INCENTIVES AND LOSS OF CONTROL IN AN OPTIMAL HIERARCHY

by

Yingyi Qian*

Department of Economics
Stanford University
Stanford, CA 94305-6072
U.S.A.

Review of Economic Studies, 61(3):527-544, July, 1994.

---

* I am most grateful to Andreu Mas-Colell and Eric Maskin for their advice and encouragement. I also wish to thank Oliver Hart, Martin Hellwig, Michael Keren, Janos Kornai, David Levhari, Paul Milgrom, Kathryn Spier, Jean Tirole, Martin Weitzman, Chenggang Xu, three anonymous referees, and seminar participants at Harvard, MIT, Stanford, Chicago, Yale, Princeton, Columbia, Michigan State, University of Pennsylvania and University of Michigan for comments on the earlier version of the paper. This is a part of Essay I of my Ph.D. dissertation entitled *Incentives and Control in Socialist Economies* submitted to Harvard University.

Abstract

This paper studies incentives and loss of control in a hierarchy model which combines and generalizes the models of Williamson, Calvo-Wellisz and Keren-Levhari. In our model of the hierarchy, the levels of effort from managers and workers, the wage scales, the span of control and, in particular, the total number of tiers are all endogenous. Using optimal control techniques, we show that in the optimal hierarchy the wage scales and effort levels decrease as one moves down the hierarchy. As the hierarchy expands with no technological progress, workers exert less effort and are paid less, top managers work harder and are paid more and the wage distribution becomes increasingly skewed.

1. Introduction

Hierarchy is a typical organizational form of business firms, not-for-profit organizations, governments, as well as centrally-planned economies. Hierarchical organizations differ from the market. In the market, buyers and sellers trade with each other on a more-or-less equal footing. In a hierarchy, employees at different hierarchical levels are vertically related: the superior has authority over his subordinates and the subordinate provided information to, and followed instructions from, his superiors.

We study a model of hierarchy which combines and generalizes the models of Williamson (1967), Calvo and Wellisz (1978, 1979), and Keren and Levhari (1979). Among many important economic problems of hierarchical organizations, the paper focuses on the one of incentives.[1] In the hierarchy that we study, subordinates may shirk on the job or divert their effort to their own interests when effort is not observed by their superiors. To mitigate such a problem, the superior spends time in monitoring the effort exerted by his immediate subordinates. The final output of the hierarchy is determined by a production function which is cumulative in the efforts of workers and managers at all levels.

The organizational design problem is to determine the number of hierarchical tiers, the span of control (i.e., the number of subordinates under the same superior) and the wage scales in the hierarchy, taking as given the amount of capital and the state of technology. The economic tradeoffs are rather complex, but the basic idea is the following: The benefit of having fewer hierarchical tiers is that there is a smaller cumulative loss across hierarchical levels and also fewer managers to pay. The cost is that remaining employees have to be paid with higher efficiency wages to induce them to work, because the effectiveness of monitoring is reduced as the result of the increased span of control. The idea echoes Simon (1976)'s analysis of the tradeoff between reducing the span of control and reducing the total number of tiers

---

[1] The literature on hierarchy is divided into at least two distinct fields: "team theory" and "incentive theory." Studies are further differentiated by the approach of the "grand contract," which includes the above cited papers and this paper, and the approach of the "network of contracts," which includes Tirole (1986), Melumad, Mookherjee and Reichelstein (1989) and McAfee and McMillan (1990).

of the hierarchy.

We solve the problem of optimal hierarchy for the objective of maximizing net revenue by using optimal control techniques, a method pioneered by Keren and Levhari (1979).[2] The optimal control techniques have proven to be quite useful, which enables us to obtain many insightful results about the hierarchical structure. Two main results are obtained with a specific monitoring and production technology. First, in the optimal hierarchy in which all managers and workers are identical *ex ante*, wages fall and efforts decrease as one moves from the top to the bottom of the hierarchy. If we interpret loss of control as a fall in the amount of effort exerted by managers and workers, then our result shows that loss of control increases down the hierarchy, and suffers most at the bottom. Second, as the size of the hierarchy increases, both efforts and wages of managers at the top increase because their marginal product increases, and both efforts and wages of workers fall because their marginal product declines. Hence, the wage ratio between the top managers and workers increases. This result implies a greater loss of control for a bigger hierarchy.[3]

We also analyze a simpler model in which effort choice is restricted to only two values, 0 and 1. In such a model we find that, to implement effort level 1 without cumulative losses through tiers in the production function, the hierarchy exhibits constant wage and constant span of control across tiers. Furthermore, an increase in the capital stock or an increase in the productivity of workers leaves the structure of the hierarchy unchanged except for the total number of tiers.

The paper is organized as follows. Section 2 introduces the model. Section 3 analyzes the simple model with only 0 and 1 effort choice and finds the closed form solution. Section 4 examines optimal wage

---

[2] Our model is technically akin to those of Keren and Levhari (1979, 1983, 1989), although their models of hierarchy are not based on incentive considerations.

[3] Throughout the paper, we have assumed away the possibility of increasing returns. However, we want to stress the point that the problem of excessive bigness does not go away, even if we had allowed for increasing returns.

scales and effort levels within a hierarchy. Comparative statics results are derived in Section 5. The final section 6 discusses applications and extensions of the model. Appendix A derives the solution to the discrete version of the model with 0 and 1 effort choices and Appendix B contains mathematical proofs.

2. The Model

Consider an economic organization that owns a capital stock K and uses a hierarchy to control the production. We can imagine that at the very bottom of the hierarchy are workers operating machines or working on the assembly lines; they are grouped into workshops. Several workshops are organized into a factory, and several factories into a firm. To simplify the analysis, we assume that everyone in the hierarchy has only one superior. Hence, the hierarchy can be represented by a tree, and the owner of the organization is represented by the root of the tree. Tiers of the hierarchy are denoted by subscript t when counted from the top to the bottom and the number of employees in tier t is denoted by x_t with $x_0=1$. The number of employees in tier t who are subordinates of a common superior is called the span of control in tier t-1, and is denoted by s_t. The span of control is assumed to be constant in any tier, which implies that $x_t=x_{t-1}s_t$. We assume that the organization maximizes the revenue from the capital stock net of payments to all employees (i.e., profit). The objective of the managers/bureaucrats and workers is to maximize their utilities, w-g(a), where w is the wage paid by the organization, a is the effort devoted to the hierarchy, and g(•) is the disutility from making such an effort.

Production Technology. We assume a fixed capital-worker ratio and denote it by k. By doing so, we assume away the problem of substitution between capital and workers and make the model tractable. For the given capital stock K, the inelastic demand for workers is equal to N=K/k, N>1.[4] The effort of employees in tier t is denoted by a_t, which is constant in any tier, but may differ across tiers. We employ a

---

[4] Implicitly we assume that the labor supply constraint is not binding.

recursive production technology similar to the one used by Williamson (1967), Beckmann (1977) and Rosen (1982), in which an intermediate product called "managerial effectiveness" is produced. In any tier t, the intermediate product y_{t-1} from the immediate superior is used as an input and combined with effort a_t to produce y_t for the immediate subordinate. This process can be written as a function $y_t=F_t(y_{t-1},a_t)$. For simplicity, a simple functional form of F_t is assumed so that $y_t=y_{t-1}a_t$, where $0 \leq a_t \leq 1$ and $a_t=1$ is the maximum effort attainable. The initial input $y_0=1$ is assumed, so $y_t=a_t a_{t-1} \bullet\bullet\bullet a_1$.[5] The output of each fully effective worker (that is, when $y_T=1$) is assumed to be $\theta>0$, the gross output of the hierarchy is therefore $\theta N y_T$, which is also the gross revenue as we normalize the price equal to 1.

Monitoring and Rewards. The managers in the hierarchy also play supervisory roles: collecting information and monitoring the inputs and outputs of their subordinates. We follow Calvo-Wellisz (1978, 1979) in modelling these activities. The superior monitors the efforts of his subordinates directly, and this monitoring requires only time and no effort. When the subordinate is checked by his superior, his effort is known precisely. When the subordinate is not checked, the effort is not known at all to the superior. Because the superior has only limited time available, he can only check his subordinates with some probability $P<1$. In general, P is a decreasing function of the span of control: the more subordinates the superior monitors, the smaller the probability of the subordinate being checked. For simplicity, we shall assume that $P=1/s$, which corresponds to the case when the superior has to spend all of his time checking one of his subordinates.[6]

---

[5] Williamson (1967) assumes that $y_t=\bar{a}^t$, where $0<\bar{a}<1$ is a constant. The production function used here is a natural generalization of Williamson's. One interpretation of the production function is plan implementation. Managers at the top first formulate a strategic plan which they want to implement. During the process of disaggregation, managers in each tier add details into the plan by using their own information and knowledge before they instruct the next tier. After many tiers of disaggregation and manipulation, the final detailed plan is implemented by the workers. Therefore, effort devoted by each level affects the final plan and, hence, the product of the entire hierarchy.

[6] Simon (1976) recognized that "the scarce resource is not information, it is the processing capacity to attend to information." In our model, the scarce resource is time.

Suppose the wages can't be negative, which represents the limited liability constraint.[7] To implement effort level a^* from his subordinates, the superior considers the following family of incentive schemes:[8]

pay w if $a \geq a^*$ is known, or if a is not known; and

pay 0 if $a<a^*$ is known.

The incentive compatibility condition requires

$$w-g(a^*) \geq P \bullet 0+(1-P) \bullet w-g(a), \text{ for all } a<a^*,$$

which yields the "efficiency wage" $w=g(a^*)/P$. With $P=1/s$, the wage function is therefore given by

$$w_t=g(a_t)s_t.$$

Hence, the employee's welfare is $u=g(a_t)s_t-g(a_t)$. When the reservation utility of the employee is 0, u also represents an economic rent (due to imperfect monitoring) earned by the employee.

The optimization problem of the organization can be summarized as

$$(2.1) \qquad \max_{s_t,\, a_t,\, T} \quad \theta N y_T - \Sigma_1^T \{g(a_t)s_t x_t\}$$

$$\text{s.t.} \quad x_t=x_{t-1}s_t$$

---

[7] Given the quasi-linear utility function $u=w-g(a)$, without the limited liability constraint the "first best" can be achieved even though monitoring is imperfect ($P<1$), as long as P is not zero. Limited liability is also a realistic assumption which reflects limited resources that agents possess.

[8] In order for such a scheme to be optimal, we must rule out the following scheme which always implements the first best effort level efficiently:
pay w if $a \geq a^*$ is known;
pay 0 if $a<a^*$ is known, or if a is not known.
This is because the incentive compatibility constraint becomes
$$pw - g(a^*) \geq 0 - g(a),$$
for any $a<a^*$. That is, $pw=g(a^*)$. Since w is paid only with probability p, and the individual rationality constraint is binding: $pw-g(a^*)=0$. However, the above scheme is applicable only when the workers can prove they are monitored, otherwise the superior has an incentive to claim he did not monitor and always pays 0.

$$y_t=y_{t-1}a_t$$

$$x_0=1,\ x_T=N, \text{ and } y_0=1.$$

For the given N and θ, the organization faces the following tradeoffs in determining the hierarchy: Reducing the total number of tiers, T, decreases cumulative losses across hierarchical levels (because in general $a_t<1$) and therefore increases the total revenue. But the span of control has to be increased, and therefore the probability that a subordinate is checked is reduced so that the wage must be increased to satisfy the incentive compatibility constraint. This is costly. On the other hand, reducing the span of control would increase the effectiveness of monitoring and hence reduce the wages, but the total number of tiers must be increased, which increases cumulative losses and reduces the gross revenue. Furthermore, the organization wants higher efforts a_t to increase the total revenue Ny_T, but it has to pay more for higher a_t.[9]

Although the discrete formulation of the problem can be solved completely in a simple case (see Appendix A), it is rather difficult to tackle the problem in general. Following Keren and Levhari (1979), we employ the following continuous approximation. Taking logarithm on both sides of

$$x_t=x_{t-1}s_t=s_ts_{t-1} \cdots s_1x_0$$

yields

$$\log(x_t)=\Sigma_0^t\log(s_\iota),$$

which has a continuous analog

$$\log(x_t)=\int_0^t\log(s_\iota)d\iota,$$

or

---

[9] In our model, the owner of the organization can't achieve the "first best" by stopping monitoring and asking each agent to pay a fixed rent p because such a contract is not enforceable. Since the wages can't be negative as we have assumed so far, $w=\theta a-p\geq 0$, or $p\leq\theta a$. Furthermore, because a is not observable without monitoring, neither is output θa. Then the agent always has an incentive to claim that output is low and thus refuse to pay p.

$$\dot{x}_t = x_t \log(s_t).$$

Similarly,

$$\dot{y}_t = y_t \log(a_t).$$

This gives the general problem in continuous form as

$$(2.2) \qquad \max_{s_t,\, a_t,\, T} \quad \theta N y_T - \int_0^T \{g(a_t) s_t x_t\} dt$$

$$\text{s.t.} \quad \dot{x}_t = x_t \log(s_t)$$

$$\dot{y}_t = y_t \log(a_t)$$

$$x_0 = 1,\ x_T = N, \text{ and } y_0 = 1.$$

3. The Case of 0 and 1 Effort Choices

We start our analysis with a simpler model, in which the choice of effort level is only between the minimum effort ($a=0$) and the maximum effort ($a=1$). The study of such a simple model not only enables us to derive a closed form solution, which itself is interesting, but also provides us with a reference model to be compared later with the general one.

Because $y_t = a_t a_{t-1} \bullet\bullet\bullet a_1$ and a_t is either 0 or 1, it does not pay for the organization to implement $(a_1, \bullet\bullet\bullet, a_T)$ in which some of a's are 0 and others are 1. Interesting results can be derived when the organization implements $(1, \bullet\bullet\bullet, 1)$. In such a case, $y_T = 1$, no cumulative loss in total managerial effectiveness. Write $g(1) = g$, the disutility of working. The maximization problem can be written as[10]

---

[10] The solution to the discrete formulation is contained in Appendix A, which has qualitatively the same properties as in Propositions 1 and 2.

(3.1) $$\max_{s_t,\, T} \quad \theta N - g \int_0^T \{s_t x_t\} dt$$

$$\text{s.t.} \quad \dot{x}_t = x_t \log(s_t)$$

$$x_0 = 1 \text{ and } x_T = N.$$

Technically, this is an "open-final-time" and "fixed-end-point" optimal control problem with one state variable (x_t) and one control variable (s_t). The Hamiltonian is

$$H(t) = -gs_t x_t + p_t x_t \log(s_t),$$

where p_t is the multiplier for equation $\dot{x}_t = x_t \log(s_t)$. The first order condition yields

$$\dot{p}_t = gs_t - p_t \log(s_t), \text{ and}$$

$$-gx_t + p_t x_t / s_t = 0.$$

We derive from the above equations

(3.2) $$p_t = gs_t, \text{ and}$$

(3.3) $$\dot{p}_t / p_t = 1 - \log(s_t).$$

Combining (3.2) and (3.3), we obtain

$$\dot{s}_t / s_t = 1 - \log(s_t), \text{ or}$$

$$(\log(s_t))' + \log(s_t) = 1.$$

The solution to this first order differential equation is

(3.4) $$\log(s_t / e) = Ce^{-t},$$

where C is a constant to be determined by the boundary conditions.

In a t-independent autonomous problem such as the present one, the transversality conditions for an "open-final-time" problem is

(3.5) $H(t) = 0.$

Combining (3.2), (3.4) and (3.5) yields

$$C=0.$$

Finally, using the boundary conditions $x_0=1$ and $x_T=N$, we obtain

Proposition 1. Assume workers are sufficiently productive: $\theta>ge$.[11] Then the optimal hierarchy has the following properties:

(i) the span of control is constant: $s_t=e$;

(ii) the wage scale is also constant: $w_t=p_t=ge$;

(iii) the number of employees in tier t is $x_t=e^t$;

(iv) the total number of tiers of the hierarchy is $T=\log N$, and the marginal increase of the total number of tiers decreases in total capital ($dT/dN=1/N$); and

(v) total net revenue, or the total return to capital, is given by $V(N,\theta)=N(\theta-ge)+ge$. Therefore, the marginal return to capital is constant ($dV/dN=\theta-ge$), so is the marginal increase as the productivity parameter θ increases ($dV/d\theta=N$).

In our model, the endogenous wage scale is completely egalitarian, there is no need to pay more to higher rank managers than to lower ones. This is in contrast with the model of Calvo and Wellisz (1979) (a simplified version is given by Holmstrom and Tirole, 1988), which obtained the result of falling wage down the hierarchy. However, in Calvo and Wellisz, there is a "bottleneck at the top" resulting from the assumption of fixed total number of tiers. In such a case, the constraint imposed by this bottleneck can be

---

[11] If $\theta<ge$, the organization would want to implement $a_t\equiv 0$. Note that, when effort choice is restricted to 0 and 1, efficiency requires that $a_t\equiv 1$ if $\theta>g$ and $a_t\equiv 0$ if $\theta<g$. Inefficiency only arises if $g<\theta<ge$, when the organization does not want to implement $a_t\equiv 1$ but it is efficient to do so.

relaxed to some extent by lowering monitoring time, or, increasing the span of control, at the top levels, while at the same time adjusting wages upward to maintain incentives. Their result can be reproduced in our model as follows.

Proposition 2. Define $c=\theta/(ge)$, and assume workers are sufficiently productive: $c>1$. Then when T is fixed and N is allowed to vary, the hierarchy will be adjusted to the point so that:

(i) the span of control decreases: $s_t=(e)c^{e^{(T-t)}}$;

(ii) the wage also decreases: $w_t=(ge)c^{e^{(T-t)}}$;

(iii) the number of employees in tier t is $x_t=(e^t)c^{e^T(1-e^{-t})}$; and

(iv) the maximum number of workers absorbed is $N=(e^T)c^{e^T-1}$.

Proof: This is the same optimal control problem as above but with a "fixed-final-time" and "open-end-point" so that the transversality condition becomes

$$p_T=\theta. \tag{3.6}$$

Combining (3.6) with equation (3.2) yields

$$s_T=\theta/g.$$

Substituting this into equation (3.4) to determine constant C:

$$C=e^T\log(\theta/ge).$$

The rest of the proof is routine.■

Indeed, when T is assumed fixed, the result of decreasing wages depends entirely on the decreasing span of control (since $w_t=gs_t$), which is in turn a response to the "bottleneck at the top tiers." But as soon as T is made a choice variable, the "bottleneck at the top" can be relaxed in a more efficient way by increasing the number of tiers instead of lowering monitoring time and raising wages at the top levels.[12] To restore the decreasing wage result under endogenous number of tiers, we need to turn to our general model.

4. Wages and Efforts in an Optimal Hierarchy

In the general model, the choice of effort level is continuous between 0 and 1. Hence the managerial effectiveness y_T is in general between 0 and 1 ($0<y_T<1$), which makes the simple model in the previous section only a boundary case. The following assumption will be maintained throughout the paper to insure an interior solution:

Assumptions. (i) $g(0)=0$ and $g(a)\rightarrow+\infty$ as $a\rightarrow1$;[13]

(ii) g is differentiable and is strictly increasing and strictly convex; and

(iii) ξ is differentiable and is strictly and monotonically increasing, where $\xi(a)=ag'(a)/g(a)$ is the elasticity of disutility.[14]

We use the Maximum Principle to solve the general problem (2.2). The Hamiltonian is

$$H(t)=-g(a_t)s_tx_t+p_tx_t\log(s_t)+q_ty_t\log(a_t),$$

where p_t and q_t are multipliers for the corresponding equations of motion. The transversality condition for the t-independent autonomous problem with "open-final-time" requires

$$H(t)=0.$$

---

[12] Propositions 1 and 2 show distinct structures of the optimal hierarchy by taking different procedures in the maximization. Note that the value of the objective has no limit when both T and N go to infinity.

[13] If g is bounded at a=1, the boundary solution a=1 is possible for large θ, and we are back to Section 3. On the other hand, if $a>1$ is allowed, it can be shown that the solution will explode.

[14] Since $g\rightarrow\infty$ as $a\rightarrow1$, it can be shown that $\xi\rightarrow\infty$ as $a\rightarrow1$ as well. Hence, it is impossible to have a constant or a monotonically decreasing ξ.

This yields the following set of first order conditions

(4.1) $\dot{p}_t/p_t=1-\log(s_t)$

(4.2) $\dot{x}_t/x_t=\log(s_t)$

(4.3) $\dot{q}_t/q_t=-\log(a_t)$

(4.4) $\dot{y}_t/y_t=\log(a_t)$

(4.5) $p_t=g(a_t)s_t$

(4.6) $\xi(a_t)p_tx_t=q_ty_t$

(4.7) $1-\log(s_t)=\xi(a_t)\log(a_t)$

as well as the boundary conditions

$$x_0=1,\ x_T=N,\ \theta N=q_T \text{ and } y_0=1.$$

Proposition 3. The optimal hierarchical organization has the following properties:

(i) wages decrease in t: $\dot{w}_t<0$;

(ii) effort levels also decrease in t: $\dot{a}_t<0$; and

(iii) the span of control is always greater than e: $s_t>e$.

Proof: Using equations (4.1) to (4.7) and the wage equation $w_t=g(a_t)s_t$.

The decreasing wage means that, within the hierarchy, the superior is paid more than his subordinates, despite the fact that they are identical in ability and in preferences. Therefore, position counts in the hierarchy. By introducing a "loss-of-control" feature across hierarchical levels à la Williamson (1967) into the Calvo and Wellisz model, we are able to restore a Calvo and Wellisz type result about the decreasing wage schedule even though the total number of tiers is endogenous. But unlike Williamson, who assumes that the loss of control in each tier, 1-a, is constant and bounded away from 0, in our model, the loss of control can be reduced at each tier by exerting more effort a_t, which is endogenous. The decreasing effort means that under the optimal wage scales, the higher tier managers exert more effort and lower tier managers less effort, and the loss of control in the hierarchy is less in the upper tiers of the hierarchy, more in the lower tiers, and mostly at the bottom. The results of the falling wage and the falling effort level can be explained intuitively. The pyramidal structure of hierarchy implies fewer employees in the upper tiers than in the lower tiers. Given that $y_T=a_0\bullet\bullet\bullet a_T$, $0<a_t<1$, for the same level of aggregate effort y_T, implementing a higher effort from and paying a higher wage to the upper level manager is cheaper than implementing a higher effort from and paying a higher wage to the lower level managers. It follows that higher tier managers exert more effort and earn higher wages.

In comparison to the clean results of falling wages and falling effort, the properties of the span of control are ambiguous. Using the first order conditions, we obtain

$$\dot{s}_t/s_t=(\dot{w}_t/w_t)-\xi(a_t)\bullet(\dot{a}_t/a_t). \tag{4.8}$$

The first term on the right hand side is always negative, and the second term is always positive, leaving the sign of $\dot{s}_t$ undetermined.[15] Because $u_t=w_t-g(a_t)=g(a_t)(s_t-1)$ we derive

---

[15] For the following classes of disutility functions, the span of control increases in all tiers down the hierarchy: (i) $g(a)=[a/(1-a)]^\lambda$, where $\lambda\geq 1$; (ii) $g(a)=a^\lambda/(1-a)$, where $\lambda\geq 1$; and (iii) $g(a)=\log(1/(1-a))$. The span of control decreases in all tiers for the following g(a):

$$g(a) = \lambda \exp\left\{ \int_\epsilon^a (1/t)\, e^{\int_0^t (G(\tau)/\tau)d\tau} dt\right\},$$

where $0<\epsilon<1$ and

$$G(a) = \begin{cases} -\log(4)/(\log(a)) & \text{if } a<1/2; \\ 1/(1-a) & \text{if } a\geq 1/2. \end{cases}$$

Proofs are contained in Qian (1990).

$$\dot{u}_t = g'(a_t)(s_t - 1)\dot{a}_t + g(a_t)\dot{s}_t. \tag{4.9}$$

Therefore, falling wage in the hierarchy does not always imply falling welfare. The utility decreases down the hierarchy if the span of control does not increase too fast.

5. Comparative Statics

This section investigates effects of changes in N and θ on all endogenous variables of the hierarchy. The difficulty involved in such a study, both mathematically and conceptually, is the fact that as N or θ changes the optimal structure of the hierarchy changes, including the total number of tiers T. Allowing T to be a choice variable is important in the study of hierarchies because if we had allowed T to be fixed, the constraint on T would have introduced a bottleneck at the top tiers, an element of decreasing returns.

Conceptually, because T will change in response to changes in N or θ, we need to draw distinctions between someone who maintains the relative position in the hierarchy with respect to the top and that with respect to the bottom. Let tiers of the hierarchy be denoted by (superscript) τ when counted from the bottom to the top so that $T=t+\tau$. When T changes, fixing t will result in a change of τ, and fixing τ will result in a change of t. Suppose the capital stock (as well as the number of workers) is doubled, and suppose, as a result, one more tier is added to the hierarchy. To study the effects on wages, effort and the span of control of the person in tier 2, we must consider the following two cases separately. In the first case, the person remains in a position of tier 2 relative to the bottom workers ($\tau=2$): A foreman remains a foreman in a larger hierarchy. The foreman is not promoted from the workers' point of view; but he is "demoted" from the point of view of the top of the hierarchy. In the second case, the person remains in a position of tier 2 relative to the top ($t=2$): A vice president remains a vice president in a larger hierarchy. Although the vice president is not demoted in a larger hierarchy, he is "promoted" from the point of view of workers.

Mathematically, the dependency of the endogenous variables on the optimal T in the solution makes the analysis of comparative statics difficult. However, it is a nice property of the model that the endogenous variables depend on N, θ and T only through $c = \theta y_T/w_T$, the ratio of the marginal product of the worker to his marginal wage.

Lemma 1. Let a^τ, w^τ, and s^τ, $0 \leq \tau \leq T$, be the solution to the optimal hierarchy. Then,

(i) $a^\tau = h(ce^\tau)$;

(ii) $s^\tau = \exp\{1-\xi(a^\tau)\log(a^\tau)\}$; and

(iii) $w^\tau = g(a^\tau)s^\tau$,

where $h=\xi^{-1}$. Furthermore, $c>1$, that is, the marginal product of the worker is greater than the marginal wage paid to the worker.

Proof: See Appendix B.

We may write $a^\tau(c)$, $s^\tau(c)$ and $w^\tau(c)$ for all τ such that $0 \leq \tau < T$. According to Lemma 1, $c(N,\theta)$ provides the organization with sufficient information for determining the wage w^τ, the span of control s^τ and the effort a^τ for any tier τ. Also because of Lemma 1, the comparative statics exercises can be broken into two parts and the chain rule of differentiation can be applied: The first is comparative statics about c, and the second is that about the endogenous variables.

Lemma 2. For any c and τ, $0 \leq \tau < T$:

(i) $da^\tau/dc = a^\tau/(c\Phi(a^\tau)) > 0$;

(ii) $ds^\tau/dc = -s^\tau\xi(a^\tau)[\log(a^\tau) + 1/\Phi(a^\tau)]/c = -\dot{s}^\tau/c$; and

(iii) $dw^\tau/dc = -w^\tau\xi(a^\tau)\log(a^\tau)/c > 0$.

Proof: Differentiation and using $\dot{s}_t/s_t=\xi(a_t)\bullet[\log(a_t)+1/\Phi(a_t)]$.■

For <u>any c and N</u>, we define T(c,N) such that

$$(5.1) \qquad \int_0^{T(c,N)} \log(s^\tau(c))\, d\tau \equiv \log N,$$

and $y_{T(c,N)}$ such that

$$(5.2) \qquad \log y_{T(c,N)} \equiv \int_0^{T(c,N)} \log(a^\tau(c))\, d\tau.$$

Note that for $c=\theta y_T/w_T$, $T(c,N)=T$ and $y_{T(c,N)}=y_T$. T(c,N) and $y_{T(c,N)}$ have the following properties:

Lemma 3. For any c,

(i) $\partial T(c,N)/\partial N = 1/(N\log(s_0)) > 0$;

(ii) $\partial y_{T(c,N)}/\partial N = (y_{T(c,N)}\log(a_0))\bullet\partial T(c,N)/\partial N < 0$; and

(iii) $\partial T(c,N)/\partial c = \{\log(s_T)/\log(s_0) - 1\}/c$.

Proof: Differentiation of (5.1) and (5.2) and using Lemma 2(ii).■

Using these properties, we are able to prove the following crucial lemma:

Lemma 4. For $c=\theta y_T/w_T$, $dc/dN < 0$ and $dc/d\theta > 0$.

Proof: See Appendix B.■

5.1. Wages and Efforts

We first study the changes of wages and efforts for a particular person in the hierarchy in response to an increase of capital stock while keeping productivity θ fixed. Propositions 4 and 5 below show that results from the case of "counting from the bottom" (using a superscript τ) and the case of "counting from the top" (using a subscript t) are in sharp contrast.

Proposition 4. For any person remaining in tier τ, $0 \leq \tau < T$, when the capital stock increases, the optimal hierarchy will be adjusted so that:

(i) his effort level decreases ($da^{\tau}/dN<0$); and

(ii) his wage payment also decreases ($dw^{\tau}/dN<0$).

The results are true in particular for the workers ($\tau=0$): $da_T/dN<0$ and $dw_T/dN<0$.

Proof: Using the chain rule of differentiation and applying Lemma 2 and Lemma 4.■

Intuitively, the marginal product of a worker (θy_T) falls as a result of an increase in N (to be proved in Proposition 7), therefore, the worker's effort as well as wage fall too. Proposition 4 actually applies to all people who occupy those positions that are frequently referred to the relative distance to the bottom of the hierarchy. For examples, workers, foremen, project managers, etc. If an organization expands without any productivity improvement, anyone who remains in the same position relative to the bottom without promotion receives a lower wage than before.

Proposition 5. For any person remaining in tier t, $0 \leq t < T$, when the capital stock increases, the optimal hierarchy will be adjusted so that:

(i) his effort level increases ($da_t/dN>0$); and

(ii) his wage payment also increases ($dw_t/dN>0$).

The results are true in particular for managers at the top ($t=0$): $da_0/dN>0$ and $dw_0/dN>0$.

Proof: See Appendix B.■

Corollary. The wage ratio between the top managers and workers increases as the hierarchy expands, assuming a constant productivity parameter θ.

Since the marginal product of the top managers (which is $\theta N y_T$) increases as a result of an increase in N (to be shown in Proposition 7), the efforts from the top managers should be expected to increase, so should their wages. Proposition 5 applies to all people who occupy those positions that often link to the relative distance to the top of the hierarchy, for example, the vice president of a firm. Top managers in a larger organization exert more effort, and get more pay than their counterparts in a smaller organization.

Since $du_0/dN=g'(a_0)(s_0-1)(da_0/dN)+g(a_0)(ds_0/dN)$, a larger size of the organization gives the top managers higher utility provided ds_0/dN is not too negative. It follows that the top managers have a built-in (endogenous) incentive to maximize size. This results can be related to several literature. First, it relates to theories of the managerial firm à la Baumol (1967) and others. We have here an endogenous explanation of managerial size maximization based on efficiency wage model of hierarchy (with endogenous number of tiers). Second, it relates to the more recent literature on mergers and acquisitions. Third, it is also consistent with the observation of the so-called "expansion drive" from leaders or managers in socialist economies (Kornai, 1980).

In the way we built the model, all the people in the hierarchy are identical *ex ante*. Proposition 3 says that in an optimal hierarchy with N fixed, people holding a higher position in the hierarchy should be paid more than people at a lower position. The above Corollary says more: As N increases, the wage ratio between the top managers and the bottom workers increases, and the distribution becomes less egalitarian.[16]

What happens if θ increases with N fixed? This will increase the effort as well as the wage payment to the person who maintains a relative position in the hierarchy either to the top or to the bottom (so to the person in between, by Proposition 3). In such a case, workers are paid more not because their responsibility is higher as a result of promotion, but because they are more productive.

Proposition 6. For any person remaining in tier τ, $0 \leq \tau < T$ (counting from the bottom), or remaining in tier t, $0 \leq t < T$ (counting from the top), when the productivity of a fully effective worker increases, the optimal hierarchy will be adjusted so that:

(i) his effort level increases ($da^{\tau}/d\theta > 0$, and $da_t/d\theta > 0$); and

(ii) his wage payment also increases ($dw^{\tau}/d\theta > 0$, and $dw_t/d\theta > 0$).

The results are true in particular for workers ($\tau = 0$) and for the top managers ($t = 0$): $da_0/d\theta > 0$, $da_T/d\theta > 0$, $dw_0/d\theta > 0$ and $dw_T/d\theta > 0$.

Proof: Using the chain rule of differentiation and applying Lemma 2 and Lemma 4 yield $da^{\tau}/d\theta > 0$ and $dw^{\tau}/d\theta$. Differentiating $\xi(a_t) = ce^{T-t}$ with respect to θ and applying Lemma 4 and using $\partial T/\partial\theta = 0$ yield $da_t/d\theta > 0$ and $dw_t/d\theta > 0$.■

5.2. Total Number of Tiers and Total Managerial Effectiveness

---

[16] Since $u_0/u_T = [g(a_0)/g(a_T)] \bullet [(s_0 - 1)/(s_T - 1)]$, the utility ratio between the top managers and workers increases as the hierarchy expands provided $(s_0 - 1)/(s_T - 1)$ does not decrease too fast.

Proposition 7. (i) As the total amount of capital increases, the total number of tiers of the hierarchy increases ($dT/dN>0$); and

(ii) as the productivity parameter θ increases, the total number of tiers of the hierarchy increases ($dT/d\theta>0$) if and only if the span of control at the bottom is larger than that at the top ($s_T>s_0$).

Proof: Differentiating $\xi(a_t) = ce^{T-t}$ with respect to N and applying Proposition 5 together with Lemma 4 imply $dT/dN>0$. Using the chain rule of differentiation and applying Lemma 3(iii) and Lemma 4 give $dT/d\theta>0$.■

Proposition 8. (i) An increase in N leads to a decline in the total managerial effectiveness ($d(y_T)/dN<0$); and

(ii) the total managerial effectiveness increases as the productivity of a fully effective worker increases ($d(y_T)/d\theta>0$).

Proof: See Appendix B.■

Proposition 8 indicates an increasing loss of control over the entire economy as N grows. Intuitively, as the capital stock increases and more tiers are added, the aggregate of effort, y_T, should decrease, since each a is less than 1. On the other hand, when technological progress leads to an increase in θ, the total managerial effectiveness increases. The reason is that when the hierarchy is more productive, the owner has more resources to spend on implementing higher effort from every employee.

5.3. **Returns to Capital**

Proposition 9. (i) The total return to capital is $V(N,\theta) = N(\theta y_T - w_T) + w_0$, which is an increasing function of the capital stock. In fact, the envelope theorem holds: $dV/dN = \theta y_T - w_T > 0$; and

(ii) the marginal return to capital, declines as the economy expands, that is, $d^2V/dN^2 < 0$. Therefore, $V(N,\theta)$ is an increasing and concave function of N.[17]

Proof: See Appendix B.

Concavity of N means that the organization exhibits decreasing returns to scale. The "fixed factor" is the limited time the owner of the organization has. If the owner of the organization had unlimited time, it would set up a two tier "hierarchy" and check everyone with probability one. In such a case, the first best would be implemented and the organization would exhibit constant returns to scale. When the owner has only limited time for monitoring, it has to face the following dilemma: on the one hand, it is in his interest to expand the organization without any limit (since $dV/dN > 0$); on the other hand, the organization becomes less and less efficient, and the marginal return to each additional unit of capital shrinks.

Proposition 10. (i) The total return to capital is an increasing function of the productivity parameter θ, and the envelope theorem also holds: $dV/d\theta = Ny_T$;

(ii) net revenue is a convex function in the productivity parameter θ, that is, $d^2V/d\theta^2 > 0$,[18] and

---

[17] Concavity of V(N) can be understood as follows. Consider the situation in which the optimal hierarchy is split into s_0 smaller hierarchies while everything else is kept the same. Now productivity is higher because of less number of tiers and savings of wages being paid to the s_0 supervisors. Re-optimization of the s_0 smaller hierarchies increase productivity even more. Therefore, the rate of return to capital must go up by making the hierarchy smaller.

[18] Suppose θ is doubled. If the hierarchy is kept the same as before, output will be doubled while the wage bill will be the same, so the net output will be more than doubled. Re-optimization increases the net output even more. Therefore, V is a convex function of θ.

(iii) the marginal return to capital increases as the productivity parameter of a fully-effective worker θ increases, that is $d^2V/dNd\theta>0$.

Proof: See Appendix B.

The result of the convexity of the net revenue function in the productivity parameter should be compared with our reference model with 0 and 1 effort levels: the marginal increase in net revenue due to the increase in the productivity parameter θ is constant there ($dV/d\theta=N$, by Proposition 1(v)). This is equally true if monitoring is perfect (i.e., $P=1$): the marginal increase in total net revenue due to the increase in the productivity parameter θ equals a constant Na^*, where a^* is the effort level under perfect monitoring. Hence the increase in θ makes loss of control less severe, and moreover, its role is magnified through the hierarchical levels.

6. Applications and Extensions

Organizations of Capitalist Firms and Socialist Economies. One motivation of the present study is comparisons of hierarchical structures of capitalist firms and socialist economies. In our view, one difference between the centrally-planned economy and a market economy is the difference between one huge hierarchy and the simultaneous existence of many small an large hierarchies competing each other in the market.

The distinctive features of the capitalist economy are the dispersed ownership of capital. Therefore, the size of hierarchies, which is endogenously determined by the market competition, will be relatively small. If the marginal cost of capital faced by a particular firm in the market is r, then in our model the firm sets up a hierarchy with capital K^* and the number of workers $N^*=K^*/k$ such that the marginal returns to capital is equal to the marginal cost:

$$dV(N^*,\theta)/dN = rk.$$

In contrast, under state-ownership in centrally-planned economies the entire economy becomes virtually one firm that uses a single hierarchy.[19] In this framework, the problem with socialist economies is that of "bigness:" with all capital owned centrally, production is organized by a hierarchy that is necessarily long and inefficient.[20]

Ability in Hierarchy. Our model can be extended into several directions. One possibility is the introduction of agents with different abilities. Ability can affect both the production and monitoring processes. For example, if the production function is modified to $y_t=y_{t-1}(a_t)^\beta$, where $\beta>0$, a smaller β represents a higher ability of the manager for reducing loss of control. On the other hand, with a generalized monitoring technology $P=\min\{\gamma/s,1\}$, a high γ manager is able to monitor more efficiently and therefore a lower efficient wages are required for his subordinates. Placing high ability agents in the higher tiers may introduce interesting effects on efforts and wages in an optimal hierarchy, which deserves further study.

---

[19] Lenin wrote in his famous book *The State and Revolution* (1917): "All citizens [in a socialist society] become employees of a <u>single</u> countrywide state 'syndicate.'" "The whole of society will become a single office and a single factory" (underline is original).

[20] It follows that there are potential gains from privatization (or making ownership decentralized), simply because the resulting decentralization reduces the amount of capital per "firm."

References

Baumol, William (1967), *Business Behavior, Value, and Growth*, revised edition, New York: Harcourt Brace Jovanovich.

Beckmann, M. (1977), "Management Production Function and the Theory of the Firm," *Journal of Economic Theory*, vol. 14, 1-18.

Calvo, Guillermo and Stanislaw Wellisz (1978), "Supervision, Loss of Control and the Optimal Size of the Firm," *Journal of Political Economy*, 86, 943-952.

Calvo, Guillermo and Stanislaw Wellisz (1979), "Hierarchy, Ability and Income Distribution," *Journal of Political Economy*, 87, 991-1010.

Holmstrom, Bengt and Jean Tirole (1988), "The Theory of the Firm," in R. Schmalensee and R. Willig (eds), *Handbook of Industrial Organization*. Amsterdam: North Holland.

Keren, Michael and David Levhari (1979), "The Optimal Span of Control in a Pure Hierarchy," *Management Sciences*, Vol. 25, pp. 1162-1172.

Keren, Michael and David Levhari (1983), "The Internal Organization of the Firm and the Shape of Average Costs," *Bell Journal of Economics*, 14, 474-486.

Keren, Michael and David Levhari (1989), "Decentralization, Aggregation, Control Loss and Costs in A Hierarchical Model of the Firm," *Journal of Economic Behavior and Organization*, (11) pp. 213-236.

Kornai, Janos (1980), *Economics of Shortage*, Amsterdam: North Holland.

McAfee, Preston and John McMillan (1990), "Organizational Diseconomies of Scale," mimeo, University of California at San Diego.

Melumad, Nahum, Dilip Mookherjee and Stefan Reichelstein (1989), "Hierarchical Decentralization of Incentive Contracts," mimeo, Stanford University.

Qian, Yingyi (1990), "Hierarchy, Loss of Control, and a Theory of State-Ownership in Socialist Economies," Essay I in *Incentives and Control in Socialist Economies*, unpublished Ph.D. dissertation, submitted to the Department of Economics, Harvard University.

Rosen, Sherwin (1982), "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, Autumn.

Simon, Herbert A. (1976), *Administrative Behavior*, 3rd ed, New York: Free Press.

Tirole, Jean (1986), "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organizations* (2), pp.181-214.

Williamson, Oliver (1967), "Hierarchical Control and Optimal Firm Size," *Journal of Political Economy*, 123-138.

Appendix A. The Case of 0 and 1 Effort Choices in a Discrete Model

The discrete problem is

$$\min \quad \Sigma_1^T s_t x_t$$

$$s_t,\, x_t,\, T$$

$$\text{s.t.} \quad x_t = x_{t-1} s_t$$

$$x_0 = 1 \text{ and } x_T = N.$$

Upon substituting for $s_t = x_t / x_{t-1}$, this problem becomes

$$\min \quad \Sigma_1^T x_t^2 / x_{t-1}$$

$$x_t,\, T$$

$$\text{s.t.} \quad x_0 = 1 \text{ and } x_T = N.$$

We solve this problem in two steps.[21] First, for fixed T, minimization with respect to $x_1, \ldots, x_T$ requires

$$2x_t^3 = x_{t-1}\, x_{t+1}^2,$$

or

$$2s_t = s_{t+1}^2,$$

or

$$\ln s_{t+1} = [\ln 2 + \ln s_t] / 2.$$

Upon solving this difference equation for $\ln s_t$, we obtain

$$\ln s_{t+1} = \ln 2 + (1/2)^t (\ln s_1 - \ln 2).$$

Since $\ln x_t = \Sigma_1^t \ln s_t$,

$$\ln x_t = t \ln 2 + 2 (\ln s_1 - \ln 2)(1 - 2^{-t}).$$

Using boundary condition $x_T = N$, we obtain

$$\ln x_t = \ln 2^t + (1 - 2^{-t})(\ln N - \ln 2^T)/(1-2^{-T}).$$

Secondly, we determine the optimal T. Substituting the above expression, the minimand becomes

---

[21] I thank Martin Hellwig for showing me this proof.

$$\Sigma_1^T \ \exp\,[2 \ln x_t - \ln x_{t-1}]$$

$$= 4(A-1)(N/A)^{A/(A-1)},$$

where $A=2^T$. This expression is minimized at $(N/A^*)^{A^*/(A^*-1)} = 1$, or equivalently, at $A^*=N$. Then for this optimal T (neglecting the integer constraint on T), we obtain:

Proposition A.1. Assume workers are sufficiently productive: $\theta>4g$. Then the optimal hierarchy has the following properties:

(i) The span of control is constant: $s_t=2$;

(ii) The wage scale is also constant: $w_t=2g$;

(iii) The number of employees in tier t is $x_t=2^t$;

(iv) The total number of tiers $T(N) = (\ln N) /(\ln 2)$ for all $N\geq 2$, and $dT/dN = 1/[N \ln 2]$; and

(v) Total net revenue is $V(N) = N\theta - 4g(N-1)$ and the marginal return is constant: $dV/dN = \theta-4g$.

We can similarly derive the solution to the discrete version of the problem for fixed T:

Proposition A.2. Define $c=\theta/(4g)$ and assume workers are sufficiently productive: $c>1$. Then when T is fixed, the hierarchy will be adjusted to the point so that:

(i) The span of control decreases in t: $s_t=2c^{A2^{-t}}$;

(ii) The wage also decreases in t: $w_t=2gc^{A2^{-t}}$;

(iii) The number of employees in tier t is $x_t=2^t c^{A(1-2^{-t})}$; and

(iv) The maximum number of workers absorbed is $N=Ac^{A-1}$.

The difference between this solution and the one obtained in the continuous model (Propositions 1 and 2) is e being replaced by 2 in (i)-(iv) and by 4 in (v) and c.

Appendix B. Mathematical Proofs

Proof of Lemma 1.

Adding up (4.1) and (4.2), we get

$$\dot{w}_t/w_t + \dot{x}_t/x_t = 1, \text{ or}$$

(B.1) $$d(w_t x_t)/dt = w_t x_t.$$

The solution to this differential equation is

(B.2) $$w_t x_t = c_1 e^t,$$

where c_1 is a constant. Using the boundary conditions $x_0=1$ and $x_T=N$, we obtain,

(B.3) $$c_1 = w_0 = w_T e^{-T} N.$$

On the other hand, adding up (4.3) and (4.4), we have

$$\dot{q}_t/q_t + \dot{y}_t/y_t = 0,$$

which gives

(B.4) $$q_t y_t = c_2,$$

where c_2 is a constant. Using the boundary condition $q_T=\theta N$, we obtain

(B.5) $$c_2 = q_0 = q_T y_T = \theta N y_T.$$

By equation (4.6), and using (A.2)-(A.5) above, we get

(B.6) $$\xi(a_t) = (q_t y_t)/(w_t x_t) = c_2/(c_1 e^t) = (\theta y_T/w_T)e^{T-t}, \text{ or}$$

(B.7) $$a^\tau = h(ce^\tau),$$

which proves (i). Parts (ii) and (iii) are similarly derived from equations (4.7) and (4.5).

Finally, assumptions on g imply

$$g'(a) > g(a)/a,$$

that is, $\xi(a)>1$ for all a. In particular, $c = \theta y_T/w_T = \xi(a_T) > 1$.■

Proof of Lemma 4.

Differentiating $w_T c = \theta y_T$ with respect to N yields

$$[w_T + c \bullet (dw_T/dc)] \bullet (dc/dN) = \theta \bullet [(\partial y_T/\partial c) \bullet (dc/dN) + (\partial y_T/\partial N)], \text{ or}$$

$$[w_T + c \bullet (dw_T/dc) - \theta \bullet (\partial y_T/\partial c)] \bullet (dc/dN) = \theta \bullet (\partial y_T/\partial N). \tag{B.8}$$

Similarly, we derive

$$[w_T + c \bullet (dw_T/dc)] \bullet (dc/d\theta) = \theta \bullet [(\partial y_T/\partial c)] \bullet (dc/d\theta) + (\partial y_T/\partial \theta)\} + y_T.$$

Note that $y_{T(c,N)}$ is not a function of θ, that is, $\partial y_T/\partial\theta=0$, hence,

$$[w_T + c \bullet (dw_T/dc) - \theta \bullet (\partial y_T/\partial c)] \bullet (dc/d\theta) = y_T. \tag{B.9}$$

Because $(\partial y_T/\partial N)<0$ (Lemma 3(ii)), it is sufficient to prove $[w_T + c \bullet (dw_T/dc) - \theta \bullet (\partial y_T/\partial c)] > 0$.

For any N, $c=\theta y_T/w_T$ must solve

$$\max_c \{\theta y_{T(c,N)} N \ - \int_0^{T(c,N)} (w^\tau x^\tau)\, d\tau\}.$$

The first order condition requires

$$\theta(\partial y_{T(c,N)}/\partial c) \quad = (1/N) \bullet \{ (\partial T/\partial c) \bullet (w^T x^T) \ + \int_0^{T(c,N)} (\partial/\partial c)(w_T e^{-T} N e^{T-\tau})\, d\tau \}.$$

Uing (B.3), we obtain

$$\theta(\partial y_{T(c,N)}/\partial c) \quad = w_T e^{-T}(\partial T/\partial c) + (1-e^{-T}) \bullet (dw_T/dc). \tag{B.10}$$

Substituting (4.7) into Lemma 2(iii) yields,

$$(1/w_T) \bullet (dw_T/dc) = [\log(s_T)-1]/c. \tag{B.11}$$

Therefore, using (B.10) and (B.11) and applying Lemma 3(iii), we derive

$$\begin{aligned} & w_T + c \bullet (dw_T/dc) - \theta \bullet (\partial y_{T(c,N)}/\partial c) \\ &= [\ w_T + (c-1) \bullet (dw_T/dc)\ ] + e^{-T} w_T\ [\ (1/w_T) \bullet (dw_T/dc) - (\partial T/\partial c)\ ] \\ &= [\ w_T + (c-1) \bullet w_T \bullet [\log(s_T)-1]/c\ ] + e^{-T} w_T\ (1/c)\{[\log(s_T)-1]-[\log(s_T)/\log(s_0) - 1]\} \end{aligned} \tag{B.12}$$

$= (w_T/c)\bullet\{ 1 + (c-1)\bullet\log(s_T) + e^{-T}\bullet\log(s_T)\bullet[\log(s_0)-1]/\log(s_0) \}$

$> 0.$■

Proof of Proposition 5.

(i) Differentiating $\xi(a_t) = ce^{T-t}$ with respect to N yields

$$\xi'(a_t)\bullet da_t/dN = e^{T-t}(dc/dN) + ce^{T-t}[(\partial T/\partial c)(dc/dN) + \partial T/\partial N]$$

$$= \xi(a_t)\ \{(1/c)[1+c(\partial T/\partial c)]\bullet(dc/dN) + \partial T/\partial N\}.$$

Applying Lemma 3(iii),

$$(\Phi(a_t)/a_t)\bullet(da_t/dN) = (1/c)\bullet(\log(s_T)/\log(s_0))\bullet(dc/dN) + \partial T/\partial N.$$

On the other hand, from (B.8)

$$dc/dN = \{\theta\bullet(\partial y_T/\partial N)\} / \{w_T + c\bullet(dw_T/dc) - \theta\bullet(\partial y_T/\partial c)\}.$$

Because of (B.12) and Lemma 3(i), it is sufficient to show

$$(1/c)\bullet\theta\bullet N\bullet\log(s_T)\bullet(\partial y_T/\partial N) + \{w_T + c\bullet(dw_T/dc) - \theta\bullet(\partial y_T/\partial c)\} > 0.$$

Substituting Lemma 3 and equation (B.12), the above expression is equal to

$$(1/c)(\log(s_T))(\theta y_T)\log(a_0)/\log(s_0) + (w_T/c)\{1+(c-1)\log(s_T) + e^{-T}\bullet\log(s_T)\bullet[\log(s_0)-1]/\log(s_0)\}$$

$$= (1/c)\log(s_T)e^{-T}w_T\ [\xi(a_0)\log(a_0) + \log(s_0) - 1]/\log(s_0) + (w_T/c)[1+(c-1)\log(s_T)]$$

$$= (w_T/c)[1+(c-1)\log(s_T)$$

$$> 0,$$

where $\xi(a_0)=ce^T$ and equation (4.7) are used.

(ii) Differentiating $1-\log(s_t)=\xi(a_t)\log(a_t)$ and $\log(w_t)=\log(g(a_t))+\log(s_t)$ with respect to N and using $da_t/dN > 0$ generate $dw_t/dN > 0.$■

Proof of Proposition 8.

(i) Differentiating $w_Tc = \theta y_T$ with respect to N and using Lemma 4 and Lemma 2(iii).

(ii) Since $\partial y_T/\partial\theta = 0$,

$$dy_T/d\theta = (\partial y_T/\partial c)(dc/d\theta) + \partial y_T/d\theta = (\partial y_T/\partial c)(dc/d\theta).$$

Because $dc/d\theta>0$, it is sufficient to show $\partial y_T/\partial c>0$.

If $\partial T/\partial c>0$, from (B.10),

$$\theta(\partial y_{T(c,N)}/\partial c) \quad = w_T e^{-T}(\partial T/\partial c) + (1-e^{-T})\bullet(dw_T/dc) > 0,$$

by Lemma 2(iii).

If $\partial T/\partial c<0$, then using Lemma 2(i),

$$(\partial/\partial c)\log(y_T) \quad = \quad (\partial/\partial c) \int_0^{T(c,N)} \log(a^\tau(c))\, d\tau$$

$$= (\partial T/\partial c)\bullet\log(a^T) \quad + \int_0^{T(c,N)} (1/a^\tau)(da^\tau/dc)\, d\tau$$

$$> 0.\blacksquare$$

Proof of Proposition 9.

(i) Using (B.2) and (B.3), we obtain

$$V(N,\theta) \quad = N\theta y_T \;-\; \int_0^T \{w_t x_t\}\, dt$$

$$= N(\theta y_T - w_T) + w_0.$$

Differentiating $V(N) = \theta N y_T \;-\; \int_0^T \{g(a_t)s_t x_t\}\, dt$ yields

$$dV/dN = \theta y_T + \theta N(dy_T/dN) - g(a_T)s_T x_T(dT/dN) - \int_0^T (d/dN)\{g(a_t)s_t x_t\}\, dt.$$

$$= \theta y_T - w_T,$$

using integration by part and equations (4.1)-(4.7), (B.2)-(B.5) and $H(t)=0$.

(ii) Differentiating $V(N) = N(\theta y_T - w_T) + w_0$ yields

$$dV/dN = (\theta y_T - w_T) + N\,[\theta\,(dy_T/dN) - (dw_T/dN)] + dw_0/dN$$

$$= (\theta y_T - w_T) + N\,(d^2V/dN^2) + dw_0/dN.$$

But $dV/dN = (\theta y_T - w_T)$ implies $N\,d^2V/dN^2 + dw_0/dN = 0$, or $d^2V/dN^2 = -(1/N)(dw_0/dN) < 0$, by Proposition 5(ii).■

Proof of Proposition 10.

(i) Similar to the proof of Proposition 9(i).

(ii) $d^2V/d\theta^2 = N\,dy_T/d\theta > 0$ by Proposition 8.

(iii) Taking a derivative of $dV/dN = \theta y_T - w_T$ with respect to θ, and using $\partial y_T/\partial\theta = 0$ and (B.9), we have

$$d^2V/d\theta dN = y_T + \theta\bullet(dy_T/d\theta) - dw_T/d\theta$$

$$= [w_T + c\bullet(dw_T/dc)-\theta\bullet(\partial y_T/\partial c)]\bullet(dc/d\theta) - (dw_T/dc - \theta\bullet(\partial y_T/\partial c))\bullet(dc/d\theta)$$

$$= [w_T + (c-1)\bullet(dw_T/dc)]\bullet(dc/d\theta) > 0,$$

by Lemma 2(iii) and Lemma 4.■