FIRM SIZE AND OPTIMAL GROWTH RATES*

Uzi SEGAL
University of Toronto, Toronto, Ont., Canada M5S 1A1

Avia SPIVAK
Ben Gurion University, Beer Sheva, Israel

Received October 1986, final version received September 1987

This paper presents a theoretical model in which, due to dissolution costs, the rate of growth of small firms tends to be higher and more variable than that of larger firms. This model also predicts that for large firms the rate of growth is fixed, as claimed by Gibrat's Law.

1. Introduction

One of the major decisions a firm must make is how to allocate the profits between dividends and retained earnings. Retained earnings reinvested in the firm provide for future growth. The rate by which the firm grows is thus an endogenous decision variable, arrived at as a result of intertemporal maximization.

This consideration is entirely overlooked in discussions of growth rate and firms' size. Simon, who contributed significantly to this literature (see Simon and Bonini (1958) and the references therein), adopted Gibrat's Law, the assumption that a single firm's rate of growth is independent of its size. He proceeded to obtain the stationary size distribution consistent with this assumption.

This paper suggests to regard the growth rate as a solution to an explicit intertemporal maximization problem. In order to avoid the agency problem, it is assumed that the firm is owned by a single shareholder who is also the manager. The owner is risk-neutral and seeks to maximize the expected discounted value of his income from the firm. The firm's growth rate is

*We are grateful to Siegfried Berninghaus, Richard Kihlstrom, Ariel Pakes, Joel Sobel, and John Sutton for their comments.

stochastic and diversified. Increasing reinvestment in the firm changes the parameters that govern the stochastic growth process so that the expected growth rate increases as well. Reinvestment of profits is thus beneficial on two counts: First, it increases the expected size of the firm and hence the total expected future profits. Secondly, bigger firms are less likely to go out of business, because the growth is diversified. Hence, investment in growth decreases the expected dissolution costs of the firm. Our modeling thus introduces increasing returns to scale, not because of technology, which shows constant returns to scale, but because units of the firm insure each other mutually against the disappearance of the firm.

Since there are scale economies, one may suspect that the rate of growth will change with the size of the firm. This indeed is found to be true. Our major result is that smaller firms on the average grow faster than bigger firms. We also show that when the size of the firm tends to infinity, the growth rate converges to a positive number. In other words, large firms satisfy Gibrat’s Law, that the firm grows at a fixed rate.

Jovanovic (1982) explained the same empirical phenomenon with a different approach. Firms enter an industry not knowing their true cost function but only the industry’s average. Every period they use the observation on the noisy cost to update their estimate of cost and their optimal production decisions. So, the more efficient firms will learn and their output will grow. The other, less efficient firms, will produce less and less and eventually will exit the industry. In this model it turns out that rates of growth for smaller firms are larger and more variable than those of bigger firms. An entirely different theoretical approach, that of Lucas (1978), determined the size and growth of firms by the optimal managerial span of control.

The empirical evidence supports our theoretical finding that for small firms growth rates slow down with size and that for large firms growth rates tend to be fixed. Mansfield (1962) found an inverse relationship between size and rate of growth for a sample of small firms. Pasigian and Hymer (1962) and Hart and Prais (1956) found for their samples of large firms that growth rates were independent of size. The most exhaustive study to date was done by Bronwyn Hall (1986). She used panel data on publicly traded firms in the U.S. manufacturing sector, following more than nine hundred firms from 1972 until 1983. She found that smaller firms do grow faster. One of her major contributions is the correction for the sample selection problem, arising from the fact that smaller firms who do not grow fast are more likely to go bankrupt and hence be excluded from the sample. Hall’s data allowed her to estimate the probability of leaving the sample and hence correct for this bias. The finding that smaller firms grow faster holds both with or without the sampling bias correction.

The rest of this paper is organized as follows. Section 2 presents the model.
Section 3 considers optimal growth under the assumption that the rate of reinvestment in the firm is endogenous and concludes the paper.

2. The model

We assume a firm whose units are stochastically identical and independent. The number of units at time $t$ is $N(t)$, and is assumed to be a random variable. Denote by $N$ the initial number of units, $N = N(0)$. Each unit has a $\lambda \, dt$ probability of changing in the following time interval $dt$ into two units and a $\mu \, dt$ probability of disappearing, $\lambda, \mu \geq 0$. Thus, the model is assumed a continuous time model. Each unit generates an income stream of $y$ as long as it exists. The number of units is changing at random discretely up and down by one unit at a time, the probability of two units changing simultaneously is zero. If the number of units goes down to zero, the firm goes out of business.

A major assumption of this model is that the firm's owners suffer a loss of $L$ when the firm dissolves. These dissolution costs are the intangibles that are implied in the existence of the firm such as reputation, the ability of people to work as a team, etc. $L$ is related to the cost of putting a new firm together and acquiring a reputation. The risk-neutral owners of the firm are interested in the expected value of the profits, discounted at a positive discount factor $\delta$ ($\delta > \lambda - \mu$ to ensure boundedness of this value).

Initially, we assume that $\lambda$, $\mu$, and $y$ are given to the firm. Given that values, the expected discounted income stream is made up of two parts: expected discounted income and expected discounted loss. We start the discussion with a firm of one unit, so $N = 1$ at $t=0$. The size of the firm as a function of time, $N(t)$, is a random variable. The expected discounted value of the income stream is $Y = E[\int_0^\infty e^{-\delta t} y N(t) \, dt] = \int_0^\infty e^{-\delta t} y E[N(t)] \, dt$. Following Harris (1963, p. 104), $E[N(t)] = e^{-(\mu - \lambda) t}$. (Note that when $\mu = \lambda$, $E[N(t)] = 1$ and when $\lambda > \mu$, $E[N(t)]$ grows exponentially with time.) Substituting and integrating with respect to time, with the assumption $\delta > \lambda - \mu$, we obtain $Y = y/(\delta + \mu - \lambda)$. Since the $N$ units grow independently, the total expected discounted income is simply $NY$. Thus, regarding the income aspect, we have constant returns to the size of the firm.

Let $\phi = \phi(\lambda, \mu, \delta; N)$ be the expected discounted loss. Assume $N$ units and let $F_N(t)$ be the distribution function for the event that all $N$ units disappear exactly at or before time $t$. Denote by $f_N(t)$ the corresponding density function. If this function is known, then $\phi = \int_0^\infty I e^{-\delta t} f_N(t) \, dt$. Because units are independent, $F_N(t) = \xi^N$, where $\xi$ is the probability that one unit will disappear before or at time $t$. Referring again to [Harris (1963, p. 104)],

$$
\xi = \frac{1 - e^{-t(\mu - \lambda)}}{1 - (\lambda/\mu) e^{-t(\mu - \lambda)}} \quad \text{for} \, \lambda \neq \mu \quad \text{and} \quad \xi = \frac{\lambda t}{1 + \lambda t} \quad \text{for} \, \mu = \lambda.
$$
Using integration by parts and the fact that $F_N(0)=0$, one obtains

$$
\phi = \begin{cases} 
L \delta \int_0^\infty e^{-\lambda t} \left[ \frac{1 - e^{-(\mu - \hat{\lambda})}}{1 - (\lambda/\mu) e^{-(\mu - \hat{\lambda})}} \right] dt & \text{for } \lambda \neq \mu, \\
L \delta \int_0^\infty e^{-\lambda t} \left[ \frac{\lambda t}{\lambda t + \mu} \right] dt & \text{for } \lambda = \mu.
\end{cases}
$$

When $\delta$ is higher, future loss is less important. An increase in $\lambda$ decreases the death probability $\zeta$ and so does a decrease in $\mu$. Formally,

$$
\frac{\partial \phi}{\partial \delta} = -L \int_0^\infty te^{-\lambda t} f_N(t) dt < 0, \quad \frac{\partial \phi}{\partial \lambda} < 0 \quad \text{if} \quad \frac{\partial}{\partial \lambda} \left[ \frac{e^{(\lambda - \mu)}}{(\lambda/\mu) e^{(\lambda - \mu)} - 1} \right] < 0,
$$

which follows from the fact that $1 - x - e^{-x} < 0$ for all $x \neq 0$. A similar argument shows that $\partial \phi / \partial \mu > 0$. When $\lambda$, $\mu$, and $\delta$ are multiplied by $\alpha$ and $\chi = \alpha t$ is substituted, the same $\phi$ integral is obtained, hence $\phi$ is homogeneous of degree 0 in $\lambda$, $\mu$, and $\delta$. This amounts to saying that $\phi$ is invariant to the units by which the time is measured.

Claim 1. $\phi$ is declining and convex in $N$, $\lim_{N \to \infty} \phi = 0$, and $\lim_{N \to \infty} N \phi = 0$.

Proof. $\partial \phi / \partial N = L \delta \int_0^\infty e^{-\delta t} \zeta^N \ln \zeta dt < 0$ because $\zeta < 1$ for all $t$. $\partial^2 \phi / \partial N^2 = L \delta \int_0^\infty e^{-\delta t} \zeta^N (\ln \zeta)^2 dt > 0$. To show that $\phi = L \delta \int_0^\infty e^{-\delta t} \zeta^N dt$ and $N \phi$ tend to zero as $N$ tends to infinity, we use the facts that for all $t$, $\zeta < 1$, $e^{-\delta t} \to 0$ as $t \to \infty$, and $\zeta$ is nondecreasing in $t$. It follows that the sequence of functions $e^{-\delta t} \zeta^N$ converge to the zero function uniformly in $t$, hence $\phi \to 0$ as $N \to \infty$. A similar proof holds for $N \phi$ by noting that $N \zeta^N \to 0$ as $N \to \infty$. Q.E.D.

The last property will be very useful in the sequel because it characterizes the returns to scale that the firm possesses. Note that the only assumption used here is that the probability of disappearance $\zeta$ is less than one. This assumption must be satisfied in every stochastic model. Note that even if the loss from bankruptcy was proportional to size, we would still get the same returns to scale properties.

From the above analysis we know that the value of a firm of size $N$ to its owners equals $NY - \phi(N)$. Since $\phi(N)$ is convex and diminishing to zero, $NY - \phi(N)$ is concave and asymptotic to the straight line $NY$. Of course, for $N=0$, $NY - \phi(N) = -L$. The interpretation is that when $N=0$ the firm is dissolved and must pay the cost $L$. We see that a sufficiently large firm has the same value to its owners in the presence or absence of dissolution cost.
This suggests that the case $L=0$ might be a useful benchmark for the behavior of firms.

3. Optimal growth

The firm can change the parameters $\lambda$ and $\mu$ by plowing back more or less of its income, thus changing the dividend stream $y$. We assume that each existing unit creates a total income stream of $\bar{y}$. The firm retains a stream of $g$ that is reinvested in the firm to obtain growth and distributes dividends $y=\bar{y}-g$. It is assumed that the sum $b=\lambda+\mu$, which governs the chances of one unit to change, is constant. Thus, an increase in $g$ increases $\lambda$ according to the relationship $g=g(\lambda)$ and decreases $\mu$ by the same magnitude. We assume that $g'^{>0}$, $g''^{>0}$, $\lim_{\lambda \to 0} g'(\lambda) = 0$, and $\lim_{\lambda \to b} g'(\lambda) = \infty$. Because the units are independent, the parameters and income stream may, in principle, vary across different units. However, because of symmetry and the convexity of $g$, the decision for all the units will be identical.

When $L=0$, the expected discounted profits of the firm are $NY = N \bar{y} / (\delta + \mu - \lambda)$. Substituting $y=\bar{y}-g(\lambda)$ and $\mu=b-\lambda$, the problem becomes $\max_{\lambda} N(\bar{y}-g(\lambda)) / (\delta + b - 2\lambda)$. The first order conditions for maximum are

$$N(2(\bar{y}-g(\lambda)) - g'(\lambda)(\delta + b - 2\lambda)) / [\delta + b - 2\lambda] = 0 \Rightarrow$$

$$g'(\lambda) = 2(\bar{y}-g(\lambda)) / (\delta + b - 2\lambda).$$

Because $\delta + b - 2\lambda = \delta + \mu - \lambda > 0$, a sufficient condition for maximum is $g''(\lambda) > 0$. This condition may take an alternative form. Denote by $g^*$, $\lambda^*=\lambda(g^*)$, and $\mu^*=b-\lambda^*$ the solution to the maximum problem and $a^*=(\bar{y}-g^*) / (\delta + \mu^* - \lambda^*)$. Then at $g^*$,

$$g'(\lambda^*) = 2a^*. \quad (1)$$

As expected, the optimal growth decision is independent of $N$ and the maximum value of the firm is simply $a^* N$.

Consider now the case where $L>0$, and denote by $V(N)$ the maximum expected discounted profit that a firm of size $N$ can attain, if it chooses a different optimal $\lambda$ for different sizes of $N$. By definition, $V(0) = -L$. Given the choice of $\lambda$, a firm starting with $N$ operating units will have a stream of income $NY = N[\bar{y}-g(\lambda)]$ until the firm changes its size to either $N+1$ or $N-1$. Denoting the random date of the change by $t_1$, the expected discounted value of the firm is

$$W(N, \lambda) = E_{t_1} \left[ \int_0^{t_1} Ne^{-\delta t} [\bar{y}-g(\lambda)] dt + (\lambda / (\lambda + \mu)) e^{-\delta t_1} V(N+1) + (\mu / (\lambda + \mu)) e^{-\delta t_1} V(N-1) \right].$$
The first part of this sum is the discounted value of the income stream that the $N$ units produce. In the second part, $\lambda/(\lambda + \mu)$ is the probability that the firm will grow in one unit, conditional upon the occurrence of a change. Because we have a continuous time model, the change affects one unit only, as the probability of more than one unit changing at exactly the same time is zero. The second and third parts are thus the discounted values of $N+1$ and $N-1$ units, weighted by their conditional probabilities. $W$ depends on $\lambda$ because the distribution of $t_1$ depends on $\lambda$, $V(N+1)$ and $V(N-1)$ are given numbers, and $\mu = b - \lambda$. If $W(N, \lambda)$ is maximized with respect to $\lambda$, $V(N)$ is obtained. So

$$V(N) = \max_{\lambda} W(N, \lambda) = \max_{\lambda} E_{t_1} \left[ \int_0^{t_1} Ne^{-\delta t} [\bar{y} - g(\lambda)] \, dt + \frac{\lambda}{\lambda + \mu} e^{-\delta t_1} V(N + 1) + \frac{\mu}{\lambda + \mu} e^{-\delta t_1} V(N - 1) \right].$$

We first calculate $W(N, \lambda)$ and then apply the optimization. Noting that

$$\int_0^{t_1} Ne^{-\delta t} [\bar{y} - g(\lambda)] \, dt = \frac{N}{\delta} (1 - e^{-\delta t_1}) [\bar{y} - g(\lambda)],$$

we obtain

$$W(N, \lambda) = E_{t_1} \left[ \frac{N}{\delta} (1 - e^{-\delta t_1}) [\bar{y} - g(\lambda)] + \frac{\lambda}{\lambda + \mu} e^{-\delta t_1} V(N + 1) + \frac{\mu}{\lambda + \mu} e^{-\delta t_1} V(N - 1) \right].$$

The probability distribution of $t_1$ is the solution to the following well-known elementary problem: given $N$ light bulbs whose life is distributed exponentially with a parameter $\beta$, what is the probability distribution of the time $t_1$ elapsing until the first bulb dies out? The cumulative distribution of this problem is $F(t_1) = 1 - \left[ 1 - \int_0^{t_1} be^{-bt} \, dt \right]^N = 1 - e^{-bNt_1}$, and the density function is $f(t_1) = Nbe^{-bNt_1}$. We can now calculate $E_{t_1} [e^{-\delta t_1}] = \int_0^{\infty} e^{-\delta t} f(t_1) \, dt_1 = \int_0^{\infty} e^{-\delta t} Nbe^{-bNt_1} \, dt_1 = Nb/(\delta + Nb)$. Substituting into (2) one obtains

$$W(N, \lambda) = \frac{N}{\delta + Nb} [\bar{y} - g(\lambda) + \lambda V(N + 1) + (b - \lambda) V(N - 1)].$$

The optimal $\lambda$ satisfies the following first order conditions:
CJ. Segal and A. Spivak, Firm size and optimal growth rates

\[ g'(\lambda) = V(N + 1) - V(N - 1), \]  

(4)

\[ g''(\lambda) > 0 \] is a sufficient condition for a maximum.

The major property that we want to establish here is the concavity of \( V \) in \( N \). This is likely to be present because the loss becomes less likely when size increases, and the incremental improvement eventually goes down to zero. As we will see later, the concavity implies that the growth rate decreases with size.

**Claim 2.** \( V(N) \) is strictly increasing.

**Proof.** Suppose that instead of \( N \) the firm has \( N + 1 \) units. If it operates \( N \) units with the former policy and the \( N + 1 \)th unit with some \( \lambda \) such that \( y > g(\lambda) \), it will have an additional expected income. Its expected loss will go down because the expected loss decreases in \( N \). Since it has a feasible policy that creates more value, and it may now choose a better one, \( V(N + 1) > V(N) \). \( \text{Q.E.D.} \)

**Claim 3.** \( a^*N - \phi(N, \lambda^*) \leq V(N) < a^*N \). Hence, \( \lim_{N \to \infty} V(N) - a^*N = 0 \). (The right hand inequality holds only for \( L > 0 \).)

**Proof.** The first inequality follows from the fact that \( V(N) \) is optimal for the loss \( L \) and \( a^*N - \phi(N, \lambda^*) \) is available. The second inequality follows from optimality of \( a^*N \) for the no loss case, and from the fact that the loss decreases the income stream. The limiting results follow from \( \lim_{N \to \infty} \phi(N, \lambda^*) = 0 \). \( \text{Q.E.D.} \)

According to Claim 3, the \( V \) function is bounded between two concave functions (see Claim 1 in section 2). One is thus inclined to believe that \( V \) is itself concave. This more difficult result is proved in the next claim.

**Claim 4.** \( V(N) \) is a concave function. It is strictly concave when \( L > 0 \).

**Proof.** We wish to show that for all \( N \geq 1 \), \( V(N + 1) - V(N) < V(N) - V(N - 1) \) and equivalently, that \( 2V(N) - V(N + 1) - V(N - 1) > 0 \). \( V(N) \geq W(N, \lambda^*) \), thus it is sufficient to show that \( 2W(N, \lambda^*) - V(N + 1) - V(N - 1) > 0 \). By eq. (3),

\[ W(N, \lambda^*) = \frac{N}{\delta + Nb} (\bar{y} - g^*) + \frac{\lambda^*N}{\delta + Nb} V(N + 1) + \frac{(b - \lambda^*)N}{\delta + Nb} V(N - 1), \]

hence
2W(N, \lambda^*) - V(N + 1) - V(N - 1)

= \frac{N\{2(y - g^*) + (2\lambda^* - b)[V(N + 1) - V(N - 1)] - \delta[V(N + 1) + V(N - 1)]\}}{\delta + Nb}.

When a firm with \(N - 1\) units receives two more units, it can always operate these two units with the strategy \(\lambda^*\), creating an extra discounted expected income of \(2(\bar{y} - g^*)/(\delta + b - 2\lambda^*)\). The two extra units also decrease the expected loss, hence the maximum value \(V(N + 1)\) satisfies

\[V(N + 1) \geq V(N - 1) + \frac{2(\bar{y} - g^*)}{\delta + b - 2\lambda^*}.\]  

(5)

By Claim 3, \(V(N + 1) < a^*(N + 1)\) and \(V(N - 1) < a^*(N - 1)\), hence

\[V(N + 1) + V(N - 1) < 2a^*N = 2N \frac{\bar{y} - g^*}{\delta + b - 2\lambda^*}.\]  

(6)

Using (5) and (6) one obtains

\[2W(N, \lambda^*) - V(N + 1) - V(N - 1)

= \frac{1}{\delta + Nb}\left\{N\left[2(\bar{y} - g^*) + (2\lambda^* - b)\frac{2(\bar{y} - g^*)}{\delta + b - 2\lambda^*}\right] - 2\delta N \frac{\bar{y} - g^*}{\delta + b - 2\lambda^*}\right\} = 0.

Q.E.D.

We now have all the elements to show that the optimal \(\lambda^*\) declines with \(N\).

**Theorem 1.** When \(L > 0\), the optimal growth rate \(\lambda(N)\) is monotonically decreasing with \(N\) and converging to the no loss growth rate \(\lambda^*\).

**Proof.** By (4), \(\lambda(N)\) is defined by the condition \(g'(\lambda(N)) = V(N + 1) - V(N - 1)\) or equivalently, \(\lambda(N) = (g')^{-1}[V(N + 1) - V(N - 1)]\). The inverse function \((g')^{-1}\) exists, is continuous, and is monotonically increasing because \(g'' > 0\). By Claim 4, \(V(N + 1) - V(N - 1)\) in monotonically decreasing with \(N\), hence \(\lambda(N)\) is monotonically decreasing. A simple application of Claim 3 yields

\[\lim_{N \to \infty} [V(N + 1) - V(N - 1)] = 2a^*.\]

Since \((g')^{-1}\) is continuous and since by eq. (1), \((g')^{-1}(2a^*) = \lambda^*\), it follows that \(\lim_{N \to \infty} \lambda(N) = \lambda^*\).  

Q.E.D.

Theorem 1 states that the inverse relationship between size and rate of growth becomes weaker as the size of the firm increases. In other words, for
sufficiently large firms the growth rate is independent of size. This result is known as Gibrat's Law, thus the Theorem shows that Gibrat's Law holds for large firms. However, the main result of the paper is that growth rate declines with size.

The results of this paper depend crucially on the existence of dissolution costs. Our concept of dissolution costs includes a subset of the bankruptcy costs discussed in the financial literature. Its significance was discussed in this literature because of its bearing on the debt-equity decisions [Modigliani-Miller (1963)]. There is an ongoing debate about the importance of bankruptcy cost. Robinchek and Myers (1966), Baxter (1967), and others underscored their importance while Haugen and Senbet (1978), Miller (1977), and Warner (1977) claimed that bankruptcy costs are rather insignificant. [See also Ang et al. (1982)]. We do not have the administrative costs of selling the firm's assets to pay for its loans because we do not allow debts. Our dissolution cost is the loss of intangibles of the firm such as reputation, organization, etc., related to the cost of putting a new firm together.

References