Preliminary and incomplete.

Structure and Returns to Scale of Real-Time Hierarchical Resource Allocation

Timothy Van Zandt*
INSEAD

29 March 2002

Abstract

Companion papers develop a model of real-time hierarchical computation of resource allocations by boundedly rational members of an administrative staff. The nodes of a hierarchy are multiperson decision-making units offices. The current paper uses a reduced form to address specific questions about organizational structure and returns to scale. We find that the possibility of decentralizing decision making within these hierarchical organizations allows for larger hierarchies. However, organization size is still bounded because the combined effect of cumulative delay and administrative costs means that in large enough hierarchies, the value of the root office's information processing is less than the office's administrative costs. We also find that as the environment changes more rapidly, optimal hierarchies become smaller and more internally decentralized. A speed-up of managerial processing, such as through improved information technology, has the opposite effect.

JEL Classifications: D83, D23
Keywords: decentralization, hierarchies, bounded rationality, resource allocation, real-time computation

Author's address:

INSEAD	Voice:	+33 1 6072 4981
Boulevard de Constance	Fax:	+33 1 6074 6192
77305 Fontainebleau CEDEX	Email:	tvz@econ.insead.edu
France	Web:	zandtwerk.insead.edu

*This research was supported in part by grant IRI-9711303 from the National Science Foundation.

Contents

1 Introduction

In bureaucratic organizations such as firms, government agencies, and militaries, some decisions are made through a hierarchical structure. Actions or control variables are associated with operatives at the bottom of the hierarchy. The upper tiers of the hierarchy perform the administrative task of coordinating the operatives in response to information about the environment. Information flows up from the operatives; each node of the hierarchy receives information from its subordinates, aggregates this information, and communicates to its superior. Decisions are recursively disaggregated through a flow of information down the hierarchy. For example, resources may be allocated this way: each node receives a budget from its superior, and divides this budget among its subordinates. The hierarchy, which may be depicted by an organization chart, is a coarse description of the structure of decision making, rather than a detailed description of the flow of information between every member (clerk, secretary, manager, technician) of the administrative staff.

This paper studies a reduced-form model of such hierarchical decision procedures. It views these procedures as ways for organizations to make effective decisions in environments that are complex relative to the cognitive abilities of any single human. The complexity is tied to the changing nature of the environment and the scale of the operations.

In a conventional model with fully rational agents, a single manager can run a firm of any size or control any number of its operatives, instantly responding to new information. Even if this agent is controlling other agents with conflicting objects and private information, the manager can use a direct revelation mechanism in which the manager receives all information and makes decisions centrally, according to a prearranged rule.

However, because humans can only make use of limited amounts of information in a given period of time, a single person can effectively control operations if their scale is small or the environment is changing slowly. By sharing the administrative task among many agents, information can be aggregated and used more quickly because the agents can perform certain tasks concurrently (Radner (1993)); this is an analogue to the speed-up entailed by parallel processing in computer systems. Still, the nature of coordination limits the speed-up that such decentralization can achieve, because not all the different tasks into which decision making can be divided can be performed concurrently. This inexorable increase in delay can lead to another form of decentralization. By having some decisions made in lower levels of a hierarchy, these decisions can be made using less aggregate and hence more recent information.

This current paper studies a reduced-form model of such decision-making hierarchies. We set up the model by first describing informally the parameters such hierarchical procedures may have and how the performance should depend on these parameters and then stating a formula for the costs of a hierarchy as a function of these parameters. We hope that some readers will become sufficiently curious (or suspicious) to read Van Zandt (2002b, 2002a), in which the formula is derived "from first principles", using an explicit model of the resource allocation problem, of the computational abilities of the administrative staff, and of the decision procedures that lie behind each hierarchical form. However, the current paper is self-contained.

Our goal is to characterize (i) the costs and benefits of decentralized decision making, (ii) the shape of the hierarchies, (iii) returns to scale, and (iv) how these properties depend on the speed at which the environment changes. We begin, in Section 3, with a quantification of the costs and benefits of decentralized decision making. The decision-theoretic value of such decentralization is that intermediate offices can use more recent information than their superiors when disaggregating allocations. There is a decision-theoretic cost: the more tiers through which the period-t allocation decisions are disaggregated, the earlier an upper-level office must inform its subordinates of their period-t allocations. The managerial cost of decentralization is that passing decisions through subordinates increases the amount of calculation that must be done.

In Section 4, we characterize the returns to scale of the hierarchical procedures. If the number of tiers is fixed, and hence decentralization of decision making is limited, then there is a bound on the optimal size of hierarchies even if the managerial wage is zero. However, when the number of tiers in the optimal hierarchies is allowed to increase with the number of operatives, computational delay alone does not inexorably lead to eventually decreasing returns to scale. This is because expected shop costs can always be reduced by joining independent hierarchies under a central office that coordinates allocations to these hierarchies. The former decentralization of decision making between independent organizations then becomes decentralized decision making within a single organization. Nevertheless, when instead the wage is positive and the hierarchies are large, the value of this coordination is lower than the administrative cost because of cumulative delay. Therefore, managerial wages combined with delay lead to a bounded size of hierarchies.

In Section 5, we characterize the spans in optimal hierarchies. We show, for example, that spans are single-peaked and that, in particular, then tend to decrease moving down the hierarchy. In Section 6, we characterize how the optimal firm size and shape of the hierarchies depend on the speed with which the environment changes. We find that organizations are smaller and more internally decentralized the more rapidly the environment changes. An increase in managerial speed, which might result from improvements in information technology, has the opposite effect. This provides a framework for the empirical study of trends in firm structure (which have involved both downsizing and mergers) and how they relate to improvements in information technology and the endogenous increases in the rate of change of firms' strategic environments.

2 A model decision-making hierarchies

We consider a model of hierarchical decision procedures, coordination mechanisms, or organizations. It is a reduced form of the model of balanced CF hierarchies, developed in

Van Zandt (2002b, 2002a) from two components: (a) a stochastic control problem, which is to allocate resources to "shops" whose valuations or payoffs change over time, and (b) a model of the information processing capabilities of the managers who decide the resource allocations.

Each organization is represented by a hierarchy, which is a tree whose leaves are the shops or recipients of resources and whose nonleaf nodes are multiperson offices that are the decision-making units. The shops' resource allocations are periodically updated through a recursive downward flow of decisions through the hierarchy. Each office is informed by its superior of the amount of resource available for the shops below it in the hierarchy, and the office decides how to divide this amount among its subordinates. These decisions require aggregate information about the shops' payoff functions, so that each office can implicitly calculate a shadow price and the marginal valuations of its subordinates. This aggregation occurs periodically through a recursive upward flow of information. Each office receives information about its subordinates and aggregates this information, both in order to pass this information to its superior and in order to allocate resources to its subordinates.

The key information processing limitation is that it takes managers time to process information. This time is costly because of managerial wages that must be paid. A more important implication, however, is that decisions are always based on old information. The organizational structure determines how effectively the organization keeps up with its changing environment. The role of the recursive disaggregation of allocations, which is a form of decentralized decision making, is as follows. If there is a single office that makes all decisions, then all resource allocations are based on very aggregate and hence very old information. By delegating some decisions to offices lower in the hierarchy, each of which is above a smaller number of shops than is the root node, these decisions are made using less aggregate and hence more recent information.

To obtain a simple model, Van Zandt (2002b, 2002a) imposes considerable symmetry by assuming that the managers are homogeneous and the shops are identical (or, more precisely, that the shops' payoff functions are i.i.d.) and by restricting attention to balanced hierarchies. A hierarchy is balanced if all shops are the same distance from the root and if each office in the same tier has the same *span* or number of subordinates.

The balanced hierarchies have a simple parameterization. We index the tiers of such a hierarchy from bottom to top by $h = 0, 1, \ldots, H$, where H is called the height of the hierarchy. We denote the number of nodes in tier h by q_h. Since there should always be a single root, we set $q_H = 1$. Since the shops are in the bottom tier, q_0 is the number of shops; sometimes this number is fixed and sometimes we allow it to vary. As a continuous approximations, we treat $s_h \triangleq q_{h-1}/q_h$ as the span of each office in tier h. To do anything useful, an office has to have at least two subordinates, and so we require that $s_h \geq 2$ for $h = 1, \ldots, H$. Otherwise we ignore integer constraints on the numbers of nodes in each tier. Then the set of hierarchies of height H is

$$Q_H \triangleq \left\{ \boldsymbol{q} = \langle q_0, q_1, \ldots, q_{H-1} \rangle \in \mathbb{R}^H \;\middle|\; q_{h-1}/q_h \geq 2 \;\forall h = 1, \ldots, H \right\}$$

and the set of hierarchies of height H with n shops is $Q_H(n) \triangleq \{\boldsymbol{q} \in Q_H \mid q_0 = n\}$.

The underlying statistical model, based on Geanakoplos and Milgrom (1991), is such that the total payoff for the policy that the hierarchy calculates can be decomposed into the no-information payoff, which we normalize to zero, and the value of information processing by each office. The value of information processing by an office depends on both the hierarchical structure and on the information that the office implicitly uses to allocate resources to its subordinates. Recall that the office uses aggregate information about the shops' payoffs. The underlying model is also such that an aggregate datum is a sufficient statistic for the data from which it is calculated. However, information processing takes times and the resulting lags affect the quality of the office's information.

We denote by d_h the delay between when an office in tier h collects information about its subordinates and when it allocates resources to these subordinate. This delay is divided between delay in aggregating information and delay in disaggregating resource allocations. It adds d_h not only to the lag of the office's own information but also to the lag of the information of offices further up the hierarchy: the aggregation delay slows down how quickly the office can pass information up to its superior; the disaggregation delay means that the superior has to inform the office of its resource allocations with a lead time. Therefore, the cumulative lag of the information that an office in tier h uses about shops below it in the hierarchy is $L_h \triangleq \sum_{\eta=1}^{h} d_\eta$.

The parameters of the stochastic processes that govern the payoff functions are $\sigma^2 > 0$, which measures the overall volatility of the environment, and $b \in (0,1)$, which measures the correlation between old and new information and hence inversely measures the speed at which the environment is changing. The value of the information processing by an office in tier h is

$$v_h \;\triangleq\; \sigma^2(s_h - 1)b^{L_h}\,. \tag{2.1}$$

Since there are q_h offices in tier h and since $q_h s_h = q_{h-1}$, the total value of the information processing by offices in tier h is

$$q_h\left(\sigma^2(s_h - 1)b^{L_h}\right) \;=\; \sigma^2(q_{h-1} - q_h)b^{L_h}\,.$$

Therefore, the payoff of the hierarchy is

$$U_H(\boldsymbol{q}) \;\triangleq\; \sigma^2 \sum_{h=1}^{H}(q_{h-1} - q_h)b^{L_h}\,.$$

To complete the model, we specify, for each tier h, the delay d_h and the managerial cost of each office in that tier. These are derived from a model of information processing. Consider first the managerial cost. The workload of calculating each resource allocation is proportional to s_h. We introduce a parameter μ that is a linear measure of how long to takes managers to perform tasks, so that the workload is μs_h. Let w be the managerial wage, so that the total managerial cost of the office is $w\mu s_h$. Since there are q_h offices in tier h and since $q_h s_h = q_{h-1}$, the total managerial cost of the offices in tier h is $w\mu q_{h-1}$ and the total managerial cost of the hierarchy is $C_H(\boldsymbol{q}) \triangleq w\mu \sum_{h=0}^{H-1} q_h$.

The delay of an office in tier h would also equal μs_h if each resource allocation were calculated by a single manager. However, by decentralizing information processing within the office, the delay can be reduced to $d_h \triangleq \mu(\alpha + \log s_h)$. The delay $\mu \log s_h$ is from the aggregation of information, which is the essence of coordination in this decision problem. It increases with the amount of data to be aggregated because the operations cannot all be performed concurrently. The delay $\mu\alpha$ is from statistical filtering of the data and disaggregation of the resource allocations. It does not depend on s_h because some steps involve a single operation no matter how many subordinates the office has and others involve operations that can be performed concurrently for all the subordinates. For example, once information has been aggregated and implicitly a shadow price has been calculated, the individual resource allocations can be computed concurrently.

Since $d_h = \mu(\alpha + \log s_h)$, we have $L_h = \mu(\alpha h + \log(s_1 \cdots s_h))$. Define $n_h \triangleq s_1 \ldots s_h = q_0/q_h$, which is the number of shops below an office in tier h. Then $L_h = \mu(\alpha h + \log n_h)$; this equals the delay $\mu\alpha h$ in having information pass up and down through h levels and the cumulative delay $\mu \log n_h$ in aggregating information about n_h shops.

We have thus defined the payoff $U_H(\boldsymbol{q})$ and the cost $C_H(\boldsymbol{q})$ of a hierarchy of height H. Its *profit* is defined to be $\Pi_H(\boldsymbol{q}) \triangleq U_H(\boldsymbol{q}) - C_H(\boldsymbol{q})$. Combining the various formulae, we have

$$\Pi_H(\boldsymbol{q}) \;=\; \sigma^2 \sum_{h=1}^{H} (q_{h-1} - q_h)(b^\mu)^{\alpha + \log q_0/q_h} - w\mu \sum_{h=0}^{H-1} q_h \,. \tag{2.2}$$

Since the height of hierarchies can vary, we also denote a hierarchy by $\mathcal{H} = \langle H, \boldsymbol{q} \rangle$, where $H \in \{1, 2, \ldots\}$ and $\boldsymbol{q} \in Q_H$. We then denote its payoff, cost, and profit by $U(\mathcal{H})$, $C(\mathcal{H})$, and $\Pi(\mathcal{H})$, respectively. For example, $\Pi(H, \boldsymbol{q})$ is the same as $\Pi_H(\boldsymbol{q})$. Finally, since the number of shops can vary, we want to allow for the special case of a hierarchy $\mathcal{H}$ that has a single shop and hence no coordination; then $H = 0$, $q_0 = 1$, and $U(\mathcal{H}) = C(\mathcal{H}) = \Pi(\mathcal{H}) = 0$.

The main exogenous parameters that interest us are σ^2, b, w, and μ. The main purpose of this paper is to characterize optimal hierarchies and how they depend on these four parameters. Note that this dependence is only through b^μ and $w\mu/\sigma^2$; that is, there are only two degrees of freedom. To simplify notation in what follows, we define $W \triangleq w/\sigma^2$ and we normalize $\mu = 1$ until Section 7, when we consider what happens to optimal hierarchies when μ varies.

3 Benefits and costs of decentralized decision making

If a hierarchy has just two tiers ($H = 1$), then the root is the only office and it allocates resources directly to the shops. If instead the hierarchy has more than two tiers, then there are offices that allocate resources to other offices—thus, decision making is decentralized. The advantage of having an office in tier $h > 1$ allocate resources to an office in tier $h-1$ rather than directly to subordinates in tier $h-2$ is that the tier-$(h-1)$ office can suballocate resources using less aggregate and hence more recent information than that used by the tier h office. We quantify the benefits and costs of such decentralized decision making by comparing the profit before and after an entire tier of offices is eliminated.

Consider, then, a hierarchy $\mathcal{H} = \langle H, \boldsymbol{q} \rangle$ with height $H \geq 2$. Let $h \in \{1, \ldots, H-1\}$. Let $\mathcal{H}'$ be the hierarchy obtained by eliminating all offices in tier h. We use the standard notation for $\mathcal{H}'$ (augmented by primes), with the following exception: The tiers in $\mathcal{H}'$ are numbered $0, 1, \ldots, h-1, h+1, \ldots, H$, so that corresponding tiers in $\mathcal{H}$ and $\mathcal{H}'$ have the same number of nodes.

Proposition 3.1 *$U(\mathcal{H}) - U(\mathcal{H}') = \Delta_b - \Delta_c$, where*

$$\begin{aligned} \Delta_b &= \sigma^2 (q_{h-1} - q_h) b^{L_h} \left(1 - b^{\log s_{h+1}}\right) \quad \textit{and} \\ \Delta_c &= \sigma^2 \sum_{\eta=h+1}^{H} (q_{\eta-1} - q_\eta) b^{L'_\eta} (1 - b^\alpha)\,. \end{aligned}$$

$C(\mathcal{H}) - C(\mathcal{H}') = w q_h$.

Proof. Observe that the value of information for offices in tiers $1, \ldots, h-1$ is the same in $\mathcal{H}$ and $\mathcal{H}'$. Hence,

$$\begin{aligned} U(\mathcal{H}) - U(\mathcal{H}') &= \left(\sigma^2 (q_{h-1} - q_h) b^{L_h} + \sigma^2 \sum_{\eta=h+1}^{H} (q_{\eta-1} - q_\eta) b^{L_\eta}\right) \\ &\quad - \left(\sigma^2 (q_{h-1} - q_{h+1}) b^{L'_{h+1}} + \sigma^2 \sum_{\eta=h+2}^{H} (q_{\eta-1} - q_\eta) b^{L'_\eta}\right). \end{aligned} \tag{3.1}$$

By substituting

$$\sigma^2 (q_{h-1} - q_{h+1}) b^{L'_{h+1}} = \sigma^2 (q_{h-1} - q_h) b^{L'_{h+1}} + \sigma^2 (q_h - q_{h+1}) b^{L'_{h+1}}$$

and rearranging, we obtain

$$U(\mathcal{H}) - U(\mathcal{H}') = \sigma^2 (q_{h-1} - q_h)(b^{L_h} - b^{L'_{h+1}}) - \sigma^2 \sum_{\eta=h+1}^{H} (q_\eta - q_{\eta+1})(b^{L'_\eta} - b^{L_\eta})\,. \tag{3.2}$$

For $\eta \geq h+1$, the offices in "tier η" in $\mathcal{H}'$ are actually in tier $\eta - 1$. Therefore,

$$L'_\eta = \alpha(\eta - 1) + \log q_0 / q_\eta = L_\eta - \alpha\,.$$

Hence, $b^{L'_\eta} - b^{L_\eta} = b^{L'_\eta}(1 - b^\alpha)$. Therefore, the second term in equation (3.2) is equal to Δ_c as defined in Proposition 3.1.

Since $L'_{h+1} = \alpha h + \log s'_{h+1}$ and $L_h = \alpha h + \log s_h$, we have $L'_{h+1} - L_h = \log s'_{h+1} - \log s_h = \log s'_{h+1}/s_h$. Since also $s_h s_{h+1} = s'_{h+1}$, we have $L'_{h+1} - L_h = \log s_{h+1}$. Therefore, $b^{L_h} - b^{L'_{h+1}} = b^{L_h}(1 - b^{\log s_{h+1}})$, and the first term in equation (3.2) is equal to Δ_b as defined in Proposition 3.1.

Since $\boldsymbol{q}'$ has q_h fewer offices than $\boldsymbol{q}$, its administrative cost is wq_h lower. □

Here is the interpretation. Treat $\mathcal{H}'$ as the status-quo hierarchy, and let some of the decision making by offices in tier $h + 1$ be decentralized to a new set of q_h offices in a new tier h. Then Δ_b is the *increase* in payoff due to the fact that offices in tier h in $\mathcal{H}$ use data that is $\log s_{h+1}$ periods more recent than the information used by offices in tier $h + 1$ in $\mathcal{H}'$, due to the additional aggregation delay of the tier-$(h + 1)$ offices. This is the benefit of decentralization. Δ_b is the *decrease* in payoff due to the fact that the cumulative lag of offices in tiers $\eta \geq h + 1$ is α periods larger in $\mathcal{H}$ than in $\mathcal{H}'$, because of additional delay in updating the prediction of the aggregate payoff of the new offices and in disaggregating resource allocations to these offices. This is the decision-theoretic cost of decentralization. Furthermore, the administrative cost is wq_h higher in $\mathcal{H}$ than in $\mathcal{H}'$ because of the cost of performing the additional operations just mentioned.

4 Returns to scale

4.1 Motivation

Although the hierarchies we study permit internal decentralization, they still resemble tightly integrated, bureaucratic organizations such as firms and governments. In reality, there appear to be limits to the scale of such integration, since economic activity is carried out by many independent organizations that interact through spot markets or not at all. It has been conjectured, at least since the 1930's, that information processing constraints are a source of such limits.[1]

A proper model of these limits would allow for market interaction, but we address them using a simpler extension to our model in which allocations can be coordinated by multiple hierarchies that do not interact at all. Such a collection of independent hierarchies is called a *forest*. The total profit of a forest $\{\mathcal{H}_1, \ldots, \mathcal{H}_M\}$ is $\sum_{m=1}^{M} \Pi(\mathcal{H}_M)$. We say that, for a given number n of shops, a forest with a total of n shops is optimal if it has the highest profit of all such forests. Is there a limit to the size of the hierarchies in optimal forests?

If there were no information processing constraints, full integration would be optimal because larger organizations can take advantage of greater gains from trade and risk sharing.

[1]See Van Zandt (1998) and Van Zandt and Radner (2001) for references and discussion.

(This benchmark model is obtained by considering a one-tier hierarchy with zero lag and zero administrative cost. The profit is then $\sigma^2(n-1)$; the per-unit profit is $\sigma^2(n-1)/n$, which is increasing in n.) Thus, it is significant if this conclusion is reversed by the presence of information processing constraints.

This exercise tends to overestimate the optimal size of organizations, because it presumes that the coordination of resource allocations is only possible within the hierarchical procedures constructed in this paper, and hence allocations between independent organizations cannot be coordinated through markets. However, our main conclusion is that there is a limit to firm size due to the combined effects of delay and administrative expenses; the bound would become smaller rather than cease to exist if we were to allow coordination within markets. A full model of the determinants of organization size, which is a question of what transactions take place within bureaucratic organizations and what transactions take place in markets, would have to include a model of decision procedures that resemble market mechanism.

4.2 Net value of the root

A tool we use fo characterizing the size of hierarchies in optimal forests is the net value V_R^{net} of the root of a hierarchy $\mathcal{H} = \langle H, \boldsymbol{q} \rangle$, which is defined as follows. Suppose that the root is eliminated, so that the s_H subhierarchies become independent hierarchies. The difference between the payoff of the original hierarchy and the total payoff of the independent subhierarchies is equal to the value of the roots information, $v_H = \sigma^2(s_H - 1)b^{L_H}$. The administrative cost falls by ws_H, because the s_H offices in tier $H-1$ are not longer subordinates of any office. Therefore, the net fall in profit due to this divestiture is

$$V_R^{\text{net}} \triangleq v_H - ws_H = \sigma^2(s_H - 1)b^{L_H} - ws_H \,. \tag{4.1}$$

If V_R^{net} is positive, we now that the subhierarchies cannot exist independently in an optimal forest; if V_R^{net} is negative, then we know that the hierarchy cannot exist in an optimal forest.

4.3 Benchmark: zero wage

As a benchmark, consider optimal forests when (a) the managerial wage is zero and (b) we limit internal decentralization by bounding the height of hierarchies.

For example, suppose we allow for no decentralization at all, so that each hierarchy has height 0. The profit as function of the number n of shops in the hierarchy is

$$\sigma^2\left(\frac{n-1}{n}\right)b^{L_1} \,.$$

We can see that if L_1, which is the center's delay, were constant as n increased, then the per-shop payoff would increase monotonically. However, $L_1 = \alpha + \log n$ and hence

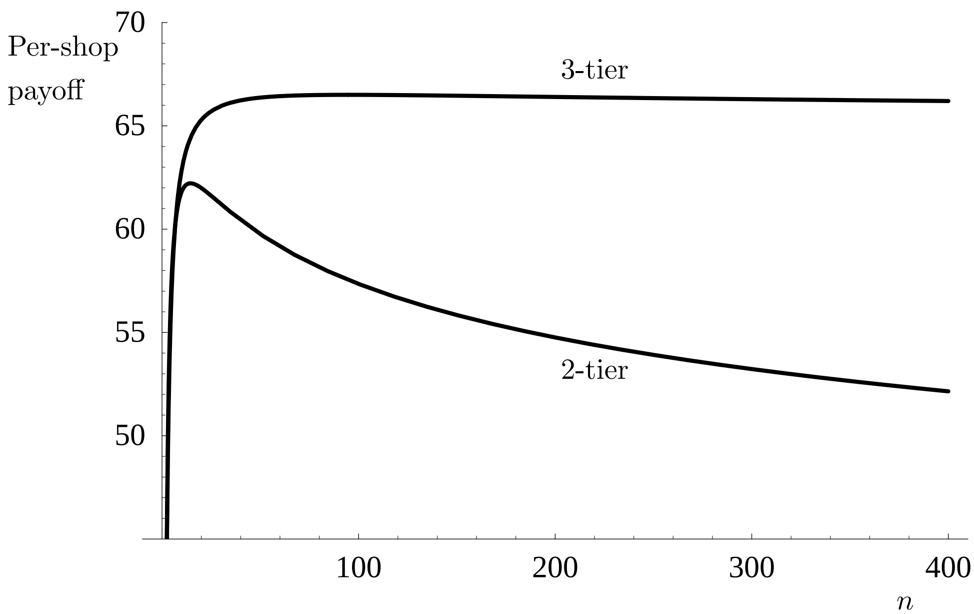

FIGURE 4.1. Per-shop payoff for 2-tier and 3-tier BCF hierarchies, as a function of organization size n. Parameter values are $\sigma^2 = 100$, $b = .95$, and $w = 0$.

$\lim_{n\to\infty} L_1 = \infty$. Therefore, $\lim_{n\to\infty} b^{L_1} = 0$ and the per-shop payoff converges, as $n \to \infty$, to the per-shop payoff when there is no coordination. That is, asymptotically the per-shop realized gains from trade are zero, because delay causes the allocations to be based on old information.

In fact, there is a limit to firm size even whenever the number of tiers of a hierarchy is bounded. For three-tier balanced hierarchies, it is only the root node whose cumulative lag increases inexorably with firm size (it is at least $\log n$). As n increases, the root node becomes irrelevant and the size of the subhierarchies under the root converges to the size that maximizes the per-shop payoff for two-tier hierarchies. However, at intermediate values of n, the root's information processing is valuable. Figure 4.1 shows the per-shop payoff for optimal two- and three-tier balanced hierarchies as a function of n when the wage is zero (for $b = .95$ and $\sigma^2 = 100$). Observe that the per-shop payoff for three-tier hierarchies is higher than for two-tier hierarchies (illustrating the benefit of decentralization), but the per-shop payoff is eventually decreasing for both classes of hierarchies.

A complete proof that there is a bound on firm size when we limit decentralization is given in Van Zandt (2002b, Section 6) for general (nonbalanced) CF hierarchies. We note simply that the proof is readily adapted to this model of balanced hierarchies.

Next suppose that we do not limit the height of the hierarchies. Suppose that a forest has M identical hierarchies of height H with a total of n shops and that these are integrated under a new root, to form a new hierarchy of height $H+1$. This integration does not affect the value of information processing of any of the offices in the existing hierarchies. The change in payoff is equal to the value of the new root's information: $\sigma^2(M-1)b^{\alpha(H+1)+\log n}$. Since this value is always positive, the integration raises the profit of the forest. We have thus proved the following proposition.

Proposition 4.1 *Suppose that $w = 0$. Then no optimal forest has two hierarchies of the same size. Therefore, there is no bound on firm size.*

Van Zandt (2002b) proves a stronger result for general CF hierarchies: that any optimal forest has a single hierarchy if $w = 0$. The proof is similar to the one given above. Since the value of the root of any CF hierarchy is positive, the payoff can be increased by merging independent CF hierarchies in a CF forest. The proof does not work for balanced hierarchies because, if two hierarchies are not identical, then merging them under a new root does not result in a balanced hierarchy.[2]

4.4 Positive managerial wage

Next we establish that there is a limit to firm size whenever $w > 0$. A necessary condition for the optimality of a forest is that the profit cannot be increased by eliminating the root of the one of the hierarchies to form independent hierarchies. If the hierarchy has height H and n shop, then the per-subordinate value of the root's information processing is approximately $\sigma^2 b^{\alpha H+\log n}$, which decreases to zero as $\lim_{n\to\infty}$. However, the per-subordinate cost $ws_H/s_H = w$ of the root's information processing is constant. Therefore, for large n, $V_R^{\text{net}} < 0$.

Proposition 4.2 *If $w > 0$, then* $\max\{1, W^{1/\log b}\}$ *is a limit to firm size.*

Proof of Proposition 4.2. From equation (4.1), the condition $V_R^{\text{net}} \geq 0$ implies

$$\begin{array}{rrcl} & s_R\left(\sigma^2 b^{\alpha H+\log n} - w\right) & \geq & \sigma^2 b^{\alpha H+\log n} \\ \Rightarrow & \sigma^2 b^{\alpha H+\log n} - w & > & 0 \\ \Rightarrow & b^{\log n} & > & w/\sigma^2 \\ \Rightarrow & n & < & W^{\frac{1}{\log b}}\,. \end{array}$$

This formula applies only when the r.h.s. is at greater than 2. Otherwise, when $W^{1/\log b} \leq 2$, the optimal firm size is 1 and there should be no administrative apparatus coordinating allocations. □

[2] A proof for balanced hierarchies could involve showing that the maximum per-shop payoff for hierarchies is an increasing function of the number of shops, but we have not attempted to prove or test this conjecture.

Thus, there are limits to the size of hierarchies because, in a large hierarchy, the central office is using such old information that the value of its decisions is less than the wages that must be paid to the agents in this office. In other words, the central office is too far removed from the daily operations of the organization, not in a spatial sense nor due to a lack of access to raw data about the daily operations, but because of the cumulative delay in aggregating information about these operations.

5 Optimal structure of the hierarchical procedures

In this section, we study the hierarchies that maximize profit for fixed height and number of shops or that maximize per-shop profit for fixed height and endogenous number of shops. When the number of shops is endogenous, we are treating the set of potential shops as if it were infinite or at least large, so that the forest that maximizes profit consists of many hierarchies, each of which maximizes per-shop profit without constraints on the number of shops. In each case, we refer to the profit-maximizing hierarchies as *optimal* hierarchies.

We characterize the optimal hierarchies by studying the first-order conditions of

$$\max_{q_1,\dots,q_{H-1}} \quad \Pi_H(q_0, q_1, \dots, q_{H-1}) \quad \text{and} \tag{5.1}$$

$$\max_{q_0,q_1,\dots,q_{H-1}} \quad \frac{1}{q_0}\Pi_H(q_0, q_1, \dots, q_{H-1}) \tag{5.2}$$

without paying attention to the integer constraints on the number of managers. The error from ignoring these constraints is smaller for lower tiers in the hierarchy, which have more managers. Leaving this caveat aside, we show, in each of the two cases, that a solution to the first-order conditions corresponds to a unique global maximum.

Consider first optimal hierarchies for fixed height and number of shops.

Proposition 5.1 Π_H *is a strictly concave function of* $\langle q_1, \dots, q_{H-1}\rangle$. *Therefore, a solution to the first-order conditions of equation (5.1) corresponds to a unique global maximum.*

PROOF. Proposition 4.4 of Van Zandt (2002a) demonstrates the strict concavity of U_H. (Note that the "proof" of Proposition 4.4 relies in part on a numerical test.) Since $\Pi_H(\boldsymbol{q}) = U_H(\boldsymbol{q}) - C_H(\boldsymbol{q})$ and C_H is linear, Π_H is strictly concave. □

Proposition 5.2 *The tier sizes* $\langle q_1, \dots, q_{H-1}\rangle$ *satisfy the first-order conditions for maximizing profit for fixed* H *and* q_0 *if and only if the spans* $\langle s_1, \dots, s_H\rangle$ *satisfy* $s_1 \cdots s_H = q_0$ *and*

$$s_h \;=\; 1 - \frac{1}{\log b} + \frac{b^{d_{h+1}}}{\log b} - \frac{W}{b^{L_h}\log b} \qquad \text{for } h = 1, \dots, H-1 \tag{5.3}$$

(recall that $W \triangleq w/\sigma^2$*).*

PROOF. See Appendix A. □

One conclusion we can derive from these first-order conditions is that the spans are single-peaked.

Corollary 5.1 *The spans $\langle s_1, \ldots, s_H \rangle$ that solve equation (5.3) and $s_1 \cdots s_H = q_0$ are single-peaked. That is, for $h = 2, \ldots, H-1$, if $s_{h-1} \geq s_h$ then $s_h > s_{h+1}$.*

PROOF. Let $h \in \{2, \ldots, H-1\}$. We prove the contrapositive: if $s_h \leq s_{h+1}$, then $s_{h-1} < s_h$. Compare s_{h-1} and s_h as defined by equation (5.3). It is always true that $L_{h-1} < L_h$; therefore, since $\log b < 0$, the final term in equation (5.3) is strictly smaller for s_{h-1} than for s_h. If $s_h \leq s_{h+1}$, then also $d_h \leq d_{h+1}$ and the second-to-last term in equation (5.3) is weakly smaller for s_{h-1} than for s_h. Therefore, $s_{h-1} < s_h$. □

Next we consider the hierarchies that maximize per-shop profit for fixed H but endogenous q_0. If H were endogenous, then such a maximization problem would have a solution if and only if $w > 0$. (We showed in Section 4 that there is a bound on firm size if and only if $w > 0$.) However, for fixed H, there is a bound on firm size even if $w = 0$ because fixing H constrains the internal decentralization of decision making.

Proposition 5.3 *A solution to the first-order conditions of equation (5.2) corresponds to a unique global maximum.*

PROOF. See Appendix A. The proof uses a change of variables to obtain a strictly-concave objective function. The strict concavity is demonstrated using a combination of analytic and numerical results. □

Proposition 5.4 *For fixed H, $\langle q_0, q_1, \ldots, q_{H-1} \rangle$ satisfy the first-order conditions for maximizing per-shop profit if and only if the spans $\langle s_1, \ldots, s_H \rangle$ satisfy equation (5.3) and*

$$s_H \;=\; 1 - \frac{1}{\log b} - \frac{W}{b^{L_H} \log b}\,. \tag{5.4}$$

PROOF. See Appendix A. □

When $w = 0$, equations (5.4) and (5.3) provide a recursive formula for the unique solution to the first-order conditions, and this solution does not depend on σ^2 or H. That is, from the top down, the optimal CF hierarchies look alike, whatever is the fixed height of the hierarchies. When $w > 0$, equations (5.4) and (5.3) do not give a recursive formula, because b^{L_h} is a function of $s_1, \ldots, s_{h-1}$. However, these equations still provide a useful characterization of the optimal span which aids in their numerical calculation and which can be used to derive further qualitative results. For example, the next proposition states that the spans of an optimal hierarchy decrease from upper to lower tiers.

Proposition 5.5 *Let $s_1, \ldots, s_H$ be the solution to equations (5.4) and (5.3), for $W < b^\alpha$. Then $s_h > s_{h-1}$ for $h = 2, \ldots, H-1$. If $w = 0$, or if $w > 0$ and H maximizes per-shop profit, then also $s_H > s_H - 1$.*

PROOF. See Appendix A. □

6 Comparative statics

When $w = 0$, the optimal hierarchies depend only on the parameter b. When $w > 0$, they also depend on the ratio $W \triangleq w/\sigma^2$. In this section, we characterize how the optimal hierarchies depend on these parameters. We are particularly interested in how organizational size and structure depend on b, which inversely measures the speed at which the environment changes.

Consider first optimal firm size. From Proposition 4.2, $\max\{1, W^{1/\log b}\}$ is a bound on firm size. If $W \geq 1$, then the optimal firm is 1: there is no information processing. Therefore, we restrict attention throughout this section to $0 \leq W < 1$. Then, since $1 < \log b < 0$, this bound is decreasing in W and increasing in b. Furthermore, as $b \uparrow 1$, the bound increases to infinity.

To actually show that firm size increases as $W \downarrow 0$ or $b \uparrow 1$, we need to derive a lower bound on firm size. Let $A\Pi(n)$ be the maximum per-shop profit for hierarchies with n shops. It follows from the existence of limits on firm size that

$$N(b, W) \triangleq \min\left\{n \in \mathbb{N} \;\middle|\; A\Pi(n) \geq A\Pi(n') \;\forall n' \in \mathbb{N}\right\}$$

exist when $w > 0$. The following proposition states in what sense $N(b, W)$ is a lower bound on the size of hierarchies in a forest.

Proposition 6.1 *There is an upper bound, which is independent of the total number of shops and applies to all optimal forests, on the number of shops in hierarchies smaller than $N(b, W)$.*

PROOF. See Appendix B. □

We can derive a lower bound on $N(b, W)$ with the following observation. It cannot be possible to raise the profit by combining under a new root several hierarchies that maximize the per-shop profit.

Proposition 6.2 *As either $b \uparrow 1$ or $W \downarrow 0$, $N(b, W) \to \infty$.*

PROOF. See Appendix B. □

We thus obtain the following comparative statistics on limits to firm size: For any positive wage, there is a limit to firm size, and when the wage-variance ratio is large enough or the environment changes quickly enough, the limit to firm size is 1. However, as the wage falls ($w \downarrow 0$) or the environment changes more slowly ($b \uparrow 1$), optimal firm size increases without bound.

Next we consider comparative statics on the spans for fixed H and endogenous n.

Proposition 6.3 *Let $H \in \mathbb{N}$. The spans $\langle s_1, \dots, s_H \rangle$ that maximize the per-shop profit are increasing in W and in b.*

PROOF. See Appendix B □

That the spans are increasing in b is as expected and is consistent with the conclusion that, asymptotically as $b \uparrow 1$, firm size is increasing in b. However, it may seem surprising that the spans, and hence firm size, are increasing in W even though we concluded that, asymptotically as $W \downarrow 0$, firm size is decreasing in W. This is a consequence of restricting the height of the hierarchy in Proposition 6.3. Merging hierarchies of height H by "firing" the root of all but one, which becomes the root of the merged hierarchy with height H, economizes on managerial costs.

So does dividing a hierarchy of height H by "firing" the root and thereby creating smaller hierarchies of height $H-1$. It is through this process that the hierarchies become smaller as W increases. Overall, we conjecture that, as W rises, over certain ranges the height of the optimal hierarchies remains fixed and the spans and size increase. Then, at certain thresholds, the height falls by 1 and the size falls as well.

7 Technological change

When we drop the normalization that $\mu = 1$, the parameter b is replaced by b^μ and the parameter w becomes $w\mu$.

Suppose that μ falls, meaning that managers become more productive, but the managerial wage stays constant. Then, with regards to our previous comparative statics result, the effect is like an increase in b and a decrease in w. Both factors lead to larger and more centralized firms. If the wage increases so that $w\mu$ stays constant, then the effect is identical to an increase in b, and once again the firms become larger and more centralized.

This raises interesting empirical questions. On the one hand, information processing has become quicker (a decrease in μ). However, as a consequence, each firm finds that its strategic environment is changing more rapidly (a decrease in b). In our model, these two trends have opposite effects on organizations.

A Proofs of results on spans of balanced hierarchies

PROOF OF PROPOSITION 5.2. Fix H. From equation (2.2),

$$\Pi_H(\boldsymbol{q}) \;=\; \sigma^2 \sum_{h=1}^{H} (q_{h-1} - q_h) b^{L_h} - w \sum_{h=0}^{H-1} q_h \,.$$

Let $h \in \{1, \dots, H-1\}$. Since $L_h = \alpha h + \log(q_0/q_h)$, $\partial L_h/\partial q_h = -1/q_h$ and

$$\frac{\partial b^{L_h}}{\partial q_h} \;=\; -b^{L_h}(\log b)/q_h \,.$$

If $\eta \neq h$, then $\partial L_\eta/\partial q_h = 0$. Therefore,

$$\frac{\partial \Pi_H}{\partial q_h} \;=\; \sigma^2(b^{L_{h+1}} - b^{L_h}) - \sigma^2(q_{h-1} - q_h) q_h^{-1} b^{L_h}(\log b) - w \,.$$

Since $L_{h+1} = d_{h+1} + L_h$ and $q_{h-1}/q_h = s_h$,

$$\frac{\partial \Pi_H}{\partial q_h} \;=\; -\sigma^2 b^{L_h}(1 - b^{d_{h+1}}) - \sigma^2 b^{L_h}(s_h - 1)(\log b) - w \,.$$

Dividing the first-order condition $\partial \Pi_H/\partial q_h = 0$ by $\sigma^2 b^{L_h}(\log b)$ and solving for s_h yields

$$s_h \;=\; 1 - \frac{1}{\log b} + \frac{b^{d_h+1}}{\log b} - \frac{W}{b^{L_h}\log b} \,.$$

We have changed the first-order conditions from a set of $H-1$ equations involving the variables $\langle q_1, \dots, q_{H-1}\rangle$ to a set of $H-1$ equations involving $\langle s_1, \dots, s_H\rangle$. The additional variable is fixed by the constraint that $s_1 \cdots s_H = q_0$. □

PROOF OF PROPOSITION 5.3. The objective function in the maximization problem is $\bar{\Pi}_H(\boldsymbol{q}) \triangleq (1/q_0)\Pi_H(\boldsymbol{q})$. We show, using a mix of analytic results and numerical tests, that a transformation of the choice variables leads to a strictly concave function.

Specifically, we define the invertible function $\langle q_0, \dots, q_{H-1}\rangle \underset{R_H}{\mapsto} \langle r_1, \dots, r_H\rangle$ by $r_h = q_h/q_0$ for $h = 1, \dots, H$. (In certain formulae involving such a vector $\boldsymbol{r} = \langle r_1, \dots, r_H\rangle$, we define $r_0 \triangleq 1$.) The Jacobian matrix $DR_H(\boldsymbol{q})$ is non-singular for all $\boldsymbol{q}$ since

$$DR_H(\boldsymbol{q}) \;=\; \begin{pmatrix} -q_1/q_0^2 & 1/q_0 & 0 & \cdots & 0 \\ -q_2/q_0^2 & 0 & 1/q_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -q_{H-1}/q_0^2 & 0 & 0 & \cdots & 1/q_0 \\ -1/q_0^2 & 0 & 0 & \cdots & 0 \end{pmatrix} .$$

We define below a function $\hat{\Pi}_H(\boldsymbol{r})$ such that $\bar{\Pi}_H(\boldsymbol{q}) = \hat{\Pi}_H \circ R_H(\boldsymbol{q})$. Then let $\boldsymbol{q}$ be a solution to $D\bar{\Pi}_H(\boldsymbol{q}) = \boldsymbol{0}$. Since $D\bar{\Pi}_H(\boldsymbol{q}) = D\hat{\Pi}_H(R_H(\boldsymbol{q})) \times DR_H(\boldsymbol{q})$ and $DR_H(\boldsymbol{q})$ is non-singular,

$D\hat{\Pi}_H(R_H(\boldsymbol{q})) = \mathbf{0}$. We show that $\bar{\Pi}_H$ is strictly concave, which implies that a solution $\boldsymbol{r}^*$ to $D\hat{\Pi}_H(\boldsymbol{r}) = \mathbf{0}$ is the unique global maximizer. Since R_H is invertible, $\boldsymbol{q}^* = R_H^{-1}(\boldsymbol{r}^*)$ is the unique global maximizer of $\bar{\Pi}_H$.

We are left with defining $\hat{\Pi}_H$ and showing that it is concave. We can write

$$\text{(A.1)} \qquad \bar{\Pi}(q_0, q_1, \ldots, q_{H-1}) \;=\; \sigma^2 \sum_{h=1}^{H} \left(\frac{q_{h-1}}{q_0} - \frac{q_h}{q_0} \right) b^{\alpha h + \log q_0/q_h} - w \left(1 + \sum_{h=1}^{H-1} q_h/q_0 \right).$$

Define $\hat{g}(s) \triangleq b^{-\log s}$ and $a \triangleq b^{\alpha}$. Then

$$\bar{\Pi}(q_0, q_1, \ldots, q_{H-1}) \;=\; \sigma_2 \sum_{h=1}^{H} a^h (r_{h-1} - r_h)\hat{g}(r_h) - w - w \sum_{h=1}^{H-1} r_h \;=:\; \hat{\Pi}_H(r_1, \ldots, r_H).$$

$\hat{\Pi}_H$ is concave if and only if

$$G_H(\boldsymbol{r}) \triangleq \sum_{h=1}^{H} a^h (r_{h-1} - r_h)\hat{g}(r_h)$$

is concave.

The proof of Proposition 4.4 in Van Zandt (2002a) shows that G_H is concave in $r_1, \ldots, r_{H-1}$. (The proof uses the restriction that $r_{h-1}/r_h \geq 2$, which holds here since $r_{h-1}/r_h = q_{h-1}/q_h$.) Although the proof relies partly on a numerical test, its analytic parts imply that

$$\frac{\partial^2 G_H}{\partial r_h^2} \;<\; 0 \qquad \text{and} \qquad \frac{\partial^2 G_H}{\partial r_h^2}\frac{\partial^2 G_H}{\partial r_{h'}^2} - \left(\frac{\partial^2 G_H}{\partial r_h \partial r_{h'}} \right)^2 \;>\; 0$$

for $h, h' \in \{1, \ldots, H-1\}$ such that $h \neq h'$. One can easily check that, although such a result was not needed in that proof, the derivation also applies when $h = H$.

These are necessary but not sufficient conditions for the Hessian matrix of G_H to be negative definite. We also tested and confirmed the strict concavity condition

$$G_H\left(\lambda \boldsymbol{r} + (1-\lambda)\boldsymbol{r}'\right) \;>\; \lambda G_H(\boldsymbol{r}) + (1-\lambda) G_H(\boldsymbol{r}')$$

in 10^8 trials, with parameters and variables chosen randomly as follows (in each case, selection is with uniform distribution on indicated range). (i) $\alpha \in (0, 10)$, (ii) $b \in (1/2, 1)$, $H \in \{2, \ldots, 24\}$, (vi) $\lambda \in (0, 1)$, (vii) $r_H, r'_H \in (0, 1/2^H)$, (viii) for $h \in \{1, \ldots, H-1\}$, given r_{h+1} and r'_{h+1}, $r_h \in (2r_{h+1}, 1/2^h)$ and $r'_h \in (2r'_{h+1}, 1/2^h)$. These ranges reflect a lower bound on q_0 of 2^H and the restriction that spans be at least 2. □

Proof of Proposition 5.4. A hierarchy that maximizes per-shop profit for fixed H but endogenous q_0 must satisfy the first-order conditions (equation (5.3)) for maximizing total profit for fixed H and q_0. The span s_H of the root must then also solve the first-order condition for maximizing per-shop profit when the remaining spans are fixed but q_0 is endogenous.

Note that

$$\frac{q_h}{q_0} = \frac{s_{h+1} \cdots s_H}{s_1 \cdots s_H} = \frac{1}{s_1 \cdots s_h} .$$

Therefore, from equation (A.1), we can write the per-shop profit as

$$\text{(A.2)} \qquad \sigma^2 \sum_{h=1}^{H} \left(\frac{1}{s_1 \cdots s_{h-1}} - \frac{1}{s_1 \cdots s_h} \right) b^{\alpha h + \sum_{\eta=1}^{h} \log s_\eta} - w \left(\sum_{h=1}^{H} \frac{1}{s_1 \cdots s_h} \right) .$$

Then the derivative of equation (A.2) with respect to s_H, set equal to 0, yields (substituting $q_H = 1$ and $q_{H-1} = s_H$)

$$\begin{aligned} \sigma^2 (q_0 s_H)^{-1} b^{L_H} + \sigma^2 \left(\frac{s_H}{q_0} - \frac{1}{q_0} \right) b^{L_H} (\log b) s_H^{-1} + w (q_0 s_H)^{-1} &= 0 \\ 1 + (s_H - 1) \log b + \frac{W}{b^{L_H}} &= 0 \\ 1 - \frac{1}{\log b} - \frac{W}{b^{L_H} \log b} &= s_H . \end{aligned}$$

□

Proof of Proposition 5.5. According to Corollary 5.1, if $s_h < s_{h+1}$, then $s_{h-1} < s_h$ (for $h = 2, \ldots, H-1$). Therefore, we only have to show (i) that $s_{H-2} < s_{H-1}$ and (ii) that $s_{H-1} < s_H$ if $w = 0$ or if $w > 0$ and H maximizes the per-shop profit.

First we show that $s_{H-2} < s_{H-1}$. If $s_{H-1} \leq s_H$, then $s_{H-2} < s_{H-1}$ follows from Corollary 5.1. Suppose instead that $s_{H-1} > s_H$. Since $s_H = 1 - (\log b)^{-1}$,

$$\text{(A.3)} \qquad 1 + (\log b)^{-1} < s_{H-1}$$

$$\text{(A.4)} \qquad 1 + (s_{H-1} - 1) \log b < 0 .$$

Since $L_{H-2} = L_{H-1} - d_{H-1}$, $b^{L_{H-2}} = b^{L_{H-1}} / b^{d_{H-1}}$. From equation (5.3) for $h = H-2$,

$$\text{(A.5)} \qquad s_{H-2} \log b = \log b - 1 + b^{d_{H-1}} - \frac{b^{d_{H-1}} W}{b^{L_{H-1}}}$$

$$\text{(A.6)} \qquad (s_{H-2} - 1) \log b + 1 - d^{H-1} = -b^{d_{H-1}} \frac{W}{b^{L_{H-1}}} .$$

From equation (5.3) for $h = H-1$,

$$\text{(A.7)} \qquad (s_{H-1} - 1) \log b + 1 - b^{d_H} = \frac{W}{b^{L_{H-1}}} .$$

Substituting equation (A.7) into equation (A.6) and rearranging, we obtain

$$\text{(A.8)} \qquad (s_{H-2} - 1) \log b + 1 - b^{d_{H-1}} = \left(1 + (s_{H-1} - 1) \log b + 1 - b^{d_H} \right) b^{d_{H-1}} .$$

Since $b^{d_{H-1}} > b^{d_H} b^{d_{H-1}}$, equation (A.8) implies

$$1 + (s_{H-2} - 1)\log b \; > \; (1 + (s_{H-1} - 1)\log b)\, b^{d_{H-1}} \,. \tag{A.9}$$

Given equation (A.4) and given $0 < b^{d_{H-1}} < 1$,

$$(1 + (s_{H-1} - 1)\log b)\, b^{d_{H-1}} \; \geq \; 1 + (s_{H-1} - 1)\log b \,.$$

This and equation (A.9) yields

$$\begin{aligned} 1 + (s_{H-2} - 1)\log b \; &> \; 1 + (s_{H-1} - 1)\log b \\ s_{H-2} \; &< \; s_{H-1} \,. \end{aligned}$$

That $s_H > s_{H-1}$ if $w = 0$ is trivial, since then $s_H = 1 - 1/\log b$, and $s_{H-1} = s_H + b^{d_H}/\log b < s_H$.

Finally, we show that $s_H > s_{H-1}$ if $w > 0$ and H maximizes per-shop profit, i.e., for hierarchies that maximize per-shop profit without restrictions. From equations (5.4) and (5.3), $s_H > s_{H-1}$ if and only if

$$-\frac{W}{b^{L_H}\log b} \; > \; \frac{b^{d_H}}{\log b} - \frac{W}{b^{L_{H-1}}\log b} \tag{A.10}$$

Multiply equation (A.10) by $b^{L_H}\log b < 0$, recalling that $b^{L_H}/b^{L_{H-1}} = b^{d_H}$, to obtain

$$(b^{d_H} - 1)W \; < \; b^{d_H} b^{L_H} \,. \tag{A.11}$$

If $b^{d_H} - 1 \leq 0$, then equation (A.11) is satisfied. Otherwise, equation (A.11) can be rewritten

$$W \; < \; \frac{b^{d_H} b^{L_H}}{b^{d_H} - 1} \,. \tag{A.12}$$

If the hierarchy maximizes per-shop profit without restrictions on the height, then $V_R^{\mathrm{net}} \geq 0$. From equation (4.1), this can be written

$$w s_H \; \leq \; \sigma^2 (s_H - 1) b^{L_H} \tag{A.13}$$

$$W \; < \; \frac{(s_H - 1) b^{L_H}}{s_H} \tag{A.14}$$

Hence, $s_H > s_{H-1}$ if the right-hand side of equation (A.14) is less than or equal to the right-hand side of equation (A.12), which we verify as follows (assuming still $b^{d_H} - 1 > 0$).

$$\begin{aligned} \frac{(s_H - 1) b^{L_H}}{s_H} \; &\overset{?}{\leq} \; \frac{b^{d_H} b^{L_H}}{b^{d_H} - 1} \,, \\ s_H b^{d_H} - b^{d_H} - s_H + 1 \; &\overset{?}{\leq} \; s_H b^{d_H} \,, \\ -b^{d_H} \; &\leq \; s_H - 1 \,. \end{aligned}$$

□

B Proofs of results on comparative statics

PROOF OF PROPOSITION 6.1. Fix b and W. Let $n^* \triangleq N(b,W)$ and let $\pi^* \triangleq A\Pi(n^*)$ be the maximum per-shop profit. Let $\pi^{\max}$ and $\pi^{\min}$ be the maximum and minimum values, respectively, of $\{A\Pi(n) \mid n < n^*\}$.

Suppose that, in an optimal forest, there are n shops in hierarchies that are smaller than n^*. The total profit of these hierarchies is at most $n\pi^{\max}$. If these shops were instead organized into hierarchies of size n^*, there would be at most $n^* - 1$ leftover shops in a smaller hierarchy and so the total profit would be at least $(n-n^*)\pi^* + n^*\pi^{\min}$. This must be no higher than $n\pi^{\max}$ since the original forest is optimal. Therefore,

$$\begin{aligned}(n-n^*)\pi^* + n^*\pi^{\min} &\leq n\pi^{\max}\\ n(\pi^* - \pi^{\max}) &\leq n^*(\pi^* - \pi^{\min})\\ n &\leq n^* \frac{\pi^* - \pi^{\min}}{\pi^* - \pi^{\max}}\,.\end{aligned}$$

□

PROOF OF PROPOSITION 6.2. Fix b and W. Let $n \triangleq N(b,W)$ and H be the size and height, respectively, of the smallest hierarchy that maximizes the per-shop profit. The change in profit by combining s_R such hierarchies under a new root equals the net value of the new root:

$$V_R^{\text{net}} = \sigma^2(s_R - 1)b^{\alpha(H+1)+\log(s_R n)} - ws_R\,.$$

Since the span of each tier is at least 2, $H \leq \log_2 n$. Therefore,

$$\begin{aligned}V_R^{\text{net}} &\geq \sigma^2(s_R-1)b^{\alpha(1+\log_2 n)+\log s_R n} - ws_R\,,\\ &= \sigma^2(s_R-1)b^{\alpha} n^{(1+\alpha/\log 2)\log b} s_R^{\log b} - ws_R\,.\end{aligned}$$

Since the profit cannot be increased by combining such hierarchies, $V_R^{\text{net}} \leq 0$ and hence

$$n^{(1+\alpha/\log 2)\log b} \leq \frac{s_R}{s_R-1} s_R^{-\log b}\frac{W}{b^\alpha}\,, \tag{B.1}$$

$$n \geq \left(\frac{s_R}{s_R-1}\frac{W}{b^\alpha}\right)^{\frac{1}{(1+\alpha/\log 2)\log b}} s_R^{\frac{-1}{\alpha+1}}\,. \tag{B.2}$$

This bound must hold for all $s_R \geq 2$. Observe first that, since $\log b < 0$, the r.h.s. of equation (B.2) converges to ∞ as $W \downarrow 0$ for any fixed s_R. Consider now the comparative statics with respect to b. Since $W < 1$, we can choose s_R such that $(s_R/(s_R-1))W < 1$. As $b \uparrow 1$, $(s_R/(s_R-1))(W/b^\alpha)$ converges to $(s_R/(s_R-1))$ and the exponent $1/((1+\alpha/\log 2)\log b)$ converges to $-\infty$. Therefore, the r.h.s. of equation (B.2) converges to ∞. □

PROOF OF PROPOSITION 6.3. The comparative statics with respect to b when $w = 0$ are the easiest to see. From equation (5.3), s_h is increasing in s_{h+1} and in b, and hence s_h is increasing in b if s_{h+1} is increasing in b. The induction is started by noting, from equation (5.4), that s_H is increasing in b.

The general case is almost as easy. Denote $\langle s_1, \ldots, s_H \rangle$ by s. We can write equations (5.3) and (5.4) as $s_h = f_h(s; b, W)$ for $h = 1, \ldots, H$. Let $f \triangleq \langle f_1, \ldots, f_H \rangle$. Each f_h (and hence f) is increasing in s, in b, and in W. Let s^0 be the spans given b and W, that is, $s^0 = f(s^0; b, W)$. Let $b' \geq b$ and $W' \geq W$, with at least one strict inequality. Define $\{s^1, s^2, \ldots\}$ by $s^t = f(s^{t-1}; b', W')$. Since f is increasing in b and W, $s^1 > s^0$. Since f is increasing in s, $s^t > s^{t-1}$ for $t \geq 2$. The monotone sequence is bounded above by the unique solution to equations (5.3) and (5.4) when the $b^{d_{h+1}}$ terms are suppressed (this solution is recursively defined starting with s_1). Therefore, it converges to a solution $s' = f(s'; b', W')$ and $s' > s^0$. □

References

Geanakoplos, J. and Milgrom, P. (1991). A theory of hierarchies based on limited managerial attention. *Journal of the Japanese and International Economies*, 5, 205–225.

Radner, R. (1993). The organization of decentralized information processing. *Econometrica*, 62, 1109–1146.

Van Zandt, T. (1998). Organizations with an endogenous number of information processing agents. In M. Majumdar (Ed.), *Organizations with Incomplete Information*. Cambridge: Cambridge University Press.

Van Zandt, T. (2002a). Balancedness of real-time hierarchical resource allocation. INSEAD.

Van Zandt, T. (2002b). Real-time hierarchical resource allocation with quadratic costs. INSEAD.

Van Zandt, T. and Radner, R. (2001). Real-time decentralized information processing and returns to scale. *Economic Theory*, 17, 497–544.