On the Size Distribution of Business Firms

Robert E. Lucas, Jr.

The Bell Journal of Economics, Volume 9, Issue 2 (Autumn, 1978), 508-523.

Stable URL:
http://links.jstor.org/sici?sici=0361-915X%28197823%299%3A2%3C508%3AOTSDOB%3E2.0.CO%3B2-S

On the size distribution of business firms

Robert E. Lucas, Jr.
Department of Economics
University of Chicago

This paper proposes a new theory of the size distribution of business firms. It postulates an underlying distribution of persons by managerial "talent" and then studies the division of persons into managers and employees and the allocation of productive factors across managers. The implications of the theory for secular changes in average firm size are developed and tested on U.S. time series.

1. Introduction

■ The designation of Herbert A. Simon as a Distinguished Fellow of the American Economic Association provides a happy occasion for his friends, colleagues, and students to gather to talk economics. For me, it also offers the opportunity of claiming a distinction for myself, one which I feel I have earned many times over yet have no degree to show for it: that of being one of Herb's students in economics, and in social science more generally, at Carnegie Tech.

In this paper, I would like to return to an area to which I contributed, rather obliquely, ten years ago, and to which Simon has contributed substantially both before and since: the size distribution of business firms. This topic is one on which important issues of economic policy are held to hinge: in wealthy economies, "bigness" is widely viewed as a menace against which government activity should, perhaps, be directed; in poor economies, "littleness" is often viewed as a sign of backwardness to be dealt with by government policy. The lack of discipline with which such issues are typically discussed reflects, I think, the lack of an adequate understanding of the forces determining firm size.

2. Existing theory and its implications

■ For many years, the theory of firm size advanced in Jacob Viner's classic paper (1932) dominated economists' thinking on the size distribution of firms. This theory predicts a unique size distribution within an *industry* under the

I would like to thank Sherwin Rosen and José Scheinkman for very helpful discussions, and Dan Gressel for his assistance. At the Conference Honoring Herbert A. Simon's Economic Contributions at Carnegie-Mellon University in October, 1977, I received helpful suggestions and corrections from many participants—Dennis Epple, Tsu Yao, and Herbert Simon in particular—as acknowledged more specifically in the text. Finally, I would like to thank two anonymous referees and Alvin Klevorick.

assumption that individual firms have *U*-shaped long-run average cost functions. In equilibrium, each firm produces at the minimum point of this curve, with firm entry or exit adapting so as to adjust total industry production to quantity demanded at the zero-profit price. The size distribution which emerges, then, is a solution to an extremum problem: allocate production over firms so as to minimize total cost.

Viner's theory proceeds under the *assumption* of product market competition. It cannot, therefore, provide a complete guide to policy directed at reducing monopoly power. It does, however, direct attention to a major potential *cost* of policies directed against large firms. By keeping firms small, relative to their most efficient (in a productive sense) scale, such a policy results in a waste of resources which must somehow be balanced against the gains from reducing monopoly inefficiency. Not surprisingly, scale economies have been a standard and often successful defense in antitrust actions.

The evidence against this nonvacuous version of the Viner theory is by now so overwhelming that few economists accept it, except perhaps as a model of *plant* or *store* size. Perhaps most fundamentally, the fact that the typical firm sells in a large number of product markets leads to enormous difficulties in applying the Marshallian industry abstraction to the firm size problem. Second, in contradiction to the theory, most changes in product demand are met by changes in firm size, *not* by entry or exit of firms. Third, the percentage rate of firm growth appears to be independent of its size (measured by sales, employment, or assets). Finally, Simon, in collaboration with Charles Bonini (1958) and Yuji Ijiri (1964), observed that by examining the distribution of firms by size at a single point in time, one can make inferences about the stochastic process which governs firm growth. This insight has been exploited in several directions, and has led to further confirmation of the law that firm growth is independent of size.

The models of Bonini, Ijiri, and Simon make no use of the hypothesis of maximizing behavior on the part of firms. Yet it is not at all difficult to reconcile their findings with a suitably modified, *constant-returns-to-scale* version of Viner's theory (cf. Lucas, 1967). On this version of the theory, *all* size distributions are competitive equilibria, and all are solutions to the extremum problem of producing at minimum resource cost.

This conclusion that the size distribution is, under competition, a matter of indifference from a welfare point of view is more far-reaching than is commonly recognized. It means that the difficult balancing between productive efficiency on the one hand and enforcement of competition on the other, as indicated by the Viner theory, is entirely unnecessary. The possible *gains* from policies directed against large firms are, because of the possibility of inefficiencies due to monopoly, positive; the *costs* in productive efficiency losses are, under constant returns, nil. Thus even a clumsy strategy of opposition to mergers and forced dissolutions will lead to beneficial or at worst, harmless, results.

These are the policy implications of the existing theory of the size distribution of business firms, under the version of that theory which is consistent with available evidence. Yet with the exception of George Stigler's discussion (1964), I am not aware of any place where these implications have been carefully developed and seriously advocated as the basis for actual practice in antitrust policy.

I suspect that one reason so few economists have followed Stigler's advocated policies against bigness lies in a suspicion that, if the existing size distribution does not uniquely solve the problem "allocate production over firms so as to minimize cost," then it uniquely solves some *other* extremum problem. Yet simply voicing this suspicion, without spelling out what this other problem is, is merely an application of the justly discredited principle that "what is, is best."

3. An alternative theory

■ In this paper I want to develop the implications of a suggestion of Henry G. Manne (1965) that the observed size distribution is a solution to the problem: allocate productive factors over managers of different ability so as to maximize output. Manne's paper was devoted mainly to explaining how takeovers via proxy fights, share purchases, and mergers can be viewed as *mechanisms* by which talented managers acquire more productive factors. My strategy will be to assume at the outset that this allocation of assets and employees over managers is *perfectly* carried out (by some mechanism which I will not specify) and then to work out some of the implications of this hypothesis. This line will, I hope, be taken as complementary to that pursued by Manne.

As a preliminary matter, it is necessary to decide whether to follow Viner and develop a model at the *industry* level, or to theorize at an economy-wide level. The former would be appropriate if managerial ability were industry-specific, as a result of "nature" or of the accumulation of industry-specific expertise. There is no doubt that these considerations matter, at some level, but the multiproduct nature of at least the largest modern firms and the mobility of top managers across industries suggest that, if one is choosing between these extremes, it is best to attack the problem at an economy-wide level.

To keep matters simple, at least initially, I shall develop a model of a closed economy with a given quantity of homogeneous capital and a given workforce which is homogeneous with respect to productivity as an employee. Each member of the workforce is also endowed with a "talent for managing" which varies across workers. A *firm* in this economy is one manager, together with the capital and labor under his or her control.[1] Resource allocation involves, first, a division of the workforce into managers and employees and, second, the allocation of factors of production across managers. After specifying the technology of production and of management, I shall study the problem of allocating resources in an output-maximizing way: the dual of the problem studied by Viner. In common with Viner's theory, the solution to this extremum problem will be a competitive equilibrium.

As with Viner's theory, this model "predicts" the size-distribution of firms, but only *given* the distribution of persons by managerial talent. Without some means of learning about the latter, then, the theory does not place restrictions on the kind of data utilized in the Simon-Bonini-Ijiri studies. On the other hand, the theory places definite restrictions on the way in which average

[1] Richard Kihlstrom and Jean-Jacques Laffont (1977) developed an equilibrium theory in which agents differ in their attitudes toward risk with the relatively least risk averse becoming entrepreneurs. I am here adopting exactly their formulation, with "attitudes toward risk" replaced by "talent for managing."

firm size should vary across economies at different levels of development. These implications form the basis for some time-series regressions.

After developing an explicit model (Sections 4 and 5) and reporting some empirical results (Section 6), I shall consider two ways in which the model might fruitfully be elaborated (Section 7). Section 8 is a conclusion.

4. An explicit model

■ In this section I shall consider a closed economy with a workforce of size N and with K units of homogeneous capital, both inelastically supplied to the market. These factors of production may be combined, in a manner to be specified in a moment, so as to produce Y homogeneous units of output. After specifying the technology, I shall study the problem of combining productive factors so as to maximize Y; the solution to this problem will also be a competitive equilibrium.

It will be helpful (though slightly artificial) to consider separately the *production technology* and the *managerial* (or entrepreneurial) *technology*. For the former, let $f(n,k)$ be the output produced with n units of labor and k of capital, under "normal" or "representative" management. Let this technology exhibit constant returns, so that we may write $f(n,k) = n\phi(r)$, where $r = k/n$, and ϕ: $R^+ \rightarrow R^+$ is a twice differentiable function, increasing and strictly concave. If everyone in the economy were capable of "normal" management at any scale of operations, then optimal and equilibrium output would be $Y = N\phi(K/N) = N\phi(R)$, letting R denote capital per capita, and equilibrium wages and capital rentals would be $w = \phi(R) - R\phi'(R)$ and $u = \phi'(R)$, respectively. Rents to managers or entrepreneurs would be zero.

The managerial technology involves two elements: variable skill or talent, and an element of diminishing returns to scale, or to "span of control." For the first, each agent is endowed with a managerial talent level x, drawn from a fixed distribution Γ: $R^+ \rightarrow [0,1]$. If agent x manages resources n and k, his "firm" produces $xg[f(n,k)]$ units of output, where g: $R^+ \rightarrow R^+$ is twice-differentiable, increasing, and strictly concave, satisfying $g(0) = 0$. That is to say, each "firm" consists of a single manager (or entrepreneur), n homogeneous employees, and k homogeneous units of capital.[2]

This description of management is a shallow one, in at least two respects. First, it does not say anything about the nature of the tasks performed by managers, other than that whatever managers do, some do it better than others. Given this assumption, however, one is led immediately to the question: why does the best manager not run everything? Therefore, I assume concavity of the function g. Second, this technology precludes pyramidal managerial structures: managers managing other managers. One could postulate a technology for such organizations without any difficulty in a mathematical sense, but without a clear idea of where one is going, this is likely to lead to an uninformative taxonomy. Let me proceed for a while at this simple level, with the understanding that what we may hope for is not serious organization theory, but perhaps some insights into why organization theory matters economically.

---

[2] Notice that *no* individual x is capable of producing under what I have called above "normal management" $f(n,k)$ at all scales of operations. Thus the normal or "representative" manager is an expositional fiction, as these terms are used here.

To complete the statement of the problem, it is convenient to think of a continuum of agents, so that the entire distribution Γ of talent is always fully represented. Then an allocation of resources is described by two functions $n(x)$ and $k(x)$, giving the labor and capital managed by agent x. Since production requires both chiefs and Indians, an allocation which yields a positive output will have $n(x) = k(x) = 0$ for some x-values (those corresponding to employees) and positive for others (those corresponding to managers). For efficient or equilibrium allocations, it will be only the most talented who manage, so that there will exist a cutoff level $z > 0$ such that if $x < z$, one is an employee, and, if $x \geq z$, one is a manager. By an *allocation*, then, I shall mean a *number* z and a pair of *functions* $n(x)$, $k(x)$:$R^+ \rightarrow R^+$ such that $n(x) = k(x) = 0$ for $x < z$ and $n(x)$, $k(x) > 0$ for $x \geq z$.

An allocation is *feasible* if it does not utilize more than the available population N and capital K. That is,

$$1 - \Gamma(z) + \int_z^\infty n(x)d\Gamma(x) \leq 1 \tag{1}$$

(the fraction engaged in managing plus the fraction engaged as employees sum to no more than one) and

$$\int_z^\infty k(x)d\Gamma(x) \leq \frac{K}{N} = R. \tag{2}$$

An *efficient* allocation is one which maximizes output,

$$\frac{Y}{N} = \int_z^\infty xg[f(n(x),k(x))]d\Gamma(x) \tag{3}$$

subject to (1) and (2).

In the Lagrangian associated with this variational problem, let w and u be the multipliers (both constant with respect to x) associated respectively with the constraints (1) and (2). Then the efficient allocation will also be the *competitive equilibrium*, with w and u being the equilibrium wage rate and rental price of capital. The income or rent to manager $x > z$ will be the residual

$$xg[f(n(x),k(x))] - wn(x) - uk(x). \tag{4}$$

The first-order conditions for this maximum problem include:

$$xg'(f)f_n(n(x),k(x)) = w, \quad x \geq z \tag{5}$$

and

$$xg'(f)f_k(n(x),k(x)) = u, \quad x \geq z \tag{6}$$

so that the marginal products of both factors are equated across firms. Recalling that $f(n,k) = n\phi(r)$, where $r = k/n$, (5) and (6) imply

$$\frac{\phi(r) - r\phi'(r)}{\phi'(r)} = \frac{w}{u}. \tag{7}$$

Thus, given the ratio of factor prices, all firms have a common capital-labor ratio $r(w/u)$ given implicitly by (7). The function $r(\cdot)$ is strictly increasing.

Given r from (7), the equilibrium scale $n(x)$ of firm x can be obtained from either (5) or (6). Using (6),

$$xg'[n(x)\phi(r)]\phi'(r) = u, \tag{8}$$

which gives employment as an implicit function $n(x,w,u){:}x \geq z$. The function $n(\cdot)$ is increasing in x and u, and decreasing in w.

The first-order condition for the cutoff value z is

$$zg[f(n(z),k(z))] = w[1 + n(z)] + uk(z). \tag{9}$$

For the marginal manager z, as for all, $k(z) = rn(z)$ with r given by (7), so that (9) may be written:

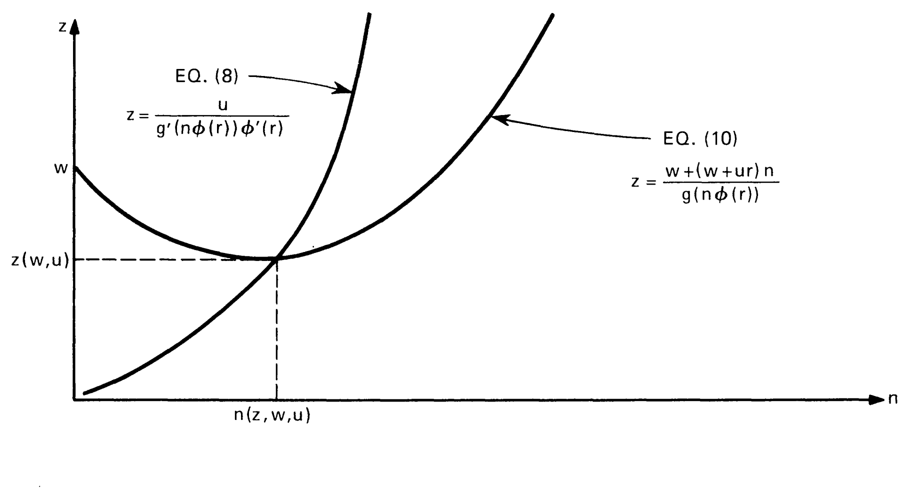
$$zg[n(z)\phi(r)] = w + (w + ur)n(z). \tag{10}$$

The right-hand side of (10) (or (9)) is total cost, consisting of a "fixed cost" w, the opportunity cost of the manager's own time, and a constant-per-employee variable cost, $(w + ur)n(z)$. The left side is output (or, with output serving as numeraire, the *value* of output). Thus (10) is a breakeven, or average-cost-equals-price condition for the marginal manager.

The marginal manager's employment level must also satisfy the *marginal-cost-equals-price* condition (8), evaluated at $x = z$. Then, given r, (8) and (10) are two equations in z and $n(z)$. Figure 1 displays these relationships in the (n,z) plane. In this plane, equation (10) gives the talent level needed to break even with a firm of employment n, as a function of n; equation (8) gives the talent level at which employment of n is profit-maximizing, also as a function of n. The relationship between these two curves is familiar from its analogue in the Viner theory. They determine a unique employment level $n(z,w,u)$ for the marginal manager: the level which *minimizes* average cost. The managerial talent level $z(w,u)$ which just breaks even at this employment level is the equilibrium cut-off.

One way to attempt to construct an equilibrium solution is as follows. First, think of the problem of maximizing output (3), subject to (1) and (2) for a *fixed* capital-population ratio R and a *fixed* managerial cutoff level z. This is a concave program, the first-order conditions for which are given by (1), (2), (7), and (8). It has a unique solution $n(x,w,u)$, $r(w/u)$ with factor prices $u(z,R)$, $w(z,R)$. To obtain it, one substitutes $n(x,w,u)$ and $r(w/u)$, obtained from (7) and (8), into the constraints (1) and (2) and solves for factor prices as continuous functions of z and R. Call the per capita output Y/N which results from this procedure $H(z,R)$. This much is a standard exercise.

FIGURE 1



Next, we find the cutoff value z which maximizes $H(z,R)$, for each fixed R. The first-order condition for *this* problem is (10), with w and u read as the solutions $w(z,R)$, $u(z,R)$ found as above, and with $n(z)$ and r equal to $n(z,w(z,R),u(z,R))$ and $r[w(z,R)/u(z,R)]$, respectively. The derivatives H_{zz} and H_{zR} needed to check on the uniqueness of a solution for z, and on the way it varies with R, are easy enough to calculate, but not very illuminating. This is not a concave problem. More structure is needed, it appears. In the next section, I shall get it by invoking Gibrat's law.

5. Implications of Gibrat's law

■ The theory of the preceding section is static, determining the size distribution of firms given the economy-wide capital-population ratio. Taken literally, the theory has *no* implications for the pattern of growth of individual firms; a *firm* is defined empirically as a collection of assets, and the matching of managers to asset collections can, according to the theory, change arbitrarily from period to period. That is, the change in assets managed by manager x may or may not be interpreted as growth of an empirically defined firm. In this section, however, I want to think of the model of Section 4 as a limiting case of a model in which there *are* costs of rearranging assets among managers, so that one can think of "manager x" and "firm x," interchangeably. Then the percentage rate of growth of employment in firm x is

$$\frac{d}{dt} \ln [n(x,w(t),u(t))] \tag{11}$$

and of assets

$$\frac{d}{dt} \ln [r(w(t),u(t))n(x,w(t),u(t))]. \tag{12}$$

It should be clear that the theory in Section 4 is consistent with a wide variety of behavior of expressions (11) and (12). A well-known feature of observed patterns in firm growth, however, is the independence of firm growth and size: Gibrat's law, or the law of proportionate effect. Though this law is certainly not implied by the model just developed, there are thus empirical reasons for giving special attention to the special case which satisfies the law. This is the objective of this section.

In the present model, Gibrat's law is the hypothesis that the derivatives with respect to x (size) of expressions (11) and (12) are zero. Carrying out the differentiation (of either one) gives

$$0 = \frac{\partial}{\partial x} \frac{n_w(x,w,u)}{n(x,w,u)} \frac{dw}{dt} + \frac{\partial}{\partial x} \frac{n_u(x,w,u)}{n(x,w,u)} \frac{du}{dt} . \tag{13}$$

If (13) holds for *all* patterns of factor price changes, (13) implies

$$\frac{\partial}{\partial x} \frac{n_w(x,w,u)}{n(x,w,u)} = \frac{\partial}{\partial x} \frac{n_u(x,w,u)}{n(x,w,u)} = 0. \tag{14}$$

It is easily checked that these two statements are in fact the same, so I will develop the wage condition only. Differentiating (8) through with respect to w and solving for n_w yield

$$n_w(x,w,u) = -r_w\left(\frac{w}{u}\right)\frac{g'(f)\phi''(r) + g''(f)[\phi'(r)]^2 n}{g''(f)\phi(r)\phi'(r)} . \tag{15}$$

Dividing by $n(x,w,u)$ and differentiating with respect to x yield

$$\frac{\partial}{\partial x}\frac{n_w(x,w,u)}{n(x,w,u)} = -r_w\left(\frac{w}{u}\right)\frac{\phi''(r)}{\phi(r)\phi'(r)}\frac{\partial}{\partial x}\left[\frac{g'(f)}{ng''(f)}\right], \tag{16}$$

recalling that the capital-labor ratio r does not vary with x. Then (14) and (16) give:

$$0 = n_x(x,w,u)\{n\phi(r)(g'')^2 - g'g'' - n\phi(r)g'g^{(3)}\}. \tag{17}$$

Since $n_x > 0$ for $x \geq z$, (17) requires that the expression in brackets be zero. (Similar development of the capital rental condition in (14) leads to the same conclusion.)

Now $n\phi(r)$ is just the argument of the function g and its derivatives, so that (17) gives a second-order differential equation in the function $g'(v)$:

$$v[g''(v)]^2 - g'(v)g''(v) - vg'(v)g^{(3)}(v) = 0. \tag{18}$$

A convenient change of variable is

$$m(v) = \ln [g'(v)], \tag{19}$$

since (18) and (19) imply

$$\frac{vm''(v)}{m'(v)} = -1. \tag{20}$$

Integrating (20) gives

$$\ln (m'(v)) = \ln (A_1) - \ln (v)$$

or, using (19)

$$\ln\left[\frac{vg''(v)}{g'(v)}\right] = \ln (A_1).$$

Integrating again yields

$$\ln [g'(v)] = A_2 + A_1 \ln (v)$$

and again

$$g(v) = \alpha v^\beta + A_3, \tag{21}$$

where $\alpha = (A_1 + 1)^{-1}e^{A_2}$ and $\beta = A_1 + 1$. We have found, in summary, that a necessary and sufficient condition for the model of Section 4 to obey Gibrat's law—in the sense of (14)—is that g takes the form given by (21). Given assumptions already imposed on g, it must also be the case that $0 < \beta < 1$ in (21), that $\alpha > 0$, and that $A_3 = 0$.

With these additional restrictions on the function g, I shall resume the construction sketched at the close of Section 4.[3] First, (8) is solved for

$$n(x,w,u) = \frac{1}{\phi(r)}\left[\frac{u}{\alpha\beta x\phi'(r)}\right]^{1/(\beta-1)}, \tag{22}$$

where r is given by (7). Next, inserting $n(x,w,u)$ from (22) and $k(x,w,u) = rn(x,w,u)$ into the constraints (1) and (2) gives

$$\left[\frac{u}{\alpha\beta\phi'(r)}\right]^{1/(\beta-1)}\frac{1}{\phi(r)}L(z) = \Gamma(z) \tag{23}$$

[3] Several errors in equations (22)–(27) and in their interpretation appeared in an earlier draft. These were pointed out to me by Dennis Epple, Tsu Yao, and a *Bell Journal* referee. In revising this section, I have made use of some notes of Tsu Yao's which, besides correcting mistakes, suggested considerable simplification.

and

$$r\Gamma(z) = R, \tag{24}$$

where

$$L(z) = \int_z^\infty x^{1/(1-\beta)} d\Gamma(x). \tag{25}$$

Now per capita output Y/N may be expressed as a function $H(z,R)$ of z and R by inserting the employment solution in (22) into equation (3):

$$\begin{aligned}\frac{Y}{N} &= \int_z^\infty xg[f(n(x),k(x))]d\Gamma(x)\\ &= \int_z^\infty x\alpha[n(x,w,u)\phi(r)]^\beta d\Gamma(x)\\ &= \int_z^\infty x\alpha\left[\frac{\alpha\beta x\phi'(r)}{u}\right]^{\beta/(1-\beta)} d\Gamma(x)\\ &= \alpha\left[\frac{\alpha\beta\phi'(r)}{u}\right]^{\beta/(1-\beta)} L(z).\end{aligned}$$

From (23),

$$\left[\frac{\alpha\beta\phi'(r)}{u}\right]^{\beta/(1-\beta)} = \left[\frac{\phi(r)\Gamma(z)}{L(z)}\right]^\beta,$$

so that

$$\frac{Y}{N} = H(z,R) = \alpha(\phi(r))^\beta(\Gamma(z))^\beta(L(z))^{1-\beta}, \tag{26}$$

where $r = R/\Gamma(z)$ from (24).

The equilibrium z will maximize $H(z,R)$ or, equivalently, $\ln H(z,R)$. The first-order condition for this problem (equivalent to (10)) is

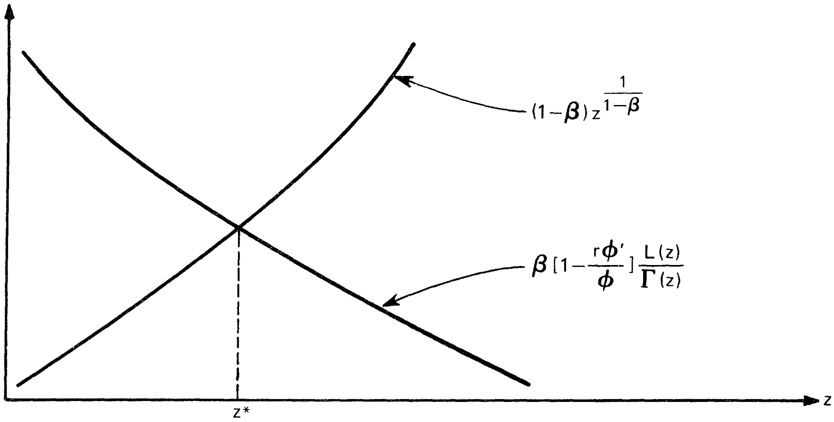
$$\begin{aligned}0 &= \frac{\partial}{\partial z}\left[\beta \ln \phi\left(\frac{R}{\Gamma(z)}\right) + \beta \ln \Gamma(z) + (1-\beta)\ln L(z)\right]\\ &= \frac{\beta}{\phi}\phi'(r)\frac{-R}{\Gamma^2}\Gamma'(z) + \frac{\beta}{\Gamma}\Gamma'(z) + \frac{1-\beta}{L}L'(z)\\ &= \frac{\beta}{\Gamma}\Gamma'(z)\left[1 - \frac{r\phi'(r)}{\phi}\right] - \frac{1-\beta}{L}z^{1/(1-\beta)}\Gamma'(z).\end{aligned} \tag{27}$$

The z-value satisfying (27) equates $(1-\beta)z^{1/(1-\beta)}$ and $\beta[1-(r\phi'(r)/\phi(r))] \times L(z)/\Gamma(z)$. Both functions are drawn in Figure 2. The first passes through the origin, and is increasing and convex. The ratio $L(z)/\Gamma(z)$ tends to infinity as $z \to 0$, and declines monotonically to zero as $z \to \infty$. With (24) the equilibrium r tends to infinity as $z \to 0$ and to R as $z \to \infty$. I shall assume that $1 - r\phi'(r)/\phi(r)$ remains bounded away from both 0 and 1 as these limits are taken, so that the asymptotic behavior of $\beta[1-(r\phi'(r)/\phi(r))]L(z)/\Gamma(z)$ is that of $L(z)/\Gamma(z)$. The derivative of $1 - r\phi'(r)/\phi(r)$ with respect to r is

$$\frac{d}{dr}\left(-\frac{r\phi'}{\phi}\right) = -\frac{\phi'(\phi - r\phi')}{\phi^2}\left[1 - \frac{1}{\sigma}\right],$$

where σ is the elasticity of substitution in production. Hence for the Cobb-

FIGURE 2



Douglas case $\sigma \equiv 1$, $1 - r\phi'/\phi$ is constant with respect to *both* z and R, and a unique solution z^* is determined as shown in Figure 2. Evidently, ln $H(z,R)$ increases to the left of z^* and decreases to the right, so that z^* is a global maximum (though H need not be concave). In this Cobb-Douglas case, the equilibrium cutoff z^* does not vary with R, so that equilibrium average employment per firm, $[1 - \Gamma(z)]^{-1}\Gamma(z)$, would not vary systematically with the per capita wealth of an economy.

For the case $\sigma < 1$ uniformly, the curve $\beta[1 - (r\phi'/\phi)]L(z)/\Gamma(z)$ continues to slope down, as in Figure 2, and shifts to the *right* with increases in per capita capital R. In this case, which considerable evidence[4] suggests as the interesting one empirically, average firm size $[1 - \Gamma(z)]^{-1}\Gamma(z)$ is predicted to be an *increasing* function of per capita wealth. For $\sigma > 1$ this relationship is reversed (provided an equilibrium exists at all).

What is the logic underlying this connection between average firm size and the degree of factor substitutability? Think of an economy, in equilibrium, receiving an addition to its stock of capital. Holding the managerial cut-off z fixed, this new capital would be allocated over existing firms, increasing r as given by (24). This change will affect *both* the wage rate and the return to the marginal manager z. The wage rate as a function of r and z can be found from (23) and (7):

$$w = \alpha\beta\left(\frac{L(z)}{\Gamma(z)}\right)^{1-\beta}\phi^{\beta-1}(\phi - r\phi').$$

Managerial rent as a function of r, z, and employment $n(x)$ is, from (4), (21) and (23),

$$\pi = \alpha(1 - \beta)\left(\frac{L(z)}{\Gamma(z)}\right)^{1-\beta}\phi^{\beta}n(x).$$

[4] Some of this evidence is reviewed and augmented in Lucas (1969). An excellent recent study by Ernst R. Berndt (1976) obtains time series estimates of σ (for U.S. manufacturing) ranging as high as 1.2, depending on the way capital costs are measured. Though Berndt prefers the high estimates, for reasons explained in his paper, I have more confidence in those in his column D, Table 2, p. 65, for reasons given in his note 24, p. 66. These are around 0.5 to 0.8.

Now the ratio of wages to marginal rents is

$$\frac{w}{\pi} = \frac{\beta}{1 - \beta} \frac{1}{n(z)} \left[1 - \frac{r\phi'}{\phi} \right].$$

With an elasticity of substitution less than one, $(d/dr)[-(r\phi'/\phi)]$ is positive, as shown above. That is, an increase in capital per capita *raises* wages relative to marginal managerial rents, or, in other words, raises the opportunity cost of managing relative to the return. This induces marginal managers to become employees, raising the equilibrium z and the average size of firms.

6. Some implications of the theory

■ The model developed in the preceding section is a variant on the production side of a Solow-type growth model. It implies an "aggregate production function" in the sense of a stable relationship between output per capita, Y/N, and capital per capita, R. To obtain this relationship, one inserts the optimal value of z as a function of R (as shown on Figure 2) into (26):

$$\frac{Y}{N} = \int_{z(R)}^{\infty} xg[n(x,R)\phi(r(R))]d\Gamma(x) = F(R). \tag{28}$$

Hence the implications of the theory for the dynamic behavior of aggregates are indistinguishable from the implications of standard, neoclassical growth theory. Whether this should be viewed as a virtue or a defect is, of course, debatable.

The model also "predicts" the full size distribution of firms, but only *given* a distribution of the (probably unobservable) managerial talent. It is more accurate to say that it predicts a particular relationship between the talent distribution and the size distribution of firms. To develop this relationship, take (7) as solved for r as a function of w/u and solve (8) for x:

$$x = \frac{u}{g'[n\phi(r)]\phi'(r)}. \tag{29}$$

Under the Gibrat's law restriction (21) and with (23), equation (29) is specialized to

$$x = \left[\frac{L(z)}{\Gamma(z)} \right]^{1-\beta} n^{1-\beta}. \tag{30}$$

The right side of (29) or (30) gives the talent level which will manage a firm with n employees at a given talent cutoff z and capital per capita R (or at given factor prices).

Now let $S(n)$ be the probability that a randomly selected firm will have fewer than n employees. Then under (30) $S(n)$ will be the probability that x is less than $(L/\Gamma)^{1-\beta}n^{1-\beta}$ *conditional on* $x \geq z$, or

$$S(n) = \frac{\Gamma\left[\left(\frac{L(z)}{\Gamma(z)} \right)^{1-\beta} n^{1-\beta} \right] - \Gamma(z)}{1 - \Gamma(z)} \tag{31}$$

for $n \geq (\Gamma/L)^{1/(1-\beta)}$ and 0 otherwise.

Any stochastic account of the development of managerial talent, leading to a limit argument motivating a particular form for Γ will, using (31), have definite implications for S. Thus, statistical and economic accounts of the size distribution need not be taken as *alternatives*, but can as well be viewed as complementary. Dennis Epple suggested the following illustration of this idea. Let Γ be a Pareto cumulative distribution function:

$$\Gamma(x) = 1 - B^{\rho}x^{-\rho}, \quad x \geq B,$$

where $B < z$. Then (31) implies that $S(n)$ is also a Pareto distribution function:

$$S(n) = 1 - z^{\rho}\left(\frac{L(z)}{\Gamma(z)}\right)^{-\rho(1-\beta)} n^{-\rho(1-\beta)} \tag{32}$$

for $n \geq (\Gamma/L)z^{1/(1-\beta)}$. An advantage of this combined statistical and economic motivation of a Pareto distribution of firm sizes is that there is a clear motivation for the value of the minimal size firm. The arbitrariness in selecting the parameter B in the talent distribution does not carry over to the predicted distribution S of firms.

A tighter implication of the Gibrat's law version of the model concerns the relationship of managerial compensation (or "rent") to firm size, as measured by employees. (This, too, I owe to Dennis Epple.) The general expression for managerial rent is given in (4). As noted in the preceding section, this implies a rent of

$$\alpha(1 - \beta)\left(\frac{L(z)}{(z)}\right)^{1-\beta} \phi^{\beta}n(x),$$

which is proportional to $n(x)$ for all x.[5]

A third implication for empirical testing is the predicted relationship between average firm size and the wealth of the economy. Average firm size is the ratio of employees to managers, or $[1 - \Gamma(z)]^{-1}\Gamma(z)$, which is an increasing function of z. According to the Gibrat's law version of the model, and provided the elasticity of substitution is less than unity, the equilibrium value of z will increase with the capital-labor ratio, R.[6]

Below, I report estimates of the parameters of

$$\ln (M_t) = \theta_0 + \theta_1 \ln (y_t) + \theta_2 t + \epsilon_t, \tag{33}$$

where M_t is employees per firm in the United States in year t; y_t is per capita GNP in constant dollars for the United States; and θ_0, θ_1, and θ_2 are fixed parameters. To obtain efficient estimators of θ_0, θ_1, and θ_2 in the presence of serially correlated residuals, the disturbances ϵ_t were assumed to follow

$$\epsilon_t = \rho_1\epsilon_{t-1} + \rho_2\epsilon_{t-2} + \eta_t, \tag{34}$$

where $\{\eta_t\}$ is a sequence of independent, normal variates with 0 mean and variance σ^2. The estimate of θ_1 will then be interpreted as an estimate of the

[5] This proportionality prediction was also noted by a referee. Available evidence on this point, obtained by D. R. Roberts (1956) and cited (and pointed out to me) by Simon (1957) is not favorable. Roberts finds managerial compensation varying with the *log* of employees. This suggests to me the necessity of incorporating hierarchical considerations, as discussed (but not carried out) in Section 7, below.

[6] Notice that this implication involves only the accuracy of the model in accounting for the *marginal*, or smallest, firm. Thus it is likely that modifications (such as the introduction of hierarchies) which affect mainly larger firms will not alter this implication.

TABLE 1

REGRESSION RESULTS*

ESTIMATES	DEPENDENT VARIABLE, TIME PERIOD		
	CONCERNS, 1900–1970	FIRMS, 1929–1963	MFG. FIRMS, 1929–1963
θ_0 (SE)	13.2 (3.80)	40.2 (7.83)	53.07 (10.59)
θ_1 (SE)	0.838 (0.071)	0.894 (0.119)	0.987 (0.172)
θ_2 (SE)	−0.004 (0.002)	−0.018 (0.004)	−0.024 (0.005)
ρ_1 (SE)	1.16 (0.12)	0.92 (0.19)	1.11 (0.17)
ρ_2 (SE)	−0.30 (0.13)	−0.25 (0.17)	−0.47 (0.14)
R_1^2	0.977	0.954	0.901
R_2^2	0.977	0.947	0.895
DW	1.84	1.78	1.85

*IN EACH COLUMN, THE INDEPENDENT VARIABLE IS ℓn[GNP/LABOR FORCE]. THE DEPENDENT VARIABLE IS ℓn[CONCERN WORKERS/CONCERNS] FOR COLUMN CONCERNS, AND SO FORTH.

sample average value of

$$\left[\frac{RF'(R)}{F(R)}\right]^{-1} \frac{\Gamma'(z)}{\Gamma(z)(1 - \Gamma(z))} z'(R),$$

which should (and does) turn out to be positive.[7]

The high level of abstraction of the theory of Sections 4 and 5 makes it difficult to match theoretically defined firms with firms as variously defined empirically. My strategy was to use as many measures of the number of firms in existence as I could find, and to hope that the choice did not make a great deal of difference. (The ones used are described in Table 2 below.) For each definition of firm, as conformable as possible an employment level was used.

Rather than attempt to measure the capital-labor ratio, R, directly, it seemed safer to use per capita GNP. The two variables are linked theoretically by (28), and using GNP per capita permits thinking of "capital" in the broadest possible terms, measurable or not. Some of the many possible economic motivations for including a trend are discussed in the next section.

Estimates of the parameters of (33) and (34) for annual U.S. data, with three dependent variables M_t, are reported in Table 1. Table 2 describes the sources and the construction of variables used. Estimates were obtained using Durbin's (1960) two-step method. Estimates of ρ_1 and ρ_2 are from the first step; R_1^2 is the R^2 from this step. Estimates of θ_0, θ_1, and θ_2 are from the second step (which is just regression (33) with ln (M_t) and ln (y_t) filtered by $1 - \hat{\rho}_1 L - \hat{\rho}_2 L^2$, where L is the lag operator). R_2^2 is the fraction of the variance of *original*

[7] This did not come as a surprise to me, nor will it to any economist with any familiarity with U.S. business population statistics. One might prefer to view the "test" results reported below as bearing on the size of the elasticity of substitution.

TABLE 2

SOURCES AND VARIABLE CONSTRUCTION

RAW SERIES	
CONCERNS	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES V–20. THOUSANDS OF FIRMS, AS DEFINED BY DUN & BRADSTREET.
	– EXCLUDES FINANCE, REAL ESTATE, INSURANCE, RAILROADS, AMUSEMENTS, MANY ONE–MAN SERVICES, PROFESSIONS, FARMERS.
FIRMS	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES V–13. THOUSANDS OF FIRMS, FROM SOCIAL SECURITY AND IRS SOURCES.
	– EXCLUDES SELF–EMPLOYED WITH NO EMPLOYEES, PROFESSIONS, FARMERS.
MFG. FIRMS	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES V–15. THOUSANDS OF MANUFACTURING FIRMS: SAME SOURCES AS "FIRMS."
GNP	– GNP IN $BILLIONS 1958.
CONCERN WORKERS	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES D–127 LESS D–137 LESS D–139 LESS RR WORKERS (SEE BELOW). THOUSANDS OF WORKERS, CONFORMING AS CLOSELY AS POSSIBLE TO EMPLOYMENT IN CONCERNS (DEFINED ABOVE).
RR WORKERS	– THOUSANDS OF RAILWAY WORKERS, CONSTRUCTED FROM SERIES IN U.S. BUREAU OF LABOR STATISTICS (1976) AND U.S. DEPARTMENT OF COMMERCE (1975).
FIRM WORKERS	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES D–127 LESS D–139. THOUSANDS OF WORKERS.
MFG. WORKERS	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES D–130. THOUSANDS OF WORKERS.
LABOR FORCE	– FROM U.S. DEPARTMENT OF COMMERCE (1975), SERIES D–1 FOR 1900–1947 AND SERIES D–12 X D–1 (1947)/D–12 (1947) FOR 1948–1970.

(that is, unfiltered) dependent variable which is explained by (33)–(34) with estimates obtained as just described. *DW* is the Durbin-Watson statistic from the second step.

The results with all three choices of dependent variable show a clear and accurately measured effect of GNP (or wealth) on average firm size. The estimated elasticity is in the range 0.8–1.0. The independent effect of trend on firm size is negative, with firm size declining (GNP held fixed) at an annual rate of about 0 to 2.5 percent. Together, these two variables do an excellent job of accounting for the secular behavior of firm size (however measured) and for large-scale cyclical movements (mainly the Great Depression).

7. Elaborations of the model

■ The extreme simplicity of the theory developed in Sections 4 and 5 matches very well the crude and aggregative data used in Section 6. If one were to attempt tests on finer data (say, quantitative histories of individual firms or entrepreneurs) it would be evident that refinement in several directions would be necessary. In this section, I shall briefly discuss two refinements that seem important and promising: elaborations to consider human capital and hierarchical managements.

The employment variable theoretically used in Section 4 was for labor of constant quality. The actual variable used in the regressions of Section 6 is for labor of improving quality. The simplest way to correct for this would be to imagine that employee and managerial quality are improving at possibly different, fixed exponential rates. This introduces a trend term (of indeterminate effect) into (33), but has no other effect on the model.

More interesting human capital considerations are suggested by commonly known features of managerial careers: people tend to move from employee to managerial status later in their careers (as opposed to immediately upon entry to the workforce, as predicted by the theory above); those that make this transition tend to be among the most skilled employees. These facts suggest the existence of a kind of human capital which is productive both in managing and in working for others, and which is accumulated most rapidly as an employee. Without trying to formulate a model embodying these features exactly, one can guess that in place of an equilibrium cutoff talent level z for managers, one would find an equilibrium *function* $\tau(x)$, the value of which is the age at which a person of talent x begins to manage. If so, and if $F(s)$ is the fraction of the workforce younger than age s, then the fraction of the workforce managing in equilibrium is

$$\int_{x_0}^{\infty} \int_{\tau(x)}^{\infty} d\Gamma(x) dF(s),$$

where x_0 is the lowest talent level ever to manage. The case studied in Sections 4 and 5 was $x_0 = z$, $\tau(x) \equiv 0$, in which case this expression was $1 - \Gamma(z)$: a decreasing function of capital per capita, R. It is likely that the equilibrium values of x_0 and $\tau(x)$ both would increase with R, so that the aggregative implications of this more elaborate theory would not differ from those of the simple version tested. On more interesting samples, however, a more refined theory along this line might be worth developing in full.

The managerial "technology" postulated in Section 4 was more complex than that usually postulated in economic models (where labor and capital are somehow costlessly combined) but considerably simpler than the complex hierarchies observed in actual business firms. To capture some of this complexity, and to broaden the term manager to include others than top executives, one would need to postulate a technology for managing managers. (See Tuck (1954) and Beckman (1977).) Again, without attempting to carry out such a construction, one can guess as to its likely implications for the aggregative data used in Section 6, and for other samples.

With diminishing returns on span-of-control built in at each managerial level, an equilibrium with multilevel management should consist of an allocation of capital and product workers over "level one" managers, an allocation of these units over level two managers, and so forth, until all resources are distributed over the top level managers. In such a scheme, it seems likely that the theory of firm size would be similar to that developed in Sections 4 and 5, with "managers" reinterpreted as "top level managers" and, of course, with considerably more complex substitution possibilities in production. Fruitful analysis along these lines would, I think, require more specific thinking about what it is, exactly, that managers do.

8. Conclusions

■ On a recent vacation in Quebec, my family and I stopped for lunch at a small, inexpensive restaurant on the St. Lawrence River. The decor, the menu, and the service in this family-run place were unique to it, and reflected a large number of managerial decisions, all solved in a way reflecting both the tastes of the owners and local prices of food and other materials. Even politics was involved: the flags of Canada and the United States flew out front; the flag of Quebec was absent. The Quebec economy is developing rapidly, however, and should we return in ten years we shall find, I imagine, a Poulet Frit Kentucky outlet in its place, with decor, menu, services, and politics identical to its twins in Montreal or Quebec City. This will occur, according to the theory developed above, because rising real wages will make working for some one else more lucrative than the return to making managerial decisions for a single, small restaurant.

If one could not think of dozens of examples of specific economic activities of this sort which, over time, are being carried out in larger and larger firms (though not necessarily larger plants or stores) one would, I suspect, have little confidence in theorizing and testing at the aggregative level of this paper. But anyone *can* think of many such examples, and to many observers, this trend to bigness seems ominous. In this paper, I have tried to show one way in which this trend (like so many other changes which seem ominous to people) can arise as a natural consequence of increasing wealth in a competitive economy. If correct, the theory implies that policies to modernize backward economies by consolidation and policies to resist conglomerates in advanced economies may both be misguided—for exactly the same reason.

References

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

BERNDT, E. R. "Reconciling Alternative Estimates of the Elasticity of Substitution." *Review of Economics and Statistics*, Vol. 58 (February 1976), pp. 59–68.

DURBIN, J. "Estimation of Parameters in Time Series Regression Models." *Journal of the Royal Statistical Society* (Series B) (1960).

IJIRI, Y. AND SIMON, H. A. "Business Firm Growth and Size." *The American Economic Review*, Vol. 54, No. 2 (March 1964), pp. 77–89.

KIHLSTROM, R. E. AND LAFFONT, J. J. "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion." Unpublished working paper, July 1977.

LUCAS, R. E., JR. "Adjustment Costs and the Theory of Supply." *Journal of Political Economy*, Vol. 75, No. 4, Part 1 (August 1967), pp. 321–334.

———. "Labor-Capital Substitution in U.S. Manufacturing" in A. C. Harberger and M. J. Bailey, eds., *The Taxation of Income from Capital*, Washington: The Brookings Institution, 1969, pp. 223–274.

MANNE, H. G. "Mergers and the Market for Corporate Control." *Journal of Political Economy*, Vol. 73, No. 2 (April 1965), pp. 110–120.

ROBERTS, D. R. "A General Theory of Executive Compensation Based on Statistically Tested Propositions." *Quarterly Journal of Economics*, Vol. 20 (1956), pp. 270–294.

SIMON, H. A. "The Compensation of Executives." *Sociometry*, Vol. 20 (March 1957), pp. 32–35.

——— AND BONINI, C. P. "The Size Distribution of Business Firms." *The American Economic Review*, Vol. 48, No. 4 (September 1958), pp. 607–617.

STIGLER, G. J. "The Case against Big Business" in E. Mansfield, ed., *Monopoly Power and Economic Performance*, New York: W. W. Norton, 1964.

TUCK, R. H. *An Essay on the Economic Theory of Rank*. Oxford: Basil Blackwell, 1954.

U.S. BUREAU OF LABOR STATISTICS. *Employment and Earnings Statistics for the U.S.* Washington, D.C.: U.S. Government Printing Office, 1976.

U.S. DEPARTMENT OF COMMERCE. *Historical Statistics of the United States: Colonial Times to 1970*. Washington, D.C.: U.S. Government Printing Office, 1975.

VINER, J. "Cost Curves and Supply Curves." *Zeitschrift für Nationalökonomie*, Vol. 3 (1932), pp. 23–46.