RAND Journal of Economics
Vol. 26, No. 4, Winter 1995
pp. 575–590

Continuous approximations in the study of hierarchies

Timothy Van Zandt*

Large organizations are typically modelled as hierarchies. Hierarchies are discrete structures (trees), but researchers frequently use continuous approximations. The purpose of this article is to study the validity of these approximations. I show that modelling hierarchies with a continuum of tiers is not a good approximation. I also show, for a particular model of balanced hierarchies, that ignoring rounding operators and integer constraints in formulae derived from the discrete model can be a valid approximation, when hierarchies are suitably large. This is made precise by bounds on the relative errors of the approximations.

1. Introduction

■ Large organizations are typically modelled as hierarchies. Although internal contracting and information processing in organizations such as firms are not as hierarchical as their organizational charts might indicate (Mintzberg, 1978; Aoki, 1992), hierarchies have the advantage of being simpler than general graphs. Balanced hierarchies, in which managers in the same tier have the same span of control, are particularly simple because they are described by a small number of parameters.

There are many hierarchical models in which authors have characterized, for example, the optimal span of control and the distribution of wages in large hierarchies and the returns to scale of hierarchical organizations. Examples are Beckmann (1960, 1977, 1983), Williamson (1967), Calvo and Wellisz (1978), Keren and Levhari (1979, 1983, 1989), Geanakoplos and Milgrom (1991), Radner (1993), Van Zandt (1994a), Qian (1994), Wang (1993), and Bolton and Dewatripont (1994). Hierarchies (trees) are discrete structures, but these articles sometimes ignore rounding operators or integer constraints in various formulae derived from discrete models. Keren and Levhari (1979, 1983, 1989), Qian (1994), and Wang (1993) have also used a formulation in which the hierarchies have a continuum of tiers. The purpose of this article is to explore the validity of these continuous approximations. I show that the continuous-tier formulation

* Princeton University.

The research assistance and comments of Archishman Chakraborty, consultations with Edward Reingold on regularity properties of trees, and the comments of two anonymous referees have been very helpful. This article was written while the author was visiting the Center for Operations Research and Econometrics at the Université Catholique de Louvain, under a CORE Research Fellowship (1993–1994). Their financial and intellectual support is gratefully acknowledged. This research has also been supported in part by grant no. SBR-9223917 from the National Science Foundation and grant no. 26 of the "Pôle d'Attraction Interuniversitaire" program of the Belgian government.

is not a good approximation. I also show that in the model of Keren and Levhari (1979, 1983), ignoring rounding operators and integer constraints in formulae derived from the discrete model can be a valid approximation, when hierarchies are suitably large.

In temporal models, it is typically possible to relate a discrete-time model to its continuous-time analog by parameterizing the discrete-time model by the length of the period and showing that the properties of the model converge to those of the continuous-time model as the length of the period converges to zero. In a finite-horizon model, decreasing the length of the period increases the number of periods, but it is not equivalent to extending the horizon. Instead, other parameters such as the per-period growth or discount rate are tied to the length of the period so that the growth or discount rate for a fixed horizon remains constant. For example, Brownian motion is the limit of a discrete-time random walk as the length of the period converges to zero while the per-period variance of the random walk is adjusted so that the variance of the process over a fixed horizon is constant.

It is not possible to treat the tiers of a hierarchy in the same way. The "distance" between tiers cannot be reduced by adjusting parameters of the model or by increasing the number of tiers in the hierarchy. The analog of the growth or discount factor in hierarchies is the span of control; this determines the rate of growth of the number of managers in the tiers. Unlike the per-period growth or discount factor that can converge to one in a temporal model as the length of the period converges to zero, the span of control is endogenous and integer-valued, and must be at least two in nearly all models of hierarchies. This leads to errors in the measure of the number of managers that is significant for the theoretical minimum span of two and that increases as the span rises to empirically relevant levels. Furthermore, these errors distort the optimization and the characterization of the optimal hierarchies.

The continuous-tier formulation is used in Keren and Levhari (1979, 1983, 1989), Qian (1994), and Wang (1993) to derive quantitative formulae characterizing the optimal hierarchies. Although such formulae would be difficult to derive in a discrete-tier model, a qualitative characterization is possible using the discrete-tier model (see Section 6). A sound qualitative characterization is better than an inaccurate quantitative characterization, and the discrete model has the advantage that the proofs are typically constructive and intuitive.

Nevertheless, the mathematical simplicity of balanced hierarchies (as opposed to general hierarchies) is enhanced by ignoring integer constraints on the span of control and on the number of managers in each tier. Note that such approximations are not always justified simply because it is possible to hire part-time managers. In hierarchical models of associative computation, for example, a manager generates a message to be processed by a superior whether the manager works full time or part time. But because these approximations involve dropping integer constraints in formulae derived from the discrete model, it is to be expected that they are valid when the variables are large. This I show for the model in Keren and Levhari (1979, 1983), and I give bounds on the relative errors of the approximations. It is impossible to show that such approximations are valid for any hierarchical model one can construct,[1] but I have chosen to study the Keren and Levhari model because it is based on the measurement of the total number of managers in a hierarchy and the sum of the spans of the tiers. These are quantities that arise in a variety of hierarchical models, such as Williamson (1967), Calvo and Wellisz (1978), and Geanakoplos and Milgrom (1991).

[1] In particular, these approximations are for balanced hierarchies. Radner (1993), Van Zandt (1994a), and Bolton and Dewatripont (1994) do not restrict attention to balanced hierarchies and do not make extensive use of continuous approximations.

When I say that an approximation is "valid," I mean the following: (i) When there are many operatives, the hierarchies that are optimal for altered formulae are approximately optimal for the true model, or there are integer hierarchies near noninteger optima that are approximately optimal. That is, the relative inefficiency of the maximizers (the ratio of the added cost to the minimum cost) converges to zero as the number of operatives increases. (ii) When there are many operatives, the measures of delay, managers, or minimized cost that are based on altered formulae, or are derived ignoring integer constraints, approximate the true values. That is, the relative error of the objective or value function (the ratio of the error to the true value) converges to zero as the number of operatives increases.

These positive results should be used with caution. One reason is that "validity" of an approximation, as defined above, does not imply that the solution to the approximate problem is close to the solution to the true problem. For example, even if an objective function of several real variables is globally convex, the integer-constrained maximizer can be arbitrarily far from the unconstrained maximizer. As is often the case when an objective function is replaced by an approximation in economic models, we must be satisfied with a characterization of approximately optimal hierarchies, as opposed to an approximate characterization of optimal hierarchies.

Furthermore, most of the approximations improve only if the span increases together with the number of operatives. However, hierarchies can grow by adding more tiers, rather than increasing the span. Spans of control in organizations are typically in the range of eight to fifteen, and in theoretical models it is common to also study hierarchies in which the span is two or three, for which the integer constraints do matter. Nevertheless, it is at least reassuring that the models with and without these approximations converge for some sizes of hierarchies, which is not true for the continuous-tier and discrete-tier models.

The article is organized as follows. In Section 2 I fix notation and terminology for hierarchies. The positive results on ignoring rounding operations and integer constraints in discrete-tier models are given in Sections 3 and 4. These are used in Section 5, where I present the negative results for the continuous-tier formulation. Section 6 concludes with comments on the usefulness of approximations in the study of hierarchies. All proofs are contained in the Appendix.

2. Hierarchies

■ Mathematically, a hierarchy is a tree. This is the only aspect of hierarchies that interests us here. However, rather than using the standard terminology of trees from computer science and discrete mathematics, I use terminology from organizations, as follows: the leaves or terminal nodes are called operatives, the internal nodes are called managers, children or immediate successors are called direct subordinates, parents or immediate predecessors are called supervisors or direct superiors, a manager's span (of control) is the number of her direct subordinates, a node's level or tier is the number of edges between the node and the root, and the maximum tier is called the height of the hierarchy. I restrict attention, once and for all, to hierarchies in which all operatives are in the same tier. This is tier H, where H is the height of the hierarchy. (The lower tiers have a higher number.) Let n be the number of operatives, which is typically given exogenously. The problem is to design the intervening hierarchy between the root and the operatives.

Let q_h be the number of managers in tier h. Tier 0, called the top tier, contains only the root ($q_0 = 1$). Tier H, called the bottom tier, contains only the n operative ($q_H = n$). For each managerial tier h, the span of tier h is $s_h \equiv q_{h+1}/q_h$. Since $q_0 = 1$,

$$q_h = s_0 \times \ldots \times s_{h-1} \tag{1}$$

for $h = 1, \ldots, H$. The maximum span of tier h, denoted by S_h, is the maximum span of the managers in tier h. This is always at least $\lceil s_h \rceil$.[2]

A hierarchy is completely balanced if all managers in the same tier have the same span. A hierarchy is said to be completely uniform if all managers have the same span. Note that if s is the common span in a completely uniform hierarchy, then $q_h = s^h$ and $n = s^H$.

These definitions of completely balanced and completely uniform hierarchies are restrictive. For example, unless n is a power of some integer, there is no completely uniform hierarchy with n operatives. Unless n is the product of H integers greater than one, there is no completely balanced hierarchy with height H and n operatives, except for those in which the span of some tiers is one. Therefore, I adopt the following generalizations:

Definition 1. A hierarchy is balanced if the maximum span of the managers in tier h is $\lceil s_h \rceil$. A hierarchy is uniform if it is balanced and if the maximum spans of the tiers differ by at most one.

In a uniform hierarchy of height H, the span of each tier is either $\lfloor n^{1/H} \rfloor$ or $\lceil n^{1/H} \rceil$, and $n^{1/H}$ is called the span of the hierarchy.

Note that there are other ways to generalize the definitions of completely balanced and completely uniform hierarchies. In an informal sense, "balanced" means that for each manager, the subtree of each direct subordinate is roughly the same size. Also, a property of uniform hierarchies is that, for fixed span, the height of the hierarchy increases logarithmically in the number of operatives. In computer science, there are various definitions of balanced s-ary trees that preserve these two properties of uniform hierarchies but do not require all terminal nodes to be on the same level. For example, a depth-balanced s-ary tree is one in which all internal nodes except perhaps some in level $H - 1$ have s children and all leaves are in levels $H - 1$ and H. I have included the restriction that all operatives be in the bottom tier so that relationship (1) holds, hence the balanced hierarchies are parameterized by the number of managers in each tier or, equivalently, by the span of each tier.

3. Integer constraints in balanced hierarchies

■ In this section I determine when ignoring rounding operators and integer constraints on the numbers of managers and spans of control in balanced hierarchies is a valid approximation, in the discrete-tier model of Keren and Levhari (1979, 1983).[3]

The Keren and Levhari model can be interpreted as one of associative computation (see Van Zandt, 1994b). Hierarchies are assumed to be balanced. The computational delay for each manager is her span. The delay for tier h is the maximum delay S_h of the managers in tier h. The performance criteria (costs) are the number Q of managers and the total delay D, where

$$Q = \sum_{h=0}^{H-1} q_h \tag{2}$$

$$D = \sum_{h=0}^{H-1} \lceil q_{h+1}/q_h \rceil. \tag{3}$$

[2] $\lceil x \rceil$ is the ceiling of x, or x rounded up. $\lfloor x \rfloor$ is the floor of x, or x rounded down.

[3] See the end of the Introduction for the definitions of valid approximation, relative inefficiency, and relative error.

The number n of operatives, the managerial wage w, and the cost $c(D)$ of delay are given. The problem is to design a hierarchy $\mathcal{H}$ (i.e., choose H and $q_1, \ldots, q_{H-1}$) to minimize the managerial wages plus the cost of delay:

$$C(\mathcal{H}) = wQ(\mathcal{H}) + c(D(\mathcal{H})).$$

The cost $c(D)$ of delay is assumed to be positive, increasing, and concave.[4] $c(D)$ may depend on n, and I write $c(D, n)$ when this dependence is important. $c_D(D)$ or $c_D(D, n)$ denotes the partial derivative of c with respect to D (or the slope of a tangent line if c is not differentiable).

The results in this section assume that the span of every tier is at least two. One can show, without using continuous approximations, that this is a property of efficient hierarchies (see Van Zandt, 1994c). For a hierarchy with $H + 1$ tiers, the number of managers is then at least $2^H - 1$.

One approximation used by Keren and Levhari is to drop the ceiling operator in the formula for the delay, replacing D by D^a:

$$D^a = \sum_{h=0}^{H-1} q_{h+1}/q_h. \tag{4}$$

Keren and Levhari (1979) show that D^a is bounded below by $e \log n$. I make frequent use of this lower bound, and the lower bound $\hat{c}(n) \equiv c(e \log n, n)$ on the cost of delay and the upper bound $\hat{c}_D(n) = c_D(e \log n, n)$ on the slope of $c(D)$. Furthermore, I assume that $\lim_{n\to\infty} \hat{c}(n) = \infty$.

Rounding up adds less than one unit of delay per managerial tier. Therefore, $D - D^a < H$, and we have Proposition 1.

Proposition 1. For all hierarchies $\mathcal{H}$ with $H + 1$ tiers,

$$\frac{D(\mathcal{H}) - D^a(\mathcal{H})}{D(\mathcal{H})} < \frac{1}{\bar{s}},$$

where $\bar{s}$ is the average span $D^a(\mathcal{H})/H$ of the tiers in $\mathcal{H}$.

The average span does not necessarily increase with the number of operatives, since a hierarchy can grow with a constant span by increasing its height. In fact, this is a very weak result because in organizations, spans of control are typically in the range of 8 to 15. In Section 4 I prove a stronger result (Proposition 4) for the case of constant span.

Let $C^a(\mathcal{H}) = wQ(\mathcal{H}) + c(D^a(\mathcal{H}))$ be the modified measure of total cost using D^a. If the wage is positive, then either the average span or the marginal costs increase with the number of operatives. This causes the relative inefficiency and the relative error to diminish as the number of operatives increases, as shown in Proposition 2.

Proposition 2. Assume $w > 0$. Let C^* be the true minimum cost $\min C(\mathcal{H})$, and let $\mathcal{H}^*$ be a solution to $\min C^a(\mathcal{H})$. Then

[4] In applications, $c(D)$ may even be bounded. For example, in information processing models it is bounded by the cost when decisions are made based only on prior information. I use concavity in the proofs to put bounds on the effect that a given change in delay has on the cost of delay.

$$\underset{\text{(relative inefficiency)}}{\frac{C(\mathcal{H}^*) - C^*}{C^*}} \quad \text{and} \quad \underset{\text{(relative error)}}{\frac{C^* - C^a(\mathcal{H}^*)}{C^*}} \quad \text{are less than} \quad \frac{\hat{c}_D(n)\log_2 n}{2\sqrt{nw}} \tag{5}$$

and are less than the inverse of the average span of $\mathcal{H}^*$.

If $\hat{c}_D(n)$ is constant, then the rate of convergence is at least $O((\log n)/\sqrt{n})$.[5] For example, this is true if $c(D, n) = (1 - e^{-\lambda D})n$ (as in Keren and Levhari (1983)), which can hold when the cost of delay is due to a declining value of information.

A second approximation is to drop the integer constraints on the number of managers in each tier, in the problem with delay D^a. Proposition 3, which is stated and proved in the Appendix, indicates that for large hierarchies there is an integer hierarchy near the noninteger solution that is approximately optimal and the cost of the noninteger solution is a good approximation for the integer-constrained minimum cost. The rates of convergence of the relative inefficiency and relative error, with respect to n and H, are $O(\log(\hat{c}(n))/\hat{c}(n))$ and $O(1/H)$, respectively.

4. Integer constraints in uniform hierarchies

■ Let $D^u(H, n)$ be the minimum delay for hierarchies with height H and n operatives. Let $Q^u(s, n)$ be the smallest number of managers for hierarchies with maximum span s and n operatives.[6] Then $D^u(H, n)$ and $Q^u(s, n)$ are each attained by a uniform hierarchy.

If $s = n^{1/H}$ for integers s and H, then $D^u(H, n)$ and $Q^u(s, n)$ are obtained only by the completely uniform hierarchy with height H and span s, and they are equal to $D^{ua}(H, n) \equiv Hn^{1/H} = s \log_s n$ and $Q^{ua}(S, n) \equiv (n - 1)/(n^{1/H} - 1) = (n - 1)/(s - 1)$, respectively.

If H is an integer but $s = n^{1/H}$ is not, then the true minimum delay $D^u(H, n)$ is greater than $D^{ua}(H, n)$. Proposition 4 says that $D^{ua}(H, n)$ is still a good approximation of $D^u(H, n)$ when s is not too small.

Proposition 4. Let H and n be integers and let $s = n^{1/H}$. Then

$$0 \leq \frac{D^u(H, n) - D^{ua}(H, n)}{D^{ua}(H, n)} < \frac{1}{8s^2} + \frac{1}{\log n}\frac{\log s}{s}. \tag{6}$$

If s is an integer but $H = \log_s n$ is not, then $Q^u(s, n) > Q^{ua}(s, n)$. Proposition 5 says that $Q^{ua}(s, n)$ is still a good approximation of $Q^u(s, n)$ when n is large and $H \geq 2$. The relative error is $O(\log(n)/n)$ for bounded s and $O(n^{-(\bar{H}-1)/\bar{H}})$ for $H \geq \bar{H}$.

Proposition 5. Let n and s be integers. Then

$$0 \leq \frac{Q^u(s, n) - Q^{ua}(s, n)}{Q^{ua}(s, n)} < \frac{1}{n - 1} + \frac{s}{\log s}\frac{\log n}{n - 1}. \tag{7}$$

Allowing H to be noninteger in D^{ua} can be a useful way to find an optimal uniform hierarchy in a model such as Keren and Levhari (1979, 1983). If the total cost is

[5] $f(x) = O(g(x))$ means that $|f(x)| \leq k|g(x)|$ for some constant k.

[6] If computation problems arrive periodically and a hierarchy can be working on more than one problem at a time, then the throughput of a hierarchy (the rate at which the hierarchy can compute problems) is of interest, and this throughput is the inverse of the hierarchy's maximum span. $Q^u(s, n)$ is an important quantity in such a model.

globally convex in H, then the integer-constrained solution is one of the two neighboring integers of the noninteger solution, and one can check each of these integers to find the optimum. It can also be convenient to use the noninteger solution to calculate the approximate delay and number of managers of the optimal solution.

Compare a uniform "hierarchy" that has a noninteger height H with a hierarchy of height $\lceil H \rceil$. The relative difference in the height is less than $1/H$. The relative difference between the spans is less than

$$\frac{n^{1/H} - n^{1/(H+1)}}{n^{1/H}} \leq 1 - n^{1/H^2} = 1 - s^{-1/\log_s n}. \tag{8}$$

One can show that $1 - s^{-1/\log_s n}$ also bounds the relative error of the delay. Similarly, the relative error of the number of managers is bounded by

$$\frac{n^{1/H} - n^{1/(H+1)}}{n^{1/(H+1)} - 1} = \frac{s - s^{H/(H+1)}}{s^{H/(H+1)} - 1}. \tag{9}$$

Therefore, this approximation is good if s is bounded and n is large, but it may not be good if s increases with n.

5. A continuum of tiers

■ Keren and Levhari (1979, 1983, 1989), Qian (1994), and Wang (1993) use an additional, fundamentally different, continuous approximation. They replace the discrete number of tiers by a continuum of tiers, analogous to the shift from a discrete-time model to a continuous-time model.

The continuous-tier model requires the approximations discussed in Sections 3 and 4. To distinguish between the error due to these approximations and the additional error due to the continuum of tiers, I ignore the integer constraints on the number of managers and on the maximum span in this section when studying both the discrete-tier and the continuous-tier models. That is, a hierarchy is described by an integer height H and a real-valued number q_h of managers in tier $h \in \{1, \ldots, H - 1\}$, the number of managers is given by Q (equation 2)), and delay is given by D^a (equation (4)).

Keren and Levhari (1979, 1983) give the following translation into a continuous-tier model. From (1), we obtain

$$\log q_h = \sum_{\eta=0}^{h-1} \log s_\eta. \tag{10}$$

The continuous analog is

$$\log q(h) = \int_0^h \log s(\eta)\, d\eta. \tag{11}$$

This yields the equation of motion

$$\dot{q}(h) = q(h) \log s(h).$$

The continuous versions of (2) and (4) are

$$Q^c = \int_0^H q(h)\, dh \tag{12}$$

$$D^c = \int_0^H s(h)\, dh. \tag{13}$$

We would like to compare (12) and (13) with their corresponding discrete formulae. I focus on uniform hierarchies because the comparison for such hierarchies is straightforward. Nonuniform hierarchies are discussed briefly at the end of this section.

If the span of a uniform hierarchy is s, then $D^c = Hs$. This is the correct delay in the discrete-tier model. Furthermore, the formula (11) for the number $q(h)$ of managers in each integer tier is correct. However, the continuous-tier model overestimates the total number of managers in tier h by averaging the number of managers in fictitious intermediate tiers whose numbers range from q_h to q_{h+1}. The discrete-tier and continuous-tier formulae for the number of managers are, respectively,

$$Q^{ua} = \frac{n - 1}{s - 1} \tag{14}$$

$$Q^{uc} = \frac{s - 1}{\log s} \sum_{h=0}^{H-1} s^h = \frac{n - 1}{\log s}. \tag{15}$$

The relative error increases with s and is independent of n:

$$\frac{Q^{uc}}{Q^{ua}} = \frac{s - 1}{\log s}. \tag{16}$$

The formulae for Q^{ua} and Q^{uc} are mathematically well defined for any span greater than one, and their ratio converges to one as the span converges to one. However, this fact is unhelpful because neither the continuous-tier nor the discrete-tier model makes any sense when the span approaches one: (i) A manager's span is integer-valued and should be at least two in nearly all models of hierarchies. (If a manager has just one subordinate, it is better to fire that manager and have the subordinate report directly to the manager's superior. In the Keren and Levhari (1979, 1983) model as presented here, the lower bound is three for most managers.) (ii) The continuous approximations of the span of control upon which Q^{ua} and Q^{uc} are based are valid only when the span is large. (iii) Even if we ignore (i) and (ii), the optimal span is at least e and both the number of managers (whether measured by Q^{ua} or Q^{uc}) and the delay increase to infinity as the span converges to one, for a fixed number of operatives.

Hence, for meaningful values of s, the error is significant. Q^{uc}/Q^{ua} is nearly two when $s = 3$ and is more than four when $s = 10$.

Note that in the discrete-tier model, our measure Q of the number of managers does not include the operatives. It is ambiguous whether the same is true for Q^c in the continuous-tier formulation. Keren and Levhari adopt the convention that Q^c does not include the operatives, and for consistency I have so far done the same. As demonstrated above, Q^c then overestimates the number of managers. However, Qian (1994) and Wang (1993) adopt the convention that Q^c does include the operatives. In this case, Q^c significantly underestimates the size of the hierarchy. When the operatives are added to the definition of Q^{ua}, Q^{uc}/Q^{ua} is approximately equal to and always less than $(s - 1)/(s \log s)$. In a continuous-time model that is a good approximation of the discrete-time model, it does not matter how we treat the boundary. This is not true in hierarchies, because there are more operatives than managers.

For example, suppose $n = 46{,}656$ and $s = 6$. Then Q^c is equal to 26,039. Keren and Levhari would say that this is the number of managers in the hierarchy, even though there are actually only 9,331 managers in the discrete-tier model. Qian (1994) and Wang (1993) would say that this is the total size of the hierarchy, including operatives, whereas the actual size of the hierarchy is 55,986 in the discrete-tier model.

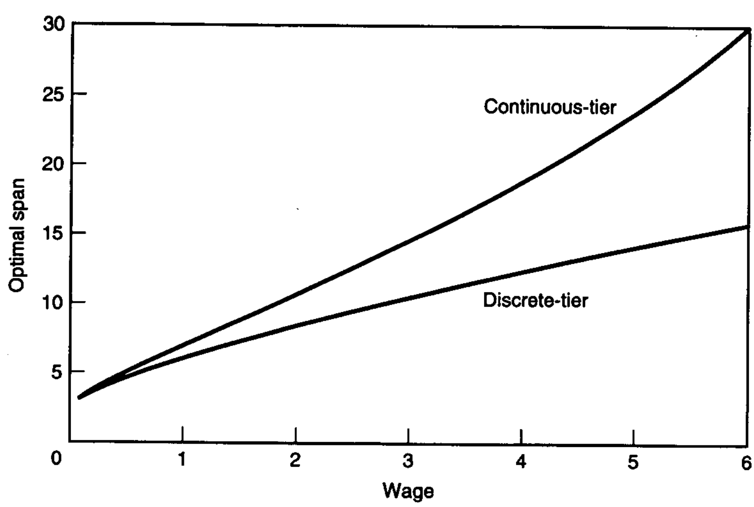
Although it is ambiguous whether Q^{uc} overestimates the number of managers or underestimates the size of the hierarchy, the distortion to the span's marginal effect on the number of managers or size of the hierarchy is unambiguous, since the number of operatives is fixed. dQ^{uc}/ds and dQ^{ua}/ds are both negative, and $dQ^{uc}/ds < dQ^{ua}/ds$. Hence, Q^{uc} overestimates the marginal managerial cost of a decrease in the span of control. In Keren and Levhari, with the restriction that the hierarchies be uniform, the first-order condition for the optimal span is $-wdQ/ds = c_D(D)dD/ds$. Therefore, the solution in the continuous-tier model is too high.

To give an idea of the magnitude of the error in Keren and Levhari, suppose that $c(D, n) = kn(1 - e^{\lambda D})$. Then the relative error for the solution and the relative inefficiency increase unboundedly in w. Figure 1 compares the solutions as a function of w. Although the error diminishes when $w \downarrow 0$, this is only because the measure of the number of managers becomes irrelevant. Furthermore, for $w = 0$ the optimal hierarchies are known to be uniform and to have a span of three, and the continuous-tier model is not helpful for this result.

For the model of Qian (1994), there is a simple example that shows how the continuous-tier model can lead to errors. In that model, output is given by a managerial production function that depends also on the level of effort of the managers. The level of effort, in turn, depends on the wage and the span of control of the superiors, whose responsibilities include monitoring. Qian considers a version in which there are only two effort levels and only two possible levels of total output. The problem reduces to one of minimizing the total wage bill, with the number n of operatives fixed. The wage in tier $h > 0$ is (proportional to) s_{h-1}, because the wage (needed to induce effort) decreases with the probability of being monitored. Hence, the total variable wage bill is

FIGURE 1

OPTIMAL SPAN FOR UNIFORM HIERARCHIES IN THE KEREN AND LEVHARI MODEL



Note: For $c(D, n) = 2n(1 - e^{-.01D})$ and $n = 100{,}000$.

$$\sum_{h=1}^{H} s_{h-1} q_h. \tag{17}$$

In the continuous-tier version, the wage bill becomes

$$\int_{h=0}^{H} s(h)q(h)\, dh. \tag{18}$$

Qian shows that the wage bill is minimized by setting $s(h) = e$ for all h, and thus $H = \log n$.

Compare this with the discrete-tier model. Qian shows that here also the optimal heirarchy has a constant span. The wage bill, as a function of the span, is

$$W^d(s) = \sum_{h=1}^{H} s^{h+1} = \frac{s^2(s^H - 1)}{s - 1} = (n - 1)\frac{s^2}{s - 1}. \tag{19}$$

This formula is exact for integer s such that n is a power of s, and otherwise it is an approximation. In the continuous-tier model, the wage bill is

$$W^c(s) = s \int_0^H s^h\, dh = \frac{s(s^H - 1)}{\log s} = (n - 1)\frac{s}{\log s}. \tag{20}$$

W^d and W^c are both convex functions. The first derivatives are

$$W^{d\prime}(s) = (n - 1)\frac{s^2 - 2s}{(s - 1)^2}$$

$$W^{c\prime}(s) = (n - 1)\frac{\log s - 1}{(\log s)^2}.$$

The solutions to the first-order conditions are, respectively, $s_d = 2$ and $s_c = e$ (and the integer solution to min $W^c(s)$ is three). Hence, the first-order conditions are distorted. One can see from the formulae for W^d and W^c that a constant error in the span translates into an efficiency loss that goes up linearly with the firm size: $W^d(2) = 4(n - 1)$ and $W^d(e) \approx 4.3(n - 1)$. Hence, again, the relative efficiency loss from using the solution to the continuous formulation and the relative error in the minimized costs do not diminish for large n.[7]

In Qian (1994) there are two ways in which the continuous-tier model distorts the optimization problem and leads to an excessive span. First, as in Keren and Levhari (1979, 1983), the continuous-tier model overestimates the marginal effect on the number of managers of a decrease in the span. Second, because Qian assumes that Q^{uc} includes the operatives and hence Q^{uc} underestimates the total size of the hierarchy, and because the wages of all members of the hierarchy depend on the span, the continuous-tier model underestimates the marginal effect of a decrease in wages that is induced by a decrease in the span.

[7] Qian (1994) also derives the solution to the discrete model in his Appendix A. He observes that even though the solutions differ, they have the same "qualitative" properties (e.g., constant span). That the qualitative properties are preserved in this example even though the solutions differ dramatically may be fortuitous and does not demonstrate that the similarity of qualitative results is a general principle.

I have focused on uniform hierarchies in this section because it makes the comparison between the continuous-tier and discrete-tier models straightforward. (For the result in Qian (1994) discussed above, this restriction is not binding.) However, having a variable span does not help the approximations. If we map a discrete-tier hierarchy with $H + 1$ tiers and spans $s_0, \ldots, s_{H-1}$ to a continuous-tier hierarchy with $H + 1$ tiers and span $s(h) = s_{\lfloor h \rfloor}$, then the discrete and continuous formulae D^a and D^c for the delay (equations (4) and (13), respectively) coincide for such hierarchies, as do the formulae (1) and (11) for q_h when h is an integer. However, the formulae Q and Q^c for the number of managers (equations (2) and (12), respectively) are distinct, since (12) becomes

$$Q^c = \sum_{h=0}^{H-1} q_h \int_0^1 s_h^t \, dt = \sum_{h=0}^{H-1} q_h (s_h - 1)/\log s_h. \tag{21}$$

In Keren and Levhari, $q_h(s_h - 1)/\log s_h$ is interpreted as the number of managers in tier h, and hence it overestimates q_h by a factor of $(s_h - 1)/\log s_h$. In Qian (1994) and Wang (1993), $q_h(s_h - 1)/\log s_h$ is interpreted as the number of managers or operatives in tier $h + 1$. Hence, it underestimates q_{h+1} by a factor of $(s_h - 1)/(s_h \log s_h)$. On a tier-by-tier basis, these errors are the same as those we derived for uniform hierarchies. Because there are more operatives than managers, or more managers in tier $h - 1$ than in any other tier, the span of the lower tiers determines the overall magnitude of the error. In all these models, when the span is variable it is increasing. Hence, the relative errors are again significant, and may even grow, as the hierarchies get larger.

6. Getting by without continuous approximations

■ The rationale for using continuous approximations is that they simplify the analysis of models. I conclude this article by commenting on how critical such simplifications are, especially in light of the negative results in Section 5.

Economic theorists are mainly interested in so-called qualitative properties of hierarchies, such as whether they are balanced, whether the span of control increases or decreases down the hierarchy, whether an increase in the managerial wage makes hierarchies taller or shorter, and whether hierarchies have increasing or decreasing returns to scale. For empirical work it may be important to know the actual magnitudes, such as the value of the optimal span of control, but it is always possible to calculate such quantities numerically, without using continuous approximations. Hence, the important question is whether continuous approximations are critical for deriving qualitative properties of hierarchies.

When studying nonbalanced hierarchies and more general graphs, there typically are not even any continuous analogs to use as approximations. Yet discrete mathematics is rich in results about such discrete structures, and Mount and Reiter (1990), Radner (1993), Van Zandt (1994a), Marschak and Reichelstein (1993, 1994), and Li (1994), among others, contain many results about hierarchies and general networks that use no continuous approximations at all. One advantage to working directly with discrete models is that proofs are often constructive and provide intuition about the results. Examples of such constructive proofs can be found in all the articles cited above.

Nevertheless, it is true that various models of balanced hierarchies, including but not limited to Williamson (1967), Beckmann (1977), and Geanakoplos and Milgrom (1991), have made beneficial use of the continuous approximations discussed in Sections 3 and 4. The positive results in those sections give us some confidence in these approximations, even when we take into account the caveats stated in Section 1.

Keren and Levhari (1979, 1983, 1989), Qian (1994), and Wang (1993) have also made powerful use of the continuous-tier formulation, and they have derived explicit formulae for the paths of the spans, wages, and other variables in hierarchies. The results of Section 5 show that these formulae are inaccurate, which also means proofs of qualitative results that use these formulae are not sound. However, although it may be difficult to derive similar quantitative formulae for the discrete-tier model, Chakraborty (1994) derives many of the qualitative results in Keren and Levhari (1979) and Qian (1994) using a discrete-tier model, and the proofs are not more difficult when the continuous approximations of Section 3 are used. The results on decreasing returns to scale in Keren and Levhari (1983) can be derived from the lower bound e log n for the delay. Qian also illustrates what can be accomplished with discrete-tier models. This suggests that the negative results of Section 5 will not hinder the study of hierarchies.

Appendix

■ I make frequent use of the following properties of the cost c of delay, which hold because c is positive and concave:

$$c(D + \Delta D) - c(D) \leq c_D(D)\Delta D$$

$$c(D)/D \geq c_D(D).$$

Proof of Proposition 2. Because $C^a(\mathcal{H}^*) \leq C^* \leq C(\mathcal{H}^*)$, the relative inefficiency and relative error are no greater than

$$RE \equiv \frac{C(\mathcal{H}^*) - C^a(\mathcal{H}^*)}{C^a(\mathcal{H}^*)}.$$

Let H^* be the height of $\mathcal{H}^*$. Rounding the span up at each tier increases delay less than H^*. Hence,

$$RE < \frac{c\big(D^a(\mathcal{H}^*) + H^*\big) - c\big(D^a(\mathcal{H}^*)\big)}{wQ(\mathcal{H}^*) + c\big(D^a(\mathcal{H}^*)\big)}. \tag{A1}$$

Because c is positive and concave in D,

$$c\big(D^a(\mathcal{H}^*) + H^*\big) - c\big(D^a(\mathcal{H}^*)\big) \leq c_D\big(D^a(\mathcal{H}^*)\big)H^*$$

and

$$c\big(D^a(\mathcal{H}^*)\big) \geq c_D\big(D^a(\mathcal{H}^*)\big)D^a(\mathcal{H}^*).$$

Therefore,

$$RE < \frac{c_D\big(D^a(\mathcal{H}^*)\big)H^*}{wQ(\mathcal{H}^*) + c_D\big(D^a(\mathcal{H}^*)\big)D^a(\mathcal{H}^*)} < H^*/D^a(\mathcal{H}^*). \tag{A2}$$

The middle expression in (A2) is decreasing in $c_D(D^a(\mathcal{H}^*))$ when the remaining terms are fixed. Therefore, since $\hat{c}_D(n) < c_D(D^a(\mathcal{H}^*))$ and $H^* < \log_2 n$,

$$RE < \frac{\hat{c}_D(n)\log_2 n}{wQ(\mathcal{H}^*) + \hat{c}_D(n)D^a(\mathcal{H}^*)}. \tag{A3}$$

Let q^*_{H-1} be the number of managers in the bottom managerial tier of $\mathcal{H}^*$. Then $Q(\mathcal{H}) \geq q^*_{H-1}$ and $D^a(\mathcal{H}) \geq n/q^*_{H-1}$. The denominator of (A3) is thus greater than or equal to $wq^*_{H-1} - \hat{c}_D(n)n/q^*_{H-1}$. This is a convex function of q^*_{H-1} whose minimum value is $2\sqrt{\hat{c}_D(n)nw}$. Therefore,

$$RE = \frac{\sqrt{\hat{c}_D(n)}\,\log_2 n}{2\sqrt{nw}}. \tag{A4}$$

Q.E.D.

Proposition 3. Let (i) $\mathcal{H}^*$ be a solution to min $C^a(\mathcal{H})$, without integer constraints on the number of managers at each tier; (ii) $\mathcal{H}^i$ be the hierarchy with an integer number of managers in each tier, obtained by rounding up the number of managers in each tier in $\mathcal{H}^*$; and (iii) C^* be the integer-constrained minimum cost min $C^a(\mathcal{H})$. Then $D^a(\mathcal{H}^i) - D^a(\mathcal{H}^*) < 2$ and $Q(\mathcal{H}^i) - Q(\mathcal{H}^*) < H^* - 1$. Furthermore,

$$\underset{\text{(relative inefficiency)}}{\frac{C^a(\mathcal{H}^i) - C^*}{C^*}} \quad \text{and} \quad \underset{\text{(relative error)}}{\frac{C^* - C^a(\mathcal{H}^*)}{C^*}} \quad \text{are less than} \quad \frac{H^*}{2^{H^*}} + \frac{1}{H^*}, \tag{A5}$$

and for $\hat{c}(n) \geq \max\{2, 2w\}$, they are also less than

$$\frac{w \log_2\left(\hat{c}(n)/w\right)}{\hat{c}(n)} + \frac{2}{e \log n}. \tag{A6}$$

Proof. Let q_h^* be the number of managers in tier h of $\mathcal{H}^*$. For each tier $h = 0, \ldots, H^* - 1$, the difference between the delay of $\mathcal{H}^i$ and $\mathcal{H}^*$ is

$$\frac{\lceil q_{h+1}^* \rceil}{\lceil q_h^* \rceil} - \frac{q_{h+1}^*}{q_h^*} < \frac{q_{h+1}^* + 1}{q_h^*} - \frac{q_{h+1}^*}{q_h^*} = \frac{1}{q_h^*}.$$

As long as the span of each tier is at least two, $q_h^* \geq 2^h$ and $\sum_{h=0}^{H^*-1} 1/q_h^* < 2$. Rounding up adds less than one manager per tier $h = 1, \ldots, H - 1$.

Because $C^a(\mathcal{H}^*) \leq C^* \leq C^a(\mathcal{H}^i)$, the relative inefficiency and relative error in (A5) are less than

$$RE \equiv \frac{C^a(\mathcal{H}^i) - C^a(\mathcal{H}^*)}{C^a(\mathcal{H}^*)}.$$

Because c is concave,

$$C^a(\mathcal{H}^i) - C^a(\mathcal{H}^*) < w(H^* - 1) + c_D\left(D^a(\mathcal{H}^*)\right)2. \tag{A7}$$

Therefore,

$$RE < \frac{w(H^* - 1) + c_D\left(D^a(\mathcal{H}^*)\right)2}{wQ(\mathcal{H}^*) + c\left(D^a(\mathcal{H}^*)\right)}. \tag{A8}$$

Because $Q(\mathcal{H}^*) \leq 2^{H^*} - 1$ and $D^a(\mathcal{H}^*) \geq 2H^*$,

$$RE < \frac{w(H^* - 1)}{w(2^{H^*} - 1)} + \frac{c_D(2H^*)2}{c(2H^*)} \tag{A9}$$

$$\leq H^*/2^{H^*} + 1/H^*. \tag{A10}$$

I conclude by deriving the upper bound (A6) from (A8). First, here is a bound on the second term in the numerator of (A8):

$$\frac{c_D\left(D^a(\mathcal{H}^*)\right)2}{wQ(\mathcal{H}^*) + c\left(D(\mathcal{H}^*)\right)} < \frac{\hat{c}_D(n)2}{\hat{c}(n)} \leq \frac{2}{e \log n}. \tag{A11}$$

Next, here is a bound on the first term in the numerator of (A8):

$$\frac{w(H^* - 1)}{wQ(\mathcal{H}^*) + c\left(D^a(\mathcal{H}^*)\right)} \leq \frac{w(H^* - 1)}{w(2^{H^*} - 1) + \hat{c}(n)} \equiv RS. \tag{A12}$$

I need to show that RS is bounded by the first term in (A6). For brevity, I write $\hat{c}$ instead of $\hat{c}(n)$. If $H^* \leq \log_2(\hat{c}/w)$, then RS is less than $w \log_2(\hat{c}/w)/\hat{c}$. I show that the same is true if $H^* > \log_2(\hat{c}/w)$, when $\hat{c} \geq 2w$. For this purpose, let $H^* = \log_2(\hat{c}/w) + \alpha$, where $\alpha > 0$. I have to verify that

$$\frac{w \log_2(\hat{c}/w)}{\hat{c}} \overset{?}{>} \frac{w\left(\log_2(\hat{c}/w) + \alpha - 1\right)}{w(2^{(\log_2(\hat{c}/w)+\alpha)} - 1) + \hat{c}} \tag{A13}$$

$$(\hat{c}2^\alpha - w + c)\log_2(\hat{c}/w) \overset{?}{>} \hat{c}\left(\log_2(\hat{c}/w) + \alpha - 1\right) \tag{A14}$$

$$2^\alpha \log_2(\hat{c}/w) \overset{?}{>} \log_2(\hat{c}/w) + \alpha - 1 \tag{A15}$$

$$2^\alpha > \alpha. \tag{A16}$$

To get from (A14) to (A15), I used $\hat{c} > w$ to drop $-w + c$ from the left-hand side, and then I factored out $\hat{c}$. To get from (A15) to (A16), I used $\hat{c} \geq 2w$ and hence $\log_2(\hat{c}/w) \geq 1$. The last inequality (A16) holds for all α. *Q.E.D.*

Proof of Proposition 4. Any hierarchy with n operatives must satisfy $\prod S_h \geq n$, and for any positive integers $S_0, \ldots, S_{H-1}$ such that $\prod S_h \geq n$, we can construct a hierarchy with n operatives such that the maximum span at tier h is at most S_h (let $q_h = \prod_{h'=0}^{h-1} S_{h'}$). Hence, $D^u(s, n)$ is equal to the value of the following integer programming problem:

$$\min_{y_0,\ldots,y_{H-1}>0} \sum_{h=0}^{H-1} y_h \tag{A17}$$

subject to

$$\prod_{h=0}^{H-1} y_h \geq n.$$

Without integer constraints, the unique solution is $y_o = \cdots = y_{H-1} = s$ and the value is Hs. With integer constraints, one can show that the value of this problem is $k\lfloor s \rfloor + (H - k)\lceil s \rceil$, where k is the largest integer such that

$$\lfloor s \rfloor^k \lceil s \rceil^{H-k} \geq n. \tag{A18}$$

To find k, find the real number such that (A18) holds with equality,

$$\lfloor s \rfloor^k \lceil s \rceil^{H-k} = S^H, \tag{A19}$$

and then round down. Raise both sides of (A19) to $1/H$, let $\lambda = k/H$, let $\hat{s} = \lfloor s \rfloor$, and let $\alpha = s - \hat{s}$. Then, if s is not an integer (the only interesting case), (A19) becomes

$$\hat{s}^\lambda(\hat{s} + 1)^{1-\lambda} = \hat{s} + \alpha \tag{A20}$$

$$\lambda \log \hat{s} + (1 - \lambda)\log(\hat{s} + 1) = \log(\hat{s} + \alpha) \tag{A21}$$

$$\lambda = \frac{\log(\hat{s} + 1) - \log(\hat{s} + \alpha)}{\log(\hat{s} + 1) - \log(\hat{s})}. \tag{A22}$$

If $H\lambda$ is an integer, then $k = H\lambda$ and the delay is $H(\lambda\hat{s} + (1 - \lambda)(\hat{s} + 1))$. Otherwise, $k = \lfloor H/\lambda \rfloor$. Rounding down adds at most one unit of delay. Therefore,

$$\begin{aligned} D^u(s, n) - D^{ua}(s, n) &< H\left(\lambda\hat{s} + (1 - \lambda)(\hat{s} + 1)\right) + 1 - H(\hat{s} + \alpha) \\ &= H(1 - \lambda - \alpha) + 1 \\ &< H/(8s) + 1. \end{aligned}$$

The last inequality comes from $1 - \lambda - \alpha < 1/(8s)$, which one can verify (for all α and s) by substituting the Taylor expansion around $\hat{s}$ for the logarithms in (A22). It then follows that

$$\frac{D^u(s, n) - D^{ua}(s, n)}{D^{ua}(s, n)} < \frac{H/(8s) + 1}{Hs} = \frac{1}{8s^2} + \frac{(\log s)/s}{\log n}. \tag{A23}$$

Q.E.D.

Proof of Proposition 5. Let s and n be integers and let $H = \lceil \log_s n \rceil$. The hierarchy with the fewest managers among those with a maximum span of s is constructed recursively by setting $q_H = n$ and $q_h = \lceil q_{h+1}/s \rceil$. The number of managers in tier $H - 1$ is

$$q_{H-1} = \left\lceil \frac{n}{s} \right\rceil < \frac{n}{s} + 1. \tag{A24}$$

The number of managers in tier $H - 2$ is

$$q_{H-2} = \left\lceil \frac{\left\lceil \frac{n}{s} \right\rceil}{s} \right\rceil < \frac{\frac{n}{s} + 1}{s} + 1 = \frac{n}{s^2} + \frac{1}{s} + 1. \tag{A25}$$

In general,

$$n/s^{H-h} \leq q_h < n/s^{H-h} + s/(s - 1).$$

Observe that $\sum_{h=0}^{H-1} n/s^{H-h} = (n - n/s^H)/(s - 1)$. Since $n/s^H \leq 1$, $(n - n/s^H)/(s - 1)$ is at least as large as $Q^{ua}(s, n)$. On the other hand,

$$\frac{n - n/s^H}{s - 1} < \frac{n - 1/s}{s - 1} = Q^{ua}(s, n) + \frac{1}{s}.$$

Thus,

$$0 \leq Q^u(s, n) - Q^{ua}(s, n) \leq 1/s + Hs/(s - 1). \tag{A26}$$

Now divide by $(n - 1)/(s - 1)$. *Q.E.D.*

References

AOKI, M. "Non-hierarchical Aspects of the Internal Organization of the Enterprise: A Partial Survey in Comparative Perspective." CEPR Publication no. 331, Stanford University, 1992.

BECKMANN, M.J. "Some Aspects of Returns to Scale in Business Administration." *Quarterly Journal of Economics,* Vol. 74 (1960), pp. 464–471.

———. "Management Production Functions and the Theory of the Firm." *Journal of Economic Theory,* Vol. 14 (1977), pp. 1–18.

———. *Tinburgen Lectures on Organization Theory.* Berlin: Springer-Verlag, 1983.

BOLTON, P. AND DEWATRIPONT, M. "The Firm as a Communication Network." *Quarterly Journal of Economics,* Vol. 109 (1994), pp. 809–952.

CALVO, G.A. AND WELLISZ, S. "Supervision, Loss of Control and the Optimal Size of the Firm." *Journal of Political Economy,* Vol. 86 (1978), pp. 943–952.

CHAKRABORTY, A. "Optimal Hierarchies with Countably Many Tiers." Mimeo, Department of Economics, Princeton University, 1994.

GEANAKOPLOS, J. AND MILGROM, P. "A Theory of Hierarchies Based on Limited Managerial Attention." *Journal of the Japanese and International Economies,* Vol. 5 (1991), pp. 205–225.

KEREN, M. AND LEVHARI, D. "The Optimum Span of Control in a Pure Hierarchy." *Management Science,* Vol. 11 (1979), pp. 1161–1172.

——— AND ———. "The Internal Organization of the Firm and the Shape of Average Costs." *Bell Journal of Economics,* Vol. 14 (1983), pp. 474–486.

——— AND ———. "Decentralization, Aggregation, Control Loss and Costs in a Hierarchical Model of the Firm." *Journal of Economic Behavior and Organization,* Vol. 11 (1989), pp. 213–236.

LI, H. "The Organizational Structure of Strategy." Mimeo, Department of Economics, University of Chicago, 1994.

MARSCHAK, T. AND REICHELSTEIN, S. "Network Mechanisms, Informational Efficiency, and the Role of Hierarchies." Haas School of Business, University of California at Berkeley, 1993.

——— AND ———. "Communication Requirements for Individual Agents in Networks and Hierarchies." Haas School of Business, University of California at Berkeley, 1994.

MINTZBERG, H. *The Structuring of Organizations.* Englewood Cliffs, N.J.: Prentice-Hall, 1978.

MOUNT, K. AND REITER, S. "A Model of Computing with Human Agents." Center for Mathematical Studies in Economics and Management Science Discussion Paper no. 890, Northwestern University, 1990.

QIAN, Y. "Incentives and Loss of Control in an Optimal Hierarchy." *Review of Economic Studies,* Vol. 61 (1994), pp. 527–544.

RADNER, R. "The Organization of Decentralized Information Processing." *Econometrica,* Vol. 61 (1993), pp. 1109–1146.

VAN ZANDT, T. "The Scheduling and Organization of Periodic Associative Computation." Mimeo, Department of Economics, Princeton University, 1994a.

———. "Organizations that Process Information with Endogenous Agents. Mimeo, Department of Economics, Princeton University, 1994b.

———. "Organizations that Process Information with Endogenous Agents: Supplementary Notes." Mimeo, Department of Economics, Princeton University, 1994c.

WANG, Y. "The Effect of Organizational Cost on the Agency Problem in Hierarchies." Mimeo, Department of Economics, University of Minnesota, 1993.

WILLIAMSON, O.E. "Hierarchical Control and Optimum Firm Size." *Journal of Political Economy,* Vol. 75 (1967), pp. 123–138.