is also relatively low, resulting in occasional breakdowns in production (e.g., product recalls) and subsequent firings. The agents who are not fired receive relatively low wages (the second best bonus). The overall picture resembles a traditional manufacturing firm with extensive specialization and division of labor, weak incentives, and low pay.

5 | THIRD BEST

We now consider the third best case where the principal can only monitor team output Q . A principal who chooses positive employment and implements efficient effort chooses n such that $1 \leq n \leq N$ and transfers $t_i(Q)$ to maximize expected profit

$$\pi(\rho)^n Q^h - \sum_{i=1}^n \{ \pi(\rho)^n t_i(Q^h) + [1 - \pi(\rho)^n] t_i(0) \} \tag{40}$$

subject to the limited liability constraints $t_i(Q) \geq 0$ for all $0 \leq Q \leq Q^h$, the participation

$$\pi(\rho)^n u(t_i(Q^h)) + [1 - \pi(\rho)^n] u(t_i(0)) - \psi \geq 0 \tag{41}$$

and incentive compatibility constraints

$$\pi(\rho)^n u(t_i(Q^h)) + [1 - \pi(\rho)^n] u(t_i(0)) - \psi \geq u(t_i(0)) \tag{42}$$

$$\pi(\rho)^n u(t_i(Q^h)) + [1 - \pi(\rho)^n] u(t_i(0)) - \psi \geq \pi(\rho)^n u(t_i(\hat{Q})) + [1 - \pi(\rho)^n] u(t_i(0)) - \psi \quad (43)$$

for all i and $0 \leq \hat{Q} < Q^h$. The incentive compatibility constraints ensure that it is a Nash equilibrium for all employed agents to exert efficient effort. Given that all other agents choose efficient effort, (42) ensures that agent i prefers efficient effort over zero effort, while (43) ensures that i prefers efficient effort over inefficient effort. The proof of the following result is similar to that for Lemmas 3 and 4 and is omitted.

Lemma 6.

- (i) For any employment level $1 \leq n \leq N$, the transfers that minimize the expected cost of implementing the efficient effort Nash equilibrium are

$$t_i(Q) = \begin{cases} h \left[\frac{\psi}{\pi(\rho)^n} \right] & \text{if } Q = Q^h \\ 0 & \text{if } 0 \leq Q < Q^h \end{cases} \quad (44)$$

for all i .

- (ii) The third best expected transfer

$$H^T(n, \psi) = \pi(\rho)^n h \left[\frac{\psi}{\pi(\rho)^n} \right] \quad (45)$$

exceeds the first and second best $h(\psi) < H^I < H^T$ for all $1 < n \leq N$ and $\psi > 0$.

The third best incentive and expected transfer have the same structure as their second best counterparts except that they depend on the probability $\pi(\rho)^n$ that all shocks are positive rather than the individual probability $\pi(\rho)$ as before. The team probability is less than the individual probability for all $n > 1$, so the third best incentive exceeds the second best incentive and the first best transfer

$$h \left[\frac{\psi}{\pi(\rho)^n} \right] > h \left[\frac{\psi}{\pi(\rho)} \right] > h(\psi). \quad (46)$$

Intuitively, a lower probability of success requires a higher reward for success to implement efficient effort. Since h is strictly convex, the third best expected transfer H^T exceeds the second best expected transfer H^I , which, in turn, exceeds the first best transfer $h(\psi)$ by Lemma 4. We make the following assumptions for the rest of the paper.

Assumption 3.

- (i) The marginal utility of money $u' \rightarrow 0$ as $t \rightarrow \infty$ and
(ii) the team probability $\pi(\rho)^n$ is decreasing in n (the second O-ring property)

The first assumption is that the marginal utility of money goes to zero as income gets arbitrarily large. Alternatively, $h' \rightarrow \infty$ as $u \rightarrow \infty$.¹⁵ This property is satisfied by many utility functions such as \sqrt{t} and $\ln t$. The second O-ring property is that the team probability $\pi(\rho)^n$ is decreasing in employment for all $n \geq 1$.¹⁶ The example below shows that our Main Example satisfies this assumption. The individual $\pi(\rho)$ and team probabilities $\pi(\rho)^n$ therefore behave quite differently. Under the first O-ring property, $\pi^n \rightarrow 0$ as $n \rightarrow \infty$, whereas $\pi \rightarrow 1$. Under the second O-ring property, $\pi(\rho)$ is increasing, whereas $\pi(\rho)^n$ is decreasing in employment n . This explains why the monitoring technology (second or third best) is so decisive for our results.

Lemma 7. For all $\psi > 0$,

- (i) the third best incentive $h(\psi/\pi^n)$ and expected transfer $H^T(n, \psi)$ are increasing in employment n and
(ii) $h(\psi/\pi^n) \rightarrow \infty$ and $H^T(n, \psi) \rightarrow \infty$ as $n \rightarrow \infty$.

An increase in employment n reduces the probability $\pi(\rho)^n$ the agent will be rewarded when he exerts efficient effort. To continue implementing efficient effort, the principal must increase the incentive. Since h is strictly convex, the increase in $h(\psi/\pi^n)$ dominates the reduction in the team probability $\pi(\rho)^n$ and the expected transfer $H^T(n, \psi)$ also increases. As $n \rightarrow \infty$ and $\pi^n \rightarrow 0$, the principal has to increase the incentive more and more because the marginal utility of money is declining and

converging to zero. The incentive and expected transfer therefore explode, which is quite different from the second best case where they both converge to their first best levels.

Substituting the third best transfers into (40), we obtain

$$\Pi^T = \pi(\rho)^n Q^h - C^T(n, \psi), \tag{47}$$

where $C^T(n, \psi) = nH^T(n, \psi)$.

Lemma 8.

(i) *The third best marginal cost of employment is given by*

$$C_n^T(n, \psi) = H^T(n, \psi)\Delta^T(n, \psi), \tag{48}$$

where

$$\Delta^T(n, \psi) = 1 + \epsilon_{\pi^n}(1 - \epsilon_h), \tag{49}$$

and ϵ_{π^n} is the elasticity of $\pi(\rho)^n$ with respect to n .

(ii) $\Delta^T > 1$ for all $n \geq 1$ and $\psi > 0$.

(iii) $C_n^T \rightarrow \infty$ as $n \rightarrow \infty$.

As before, the marginal cost of employment can be decomposed into two components: the first H^T is that each additional hire receives the expected transfer, while the second Δ^T captures the effect of the marginal hire on H^T .¹⁷ Since $H^T = \pi^n h$, Δ^T depends on the elasticities ϵ_{π^n} and ϵ_h . As before, the inverse utility of money is elastic $\epsilon_h > 1$ because h is strictly convex. The difference is that $\epsilon_{\pi^n} < 0$ because the team probability π^n is decreasing in employment n . It follows that $\Delta^T > 1$, so the marginal cost of employment exceeds the expected transfer. This is similar to the textbook monopsony model where the marginal cost of employment exceeds the wage because the firm must increase the wage to all workers, old and new, to hire additional labor. The difference is that in our model, the explanation is in terms of moral hazard rather than a traditional upward-sloping supply curve. Since $\Delta^T > 1$ and $H^T \rightarrow \infty$, $C_n^T = H^T \Delta^T \rightarrow \infty$ as $n \rightarrow \infty$.

Proposition 4.

(i) *The third best marginal cost of employment C_n^T is strictly greater than the first best $h(\psi)$ and second best C_n^I marginal costs of employment for all $n \geq 1$ and $\psi > 0$.*

(ii) *Assuming unique solutions for employment, the third best employment level is at least weakly less than the first and second best employment levels.*

The first part of the proposition follows from previous results: $H^T \geq H^I > h(\psi)$, $\Delta^T > 1 > \Delta^I$, and therefore

$$C_n^T = H^T \Delta^T > C_n^I = H^I \Delta^I, \tag{50}$$

$$C_n^T = H^T \Delta^T > C_n^F = h(\psi). \tag{51}$$

Intuitively, the third best marginal cost of employment C_n^T exceeds the third best expected transfer H^T because the team probability π^n is decreasing in employment n , which necessitates an increase in the third best incentive. The third best expected transfer H^T exceeds the second best expected transfer H^I because the team probability π^n is less than the individual probability π , so the third best incentive must be greater. Finally, the second best marginal cost C_n^I is less than the second best expected transfer H^I because the individual probability π is increasing in employment, which allows the principal to reduce the second best incentive. The ranking is therefore clear:

$$C_n^T > H^T \geq H^I > C_n^I. \tag{52}$$

The comparison between C_n^T and the first best marginal cost $h(\psi)$ is similar.

Since expected revenue as a function of employment is the same for all three institutions, the comparison of optimal employment levels in the second part of the proposition follows immediately from the comparison of the marginal costs in the first part. The uniqueness assumption simplifies the exposition. Otherwise we would have to make a more complicated statement similar to that in Proposition 3 above.

Main Example. We illustrate these ideas using our Main Example. Recall that the first best cost function $C^F = nh(\psi)$ is linear in employment n and the second best cost function C^I is asymptotically linear in employment because

$$\frac{C^I}{C^F} = \frac{nH^I}{nh(\psi)} = \frac{H^I}{h(\psi)} \rightarrow 1 \quad (53)$$

as $n \rightarrow \infty$. In our Main Example, the team probability is $\pi^n = \frac{1}{nD}$ and $C^T = D\psi^2 n^2$, which dominates the returns to employment $Q^h = Bn^{1+\gamma}$ when $0 \leq \gamma < 1$. It is only in the third best case, when the principal can only monitor team output Q , that moral hazard alone can limit the size of the firm and the extent of the division of labor.

The third best expected profit is given by

$$\Pi^T = \pi^n Bn^{1+\gamma} - C^T(n, \psi) = \frac{B}{D} n^\gamma - D\psi^2 n^2, \quad (54)$$

with critical point

$$n_T^* = \left(\frac{B\gamma}{2D^2\psi^2} \right)^{\frac{1}{2-\gamma}} \quad (55)$$

when $0 \leq \gamma < 2$. When the constraint $n \leq N$ does not bind, the third best employment level will be an integer on either side of n_T^* . We assume

$$B\gamma > 2D^2\psi^2, \quad (56)$$

which implies $n_T^* > 1$, so the third best employment level is at least one after accounting for the integer constraint. To simplify the exposition, we ignore the integer constraint from now on. The third best employment level is therefore¹⁸

$$n^T = \begin{cases} \min\{n_T^*, N\} & \text{if } 0 < \gamma < 2 \\ N & \text{if } \gamma \geq 2. \end{cases} \quad (57)$$

We now return to the first best problem. The critical point (21) satisfies $n_F^* > 1$ under the condition (56). The first best employment level is therefore

$$n^F = \begin{cases} \min\{n_F^*, N\} & \text{if } 0 < \gamma < 1 \\ N & \text{if } \gamma \geq 1. \end{cases} \quad (58)$$

We can now compare the two institutions. We first note that if $\gamma \geq 2$, then $n^T = n^F = N$. Next, we consider the case where $0 < \gamma < 1$, so that $\frac{1}{1-\gamma} > 1$ and $\frac{1}{2-\gamma} < 1$. Since $D > e \approx 2.718$,

$$1 < \frac{B\gamma}{2D^2\psi^2} < \frac{B\gamma}{D\psi^2} \quad (59)$$

and $n_T^* < n_F^*$. If $N > n_T^*$, then $n^T < n^F$. In particular, if $n_T^* < N \leq n_F^*$, then the first best firm is limited by the extent of the market $n^F = N$, while the second best firm is limited by moral hazard alone. If $N \leq n_T^*$, then $n^T = n^F = N$. Finally, we consider the case where $1 \leq \gamma < 2$ so that $n^F = N$. If $n_T^* < N$, then $n^T < n^F$ but if $n_T^* \geq N$, then $n^T = n^F = N$. In all cases, the third best firm is at least weakly smaller than the first best firm.

6 | CONCLUSION

In this paper, we developed a stochastic production function that incorporates specialization and division of labor and the possibility that negative shocks or shirking can have major adverse consequences for the firm. The purpose of the paper was to show

that this O-ring production function has important implications not only for incentive contracting but also for the size of the firm in terms of employment and the extent of the division of labor.

Under the first best monitoring technology, there is no moral hazard problem because the principal can directly monitor the agents' choices. In this case, the O-ring production function has limited effects because the principal does not provide any incentives. Instead, the principal provides full insurance that implies a constant transfer regardless of output or employment. The firm therefore behaves like a perfectly competitive wage-taker despite the fact that it is a monopolist. In the absence of any constraints, the first best employment level would be infinite because of the increasing returns to employment due to specialization and division of labor. The size of the first best firm and the extent of the division of labor are bounded either by the extent of the labor market or the first O-ring effect, which states that the team probability that all shocks are positive converges to zero as employment increases without bound.

Under moral hazard, the O-ring production function affects incentive contracting in two ways. First, the optimal contract is a Mirrlees contract where agents are rewarded when output is high (individual or team output, depending on the monitoring technology) and fired otherwise. Second, the incentive and expected transfer depends on the probability of success (individual or team) that, in turn, depends on employment. The expected transfer then determines the cost of employment through which moral hazard and agency costs can affect the size of the firm and the extent of the division of labor.

In the second best case, the principal cannot monitor choices but she can monitor individual output. The relevant probability is therefore the individual probability of a positive shock, which is increasing in employment because fewer assigned tasks reduce the chance of errors. It follows that the second best incentive and expected transfer are *decreasing* in employment and we obtain the novel result that the second best marginal cost of employment is *less* than the expected transfer. In fact, the second best marginal cost of employment can be less than first best (and even negative), so the second best firm can be larger than first best. As a result, the second best firm tends to be large, with a high degree of specialization and division of labor, weak incentives, and low expected pay. As employment increases, in the limit the moral hazard problem vanishes and the second best incentive, expected transfer, and marginal cost of employment all converge to their first best levels. In our model, agency costs cannot bound the size of the firm or the extent of the division of labor when the principal can monitor individual output.

In the third best case, the principal can only monitor team output and the relevant probability is the team probability that all shocks are positive. Whereas the individual probability increases to one with employment, under the O-ring property, the team probability declines to zero. It follows that the third best incentive and expected transfer are *increasing* in employment, so the third best marginal cost of employment *exceeds* the expected transfer like a traditional monopsonist. As employment increases, the moral hazard problem, the third best incentive, expected transfer, and the marginal cost of employment all explode. As a result, the third best firm tends to be small (smaller than the other two institutions), with a low degree of specialization and division of labor, strong incentives, and high expected pay relative to the agents' outside option. In our model, agency costs can only bound the size of the firm and the extent of the division of labor when the principal can only monitor team output.

ENDNOTES

- ¹ Note that the agents, and not the tasks, are the source of the negative shocks. It is not clear why the number of probabilities would equal the number of agents if the tasks themselves were the sole source of the negative shocks. I thank an anonymous referee for remarks on this issue.
- ² Shah, Ball, and Netessine (2016) find that automobile recalls are positively associated with product variety, plant utilization, and the interaction between the two. The finding that recalls are increasing in product variety at the plant level is analogous to our assumption that the probability of a negative shock is increasing in task variety at the agent level.
- ³ I thank the referee for remarks on this issue.
- ⁴ We assume that τ_{is} is a measurable function on S_i such that $0 \leq \tau_{is} \leq 1$ for all i and $s \in S_i$.
- ⁵ Note that $\pi' < 0$ and $\pi(1) > 0$ imply that $\pi(\rho) > 0$ for all $\rho \in [0, 1]$.
- ⁶ We make the standard assumption that an indifferent agent chooses as the principal directs. If an agent does not take the actions specified in the contract, then he will be fired and receive his outside option (zero). Since the participation constraints bind (see the proof of Lemma 2 in the Appendix), the agent is indifferent and will therefore adhere to the contract as directed by the principal.
- ⁷ We have normalized the price of output to be 1. Otherwise it would appear in the expression for B .
- ⁸ This would be the case if agents had a positive instead of zero outside option.
- ⁹ A simple sufficient condition for the first best problem is $\pi(1)B > h(\psi)$ that ensures positive expected profit from hiring at least one agent.
- ¹⁰ See Laffont and Martimort (2002, p. 152).

- ¹¹ Since the principal chooses n from a finite set $\{0, 1, \dots, N\}$, a solution to the first best problem exists but we cannot ensure uniqueness even if Π^F is strictly quasi-concave in n . When the solution is unique and equal to n^0 , we write $n^F = n^0$ instead of $n^F(B, \gamma, \psi) = \{n^0\}$.
- ¹² For example, if the agent exerts efficient effort and the shock is negative, then his payoff will be $u(0) - \psi = -\psi$ because $u(0) = 0$. If he is fired, then he receives his outside option (zero) and his payoff is $0 - \psi = -\psi$.
- ¹³ See Bolton and Dewatripont (2005, p. 139).
- ¹⁴ For simplicity, we refer to C_n^I as the marginal cost of employment when, in fact,

$$C^I(n+1, \psi) - C^I(n, \psi) = \int_n^{n+1} C_n^I dn. \quad (60)$$

- ¹⁵ The fact that h is strictly convex does not imply that $h' \rightarrow \infty$. For example, $f(x) = x - \log(x+1)$ is increasing and strictly convex but its derivative converges to one as $x \rightarrow \infty$.
- ¹⁶ All we really need is that the team probability is eventually decreasing in employment. With the appropriate modifications, our results would hold if $\pi(\rho)^n$ was decreasing in n for all $n \geq n^0$ for some $n^0 \geq 1$.
- ¹⁷ In this section, ϵ_{π^n} is the elasticity of π^n with respect to n , whereas in the previous section, ϵ_π is the elasticity of π with respect to ρ . This explains why $\Delta^I = 1 - \epsilon_\pi(1 - \epsilon_h)$ (note the minus sign) whereas $\Delta^T = 1 + \epsilon_{\pi^n}(1 - \epsilon_h)$.
- ¹⁸ The condition (56) implies that $n^T = N$ when $\gamma = 2$ and $n^F = N$ when $\gamma = 1$.

REFERENCES

- Auriol, E., Friebel, G., & Pechlivanos, L. (1999). *Teamwork management in an era of diminishing commitment*. CEPR Discussion Paper no. 2281.
- Becker, G. S., & Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics*, 107, 1137–1160.
- Bolton, P., & Dewatripont, M. (2005) *Contract theory*. Cambridge: MIT Press.
- Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16), 386–405.
- Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, 13, 324–340.
- Holmström, B., & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55, 303–328.
- Kremer, M. (1993). The O-ring theory of economic development. *Quarterly Journal of Economics*, 108, 551–575.
- Laffont, J.-J., & Martimort, D. (2002). *The theory of incentives: The principal-agent model*. Princeton: Princeton University Press.
- Liang, P. J., Rajan, M. V., & Ray, K. (2008). Optimal team size and monitoring in organizations. *Accounting Review*, 83, 789–822.
- Milgrom, P., & Roberts, J. (1992). *Economics, organization, and management*. Englewood Cliffs, NJ: Prentice-Hall.
- Rauh, M. T. (2014). Incentives, wages, employment, and the division of labor in teams. *RAND Journal of Economics*, 45, 533–552.
- Shah, R., Ball, G., & Netessine, S. (2016). *Plant operations and product recalls in the automotive industry: An empirical investigation*. INSEAD Working Paper No. 2016/01/TOM.
- Smith, A. (1904). *An inquiry into the nature and causes of the wealth of nations*. London: Methuen & Co., Ltd.,
- Ziv, A. (1993). Performance measures and optimal organization. *Journal of Law, Economics, and Organization*, 9, 30–50.

How to cite this article: Rauh M. The O-ring theory of the firm. *J Econ Manage Strategy*. 2018;27:82–101. <https://doi.org/10.1111/jems.12216>

APPENDIX

Main Example. We first show that the Main Example satisfies Assumptions 1. Write $\pi(\rho) = e^{\rho \log(\rho/D)} > 0$. Using L'Hôpital's rule,

$$\rho \log \frac{\rho}{D} = \frac{\log \frac{\rho}{D}}{\frac{1}{\rho}} \rightarrow 0 \quad (A1)$$

as $\rho \rightarrow 0$ so π is continuous at $\rho = 0$. It is smooth (i.e., infinitely differentiable) on $(0, 1]$ with derivative

$$\pi' = \pi \left[1 + \log \left(\frac{\rho}{D} \right) \right] < 0 \quad (A2)$$

for all $D > e$. Assumption 3 holds because $\pi(\rho)^n = \frac{1}{D^n}$.

Proof of Lemma 1. Let $T_i = 1$. Given $\tau_{is} \geq 0$ for some task $s \in S_i$, the optimal allocation of τ_{is} between e_{is} and l_{is} is

$$e_{is} = \frac{\tau_{is}}{1 + \gamma} \quad \text{and} \quad l_{is} = \frac{\gamma \tau_{is}}{1 + \gamma}. \quad (\text{A3})$$

In that case,

$$q_{is}^+ = \frac{A}{\gamma} \left(\frac{\gamma \tau_{is}}{1 + \gamma} \right)^{1+\gamma} = B \tau_{is}^{1+\gamma}. \quad (\text{A4})$$

Let τ_{is} be an integrable function of $s \in S_i$ that satisfies $\tau_{is} \geq 0$ and (4) with $T_i = 1$. Let $I_i = \inf_{s \in S_i} \tau_{is}$. Suppose there exists a subset S'_i of S_i of positive measure such that $\tau_{is} > I_i$ for all $s \in S'_i$. In that case, the principal could adjust τ_{is} such that $\tau_{is} = I_i$ for all $s \in S'_i$ and then increase I_i slightly so as to preserve (4). This shows that an optimal assignment τ_{is} must be almost everywhere constant. It is therefore optimal to divide $T_i = 1$ evenly $\tau_{is} = 1/\rho$ across tasks. \square

Proof of Lemma 2. The principal's problem is to choose $\bar{t}_i \geq 0$ and $\underline{t}_i \geq 0$ to minimize

$$\pi(\rho)\bar{t}_i + [1 - \pi(\rho)]\underline{t}_i \quad (\text{A5})$$

subject to the participation constraint (16). Since u is not necessarily differentiable at $t_i = 0$, we first consider the case where the domain of the objective function and participation constraint is $(\bar{t}_i, \underline{t}_i) \in (0, \infty) \times (0, \infty)$. Note that the objective function is linear and the constraint is strictly concave, so the standard concave programming theorem applies. The Lagrangean and first-order conditions are

$$\mathcal{L} = -\pi\bar{t} - (1 - \pi)\underline{t} + \lambda [\pi u(\bar{t}) + (1 - \pi)u(\underline{t}) - \psi], \quad (\text{A6})$$

$$\mathcal{L}_{\bar{t}} = -\pi + \lambda \pi u'(\bar{t}) = 0, \quad (\text{A7})$$

$$\mathcal{L}_{\underline{t}} = -(1 - \pi) + \lambda(1 - \pi)u'(\underline{t}) = 0, \quad (\text{A8})$$

$$\mathcal{L}_{\lambda} = \pi u(\bar{t}) + (1 - \pi)u(\underline{t}) - \psi \geq 0, \quad (\text{A9})$$

where we omit the argument ρ in π , drop the i subscript, and omit the complementary slackness condition. From the first two first-order conditions, we observe that $\lambda > 0$ and $\bar{t} = \underline{t}$ because $u'' < 0$. Since the participation constraint binds, we have that $\bar{t} = \underline{t} = h(\psi)$. We now check the boundary. One potential solution is $\underline{t} = 0$ and $\bar{t} = h(\psi/\pi)$. We show that this entails higher expected cost in the proof of Lemma 4. Another potential solution is $\bar{t} = 0$ and $\underline{t} = h[\psi/(1 - \pi)]$ but

$$\pi(0) + (1 - \pi)h\left(\frac{\psi}{1 - \pi}\right) = \pi h(0) + (1 - \pi)h\left(\frac{\psi}{1 - \pi}\right) > h\left[\pi(0) + (1 - \pi)\frac{\psi}{1 - \pi}\right] = h(\psi) \quad (\text{A10})$$

since h is strictly convex. \square

Proof of Proposition 1. Write expected profit as

$$\Pi^F = n [\pi(\rho)^n B n^\gamma - h(\psi)]. \quad (\text{A11})$$

If $L < \frac{h(\psi)}{B}$ then $\lim_{n \rightarrow \infty} \Pi^F = -\infty$ and there exists an \hat{N} such that $\Pi^F < 0$ for all $n \geq \hat{N}$. If $\hat{N} < N$, then $n < \hat{N} < N$ for all $n \in n^F(B, \gamma, \psi)$. If $L > \frac{h(\psi)}{B}$, then $\lim_{n \rightarrow \infty} \Pi^F = \infty$ and a similar argument shows that the constraint $n \leq N$ binds when N is sufficiently large. \square

Proof of Lemma 3. The principal's problem is

$$\min_{t_i(q_i)} \pi(\rho)t_i(q^h) + [1 - \pi(\rho)]t_i(0) \quad (\text{A12})$$

subject to $t_i(q_i) \geq 0$ for all $0 \leq q_i \leq q^h$, (25), (27), and (28). We first consider the problem neglecting the constraints (28). Since $u(t_i) \geq 0$ for all $t_i \geq 0$, (27) implies (25). We can therefore drop the latter constraint because it is redundant. Since (27) reduces to

$$u(t_i(q^h)) - u(t_i(0)) \geq \frac{\psi}{\pi(\rho)}, \quad (\text{A13})$$

the solution is clear. The omitted constraints are then satisfied by $t_i(\hat{q}_i) = 0$ for all $0 \leq \hat{q}_i < q^h$. \square

Proof of Lemma 4. Since the participation constraint (25) binds and u is strictly concave,

$$\psi = \pi(\rho)u(t_i(q^h)) + [1 - \pi(\rho)]u(t_i(0)) < u[\pi(\rho)t_i(q^h) + [1 - \pi(\rho)]t_i(0)]. \quad (\text{A14})$$

Since h is increasing,

$$h(\psi) < \pi(\rho)t_i(q^h) + [1 - \pi(\rho)]t_i(0) = \pi(\rho)h\left[\frac{\psi}{\pi(\rho)}\right] = H^I(\rho, \psi). \quad (\text{A15})$$

Assumption 2 implies that h is twice continuously differentiable on $(0, \infty)$ with $h' > 0$ and $h'' > 0$. Since h is strictly convex,

$$h'(u_1)(u_2 - u_1) < h(u_2) - h(u_1) \quad (\text{A16})$$

for all $u_1 \neq u_2$. Setting $u_1 = \psi/\pi$ and $u_2 = 0$,

$$h(\psi/\pi) - \frac{\psi}{\pi}h'(\psi/\pi) < 0 \quad \text{and} \quad 1 - \frac{\psi}{\pi} \frac{h'(\psi/\pi)}{h(\psi/\pi)} < 0. \quad (\text{A17})$$

Differentiating $H^I(\rho, \psi)$ with respect to ρ ,

$$H^I_\rho = H^I_\pi \pi' = \left(h - \frac{\psi}{\pi}h'\right)\pi' > 0, \quad (\text{A18})$$

which completes the proof. \square

Proof of Proposition 2. Write expected profit as $\Pi^I = n[\pi(\rho)^n B n^\gamma - H^I(n, \psi)]$. Since $H^I(n, \psi) \rightarrow h(\psi)$ as $n \rightarrow \infty$ for all $\psi > 0$, the proof is identical to that for Proposition 1. \square

Proof of Lemma 5. Differentiating C^I with respect to n ,

$$C^I_n = nH^I_\rho \left(-\frac{1}{n^2}\right) + H^I = H^I - \frac{1}{n}H^I_\rho = \pi h - \frac{1}{n} \left(h - \frac{\psi}{\pi}h'\right)\pi' \quad (\text{A19})$$

$$= \pi h \left[1 - \frac{1}{n} \frac{\pi'}{\pi} \left(1 - \frac{\psi}{\pi} \frac{h'}{h}\right)\right] = H^I \Delta^I. \quad (\text{A20})$$

In the proof of Lemma 4, we showed that $1 - \epsilon_h < 0$ so $\Delta^I < 1$. \square

Proof of Proposition 3. Let $\bar{n}_+^I = \max_n \mathbb{N}_+^I$ and $\bar{n}^F = \max_n n^F$, where we omit the arguments B , γ , and ψ . If $\bar{n}_+^I \leq \bar{n}^F$, the conclusion follows. Assume $0 \leq \bar{n}^F < \bar{n}_+^I$ and suppose the second best firm chooses $0 \leq \bar{n}^F < n \leq \bar{n}_+^I$, where $n = \bar{n}^F + k$ for some positive integer k . This increase in employment increases C^F by $kh(\psi)$ and reduces Π^F . Since the effect on expected revenue is the same for both institutions and

$$C^I(n, \psi) - C^I(\bar{n}^F, \psi) = \int_{\bar{n}^F}^n C^I_n(n, \psi) dn \geq \int_{\bar{n}^F}^n h(\psi) dn = kh(\psi), \quad (\text{A21})$$

the increase in employment also reduces Π^I . \square

Proof of Lemma 7. Let $p(n) = \pi(\rho)^n$. Statement (i) follows from the derivative

$$\frac{dH^T(n, \psi)}{dn} = p' \left(h - \frac{\psi}{p} h' \right) > 0. \tag{A22}$$

We now prove (ii). Since $p(n) \rightarrow 0$ and h is increasing and strictly convex, $h \rightarrow \infty$ as $n \rightarrow \infty$ and

$$\lim_{n \rightarrow \infty} ph(\psi/p) = \lim_{n \rightarrow \infty} \frac{h(\psi/p)}{\frac{1}{p}} = \lim_{n \rightarrow \infty} \psi h'(\psi/p) = \infty \tag{A23}$$

by L'Hôpital's rule and Assumptions 3. □

Proof of Lemma 8. Differentiating C^T with respect to n ,

$$C_n^T = nH_n^T + H^T = np' \left(h - \frac{\psi}{p} h' \right) + ph \tag{A24}$$

$$= ph \left[1 + \frac{np'}{p} \left(1 - \frac{\psi}{p} \frac{h'}{h} \right) \right] = H^T [1 + \epsilon_p(1 - \epsilon_h)], \tag{A25}$$

which proves (48). Since $1 - \epsilon_h < 0$ and $p' < 0$ and $\epsilon_p < 0$ for all $n \geq 1$, we have $\Delta^T > 1$ for all $n \geq 1$ which proves (ii). Since $H^T \rightarrow \infty$ as $n \rightarrow \infty$, this proves (iii). □

Proof of Proposition 4. The proof is similar to that for Proposition 3. □