Division of Research
School of Business Administration
The University of Michigan

1st draft: November 1990
this draft: October 1991

RETURNS TO SCALE WHEN PRODUCTION IS A PROBABILITY

Working Paper #665

Kerk L. Phillips
The University of Michigan

Kerk L. Phillips, School of Business Administration, University of Michigan, Ann Arbor, MI 48109-1234,
email: UserLM21@UMICHUM

This paper evolved out of research for the third chapter of my dissertation at the University of Rochester. I have benefited from discussions with many people including: Ron Jones, Barbara Mace, Scott Masten, Ken McLaughlin, Ric Pace, Craig Pirrong, Stephanie So, Alan Stockman, Valerie Suslow, and the informal reading group on growth at the University of Rochester. Any errors are mine and I apologize for them profusely.

Abstract

In this paper I consider the case where factor inputs are productive because they increase the probability of obtaining a reward. This type of production is intended to capture R&D, but is applicable to any case where a benefit of fixed value is targeted. I show that production functions with homogeneity of degree one imply synergies, i.e. large firms have a cost advantage over small firms. It follows that production functions with no synergies exhibit decreasing returns to scale. In this case assumptions of industrial organization will affect aggregate production. I show that in some cases free entry can lead to overproduction.

1. Introduction

What is meant by constant returns to scale? The precise definition is that the production function is homogeneous of degree one in all inputs. In practical terms we usually think in terms of the implications of this homogeneity; that average cost is unrelated to firm size. Constant returns to scale is thought of as yielding competitive markets and implies that while aggregate production may be determinate, production by any given firm in a market is not.

We usually think of production as yielding a measurable quantity of goods. Inputs are productive because increasing them increases the number of goods produced. In many situations, however, it is more appropriate to think of production as yielding a probability. Inputs are productive not because they increase the number of goods, but because they increase the probability a reward will be obtained. This seems especially applicable in the case of R&D. Recent models of endogenous growth have focused on monopolistic competition where firms engage in R&D races[1]. If successful firms have a design in hand that allows them to undercut competitors and capture an entire market. In the quality ladders literature the expected net present value of monopoly profits conditional on having a design in hand does not depend on the resources previously expended on R&D. This implies that inputs into R&D are productive only because the raise the probability of success.

What does returns to scale mean in this context? If the probability of success is modelled as homogeneous of degree one large firms will have a cost advantage over small firms. Suppose there is a firm that hires inputs such that is has a 50% chance of success. With homogeneity of degree one if the firm were to double in

---

1 see especially Grossman & Helpman (1991a & 1991b) and Segerstrom (1991)

size it would have a 100% chance of success. If a second firm identical to the first were to enter its probability of success would be 50% also, but the aggregate probability of success would be 75%. This example, formalized below, illustrates that constant returns to scale in this context implies synergy or an efficiency gain to large firms.

In some cases it may be reasonable to allow synergies in the production process, just as it may be reasonable to assume increasing returns to scale for some types of production. The concept of duplicability, that output can be doubled if all inputs are duplicated is a powerful reason for modelling most production as constant returns to scale. In the context of a probability duplicability implies decreasing returns to scale, but no synergies. Unless there are strong reasons for assuming synergies, therefore, it seems appropriate to model the production of probabilities with decreasing returns to scale. This will open up a variety of issues related to industrial organization that are avoided with constant returns to scale.

Randomness in the production function is not new. For example, Brock & Mirman (1972) consider random "productivity shocks." Much of the existing literature on R&D incorporates randomness into production. Dixit (1988), for example, uses continuous time and models uncertainty via a hazard function. However, this paper is unique in that it focuses explicitly on the issues of returns to scale and their impact on market organization.

The remainder of this paper is organized as follows. Section 2 discusses the relation between returns to scale and synergy in greater detail. I demonstrate that constant returns to scale implies synergies as it implies the avoidance of redundancies. I also note that in continuous time constant returns to scale does not imply synergies as any redundancies are second-order small. Section 3 discusses the industrial organization implied by a production function that exhibits no synergies. The industrial organization that results depends on how benefits are divided among firms when more than one is successful.

2. What is Constant Returns to Scale?

Let me begin this section by defining returns to scale and synergy. I will use the common mathematical definitions of returns to scale. That is increasing returns to scale is homogeneity of degree greater than one in all inputs, constant returns is degree one exactly, and decreasing returns is degree less than one. I define synergy as the case where average cost is lower for a large firm producing the same level of output as several identical small firms. If average costs are equal then no synergy exists. If several small firms can produce at lower cost than a single large firm then negative synergies exist.

Firms in this paper attempt to obtain a prize. Increasing input makes the probability of success rise without increasing the value of the prize. In addition, there is only a single prize which must be shared in some way if more than one firm succeeds.

Suppose there is a single firm which hires a vector of inputs, x. Eq. (2.1) gives the production function, where y is the value of the prize if successful.

$$Y = \pi(x)\, y \tag{2.1}$$

This paper concentrates on the functional form of $\pi(x)$. Suppose the function is homogeneous of degree one in all elements of x; this will imply synergy. To see this suppose there initially exists a single firm hiring x^*. Now suppose employment of resources is increased by a factor of $\lambda > 1$. There are two ways employment could rise: the single firm could increase its employment by λ, or the number of firms could increase by λ. If the aggregate probability of success is higher in the first case than the latter, it means the single firm could have hired less output and had the same probability of success as a collection of smaller firms. Thus it can produce at lower cost which is my definition of synergy.

If the single firm expands employment by λ output rises by the amount in eq. (2.2).

$$Y_s = \pi(\lambda x^*)\, y = \lambda\, \pi(x^*) \tag{2.2}$$

If the single firm is duplicated λ times aggregate output is given by eq. (2.3).

$$Y_m = \{\, 1 - [\, 1 - \pi(x^*)]^{\lambda} \}\, y \tag{2.3}$$

Since $Y_s > Y_m$ synergies exist.

The intuition is fairly straightforward. Constant returns to scale in eq. (2.2) implies that each additional unit of inputs has the same productivity as previous ones. This is also the case in eq. (2.3). In eq. (2.3), however, there are cases of redundancy. That is, there are some cases where more than one firm may be successful. As the size of the industry is expanded and the overall probability of success rises, the marginal product of additional factors gets smaller, not because their probability of success is lower than the factors already employed, but because the probability that more than one firm will be successful and therefore redundant is increasing. This means that the function has decreasing returns to scale. Eq. (2.2) in effect assumes that a single firm can avoid redundancies while eq. (2.3) assumes that a collection of firms cannot.

This can be illustrated by Figure 1. Suppose there are two firms , A & B, each attempting to produce a particular good. Each firm consists of factor inputs in identical proportions and has some probability of success. The square represents all possible outcomes and has an area of one. The aggregate probability of success is the sum of areas I, II and III. If B were not attempting to produce the good, the aggregate probability of success would be I and II. Thus, the contribution of B to aggregate success is area III. Area I is a redundancy where both firms develop the good.

It is important to note that there is one case where homogeneity of degree one does not imply synergies. If one considers probability in continuous time (2.2) can be rewritten as:

$$\lambda\ \pi(x^*)dt\ y \tag{2.4}$$

And (2.3) can be rewritten as:

$$\{\ 1 - [1-\pi(x^*)dt]^{\lambda}\ \}\ y \tag{2.5}$$

Since dt is a very small number both (2.4) and (2.5) are close to zero.

Take the derivative of (2.4) with respect to λ and one gets:

$$\pi(x^*)dt\ y \tag{2.6}$$

Take the derivative of (2.5) with respect to λ and one gets:

$$-\ln[1-\pi(x^*)dt]\ \{\ [1-\pi(x^*)dt]^{\lambda}\ \}\ y \tag{2.7}$$

Since $\pi(x^*)dt$ is very small this simplifies to (2.6).

Thus, (2.4) and (2.5) reduce to the same formula.

Most of the literature on growth via innovations has used a continuous time specification for just this reason. Nonetheless, there are cases when one may wish to model in discrete time.[2]

In continuous time, the probability of success over a period of time, dt, is small, but the probability of more than one firm succeeding is second-order small. In terms of Figure 1, as dt gets small the probabilities π_A and π_B approach zero. The total area of success is well approximated by areas II and III which shrink to the line segments of the left and bottom borders, respectively. Area I shrinks to a

---

2 Two important reasons being that continuous time models imply instantaneous depreciation of past R&D efforts and perfectly elastic input demands. Continuous time also eliminates ties in the R&D race, while we do observe near simultaneous introduction of new and competing goods.

single point, the bottom left corner. This means that an additional unit of inputs gives the same additional product as any previously hired one.

Viewing production as a probability of success implies that only if there is some benefit from grouping factors together within firms will the production function be homogeneous of degree one. Production functions like eq. (2.4) which have no synergies also have decreasing returns to scale. If production is modelled as homogeneous of degree one, then monopolies will arise. If production is modelled as in (2.4) then there is still the unresolved issue of how the prize is divided when more than one firm is successful. The rules for this division will impact on aggregate employment and, thus, probabilities of success.

3. Industrial Organization

Consider a alternative specification of the production function that gives no advantage to large firms. If the function $\pi(x)$ is given by eq. (3.1) then large firms have no cost advantage over small ones, but there are decreasing returns to scale in the aggregate. Eq. (3.1) is a continuous version of eq. (2.3) where x varies, as opposed to λ.

$$\{ 1 - \exp[-f(x)] \} = \pi(x) \tag{3.1}$$

$f(x)$ is increasing in all elements of x and homogeneous of degree one.

I now consider the behavior of firms when they compete with each other for success. Consider the case of a firm in an industry where all other competing firms combined hire a vector of inputs x_0. Since $f(x)$ is homogeneous of degree one and all firms face the same factor prices they will hire factors in identical proportions and one could view x as a scalar showing how many bundles of fixed proportion inputs are hired. If the units of labor are appropriately normalized the aggregate probability of success for all other firms can be written as in (3.2).

$$\pi(x_0) = \{ 1 - \exp[-x_0] \} \tag{3.2}$$

If the firm is unsuccessful, it gets nothing. If the firm is successful and all other firms are not then the firm gets the full benefit, y. If, however, the firm and some subset of the other firms are simultaneously successful the benefit must be divided somehow. This division may depend on exactly how many other firms are successful and its expected value may depend on the firm's hiring decision. I denote this value v. Thus the firm's problem can be written as:

$$\max_{x} \{ 1-\exp[-x] \} \left[(1-\pi_0)\, y + \pi_0\, v(x) \right] - wx \tag{3.3}$$

The best way to model v will depend on the problem at hand. In the R&D example I am considering if two or more firms are simultaneously successful they will each have a blueprint in hand allowing them to produce an identical good. In the Bertrand equilibrium they behave competitively and profits will be zero; thus, v=0. If firms agree to behave as a cartel, they may divide the benefit up evenly or according to some other rule. With patent laws, the firm that is lucky enough to apply for a patent soonest will receive y and all others nothing. For expectational purposes this is the same <u>ex ante</u> as dividing up y.

Suppose v=0. The first-order condition is:

$$\exp[-x]\, (1-\pi_0)\, y = \exp[-(x+x_0)]\, y = w \tag{3.4}$$

Eq. (3.4) shows that the marginal product of an input bundle depends on the the aggregate employment of inputs and not merely the firm's employment. Since there is diminishing returns to scale marginal cost exceeds average cost and firms will earn profits as long as x is positive. It is also the case that a single large firm will enjoy more profit, on average, than a collection of small firms. This is because the competitive behavior which drives v to zero will be eliminated. If a single firm did exist and hired the aggregate optimum amount of labor there is no incentive for new firms to enter. Even though a large firm will earn positive profits, the marginal product of additional inputs is the same for an entering firm and an established firm.

There is a clear incentive in this situation for firms to collude. If they can agree to not price competitively once the designs are in hand, they can divide up y and earn positive profits. However, once firms agree to collude the incentives for R&D change.

Suppose, for example, that firms agree to divide y equally among successful firms. In this case v may depend on the number of other firms in the agreement and the amounts of inputs each hires. It will not depend on x, however, as the firm's own hiring decision is assumed not to affect either the number of other firms or the amount of inputs they hire. The first-order condition for this case is given by (3.4).

$$\exp[-x]\ \{(1-\pi_0)y + \pi_0 v\} = \exp[-(x+x_0)]\ y + \exp[-x](1-\exp[-x_0])\ v = w \tag{3.5}$$

If there were only one other firm v would be (1/2)y. Comparing (3.4) and (3.5) shows that there is an extra incentive here for firms to hire inputs. The marginal product of inputs includes the term, $\exp[-(x-x_0)]y$ from (3.4) and an additional one, $\exp[-x](1-\exp[-x_0])v$, which reflects the benefits that accrue as a firm is able in increase its share of profits in the cases when y is divided. This extra benefit comes at the expense of other firms and is also decreasing in x, which implies that small firms have a greater incentive to expand than large firms.

Suppose initially there is a single firm. Since it is the only firm engaging in R&D it views x_0 as zero and (3.4) is its profit maximizing condition. Now imagine a firm that is considering entry. It views x_0 as fixed and must decide if it should hire the first unit of labor. It evaluates eq. (3.4) at x=0. Since eq. (3.4) holds with equality and $\exp[-x](1-\exp[-x_0])v$ is positive, it must be that the middle term in eq. (3.5) is greater than w. Thus, the firm would enter. There are incentives to enter that exceed the incentives to expand. In this case the market will eventually be characterized by a large number of small firms and the aggregate level of employment will be higher than either the competitive or pure monopoly case.

These incentives to enter can be illustrated by a different example. Suppose there were a large school of fish in a lake. The number of fish is known, but not their location. Think of the inputs as individual fishing boats. A single fishing firm may send out many boats. The expected marginal product of an additional boat is not the number of fish it is expected to bring in, but rather, the expected number of fish it will catch that the firm's existing boats will not. If there is free entry a new firm will enter while an existing firm will not because the expected marginal product of an single entering boat is the number of fish it is expected to bring in. This will include some fish that would have been caught by other firms. The equilibrium will be an industry consisting of many single-boat firms and will result in overfishing.

The situation is akin to rent seeking. Some inputs are employed, not because they contribute to overall production (in this case an aggregate probability), but because they redirect revenue from one party to another.

Which market arrangement is the best in a social sense? An ex ante perception of shared profits will result in over-employment of inputs. The aggregate marginal product of R&D will be less than the marginal cost. However, this cost is offset by the fact that greater employment of inputs translates to higher rates of growth. A complete specification of preferences, especially rates of time preference, is necessary to determine which effect dominates.

4. Conclusions

I have shown that when inputs are productive in the sense of increasing probabilities of successfully capturing a reward, homogeneity of degree one is inconsistent with divisibility of production and implies synergies. I have also shown that if one adopts a production function which captures divisibility of production, that issues of market organization become important. In particular, the distribution scheme adopted for cases when more than one firm succeeds is of vital importance to determining both the number of firms and the aggregate amount of inputs that are used.

References

Aghion, Philippe, and Peter Howitt. (1989) "A Model of Growth through Creative Destruction." Department of Economics M.I.T., working paper no. 527.

Brock, William A. and Leonard J. Mirman. (1972) "Optimal Growth and Uncertainty: The Discounted Case." Journal of Economic Theory vol. 4 pp. 479-513.

Cheng, Leonard K., and Elias Dinopoulos. (1991a) "A Schumpeterian Model of Economic Growth and Fluctuations." University of Florida, manuscript.

__________. (1991b) "International Schumpeterian Business Cycles and Economic Growth." University of Florida, manuscript.

Dixit, Avinash. (1988) "A General Model of R&D Competition and Policy." RAND Journal of Economics vol. 19 no. 3 pp. 317-326.

Fölster, Stefan (1990) "Firms' Choice of R&D Intensity in the Presence of Aggregate Increasing Returns to Scale." Journal of Economic Behavior and Organization vol. 13 no. 3 pp. 387-403.

Grossman, Gene M., and Elhanan Helpman. (1991a) "Quality Ladders in the Theory of Growth." Review of Economic Studies vol. 63 pp. 43-61.

__________. (1991b) "Quality Ladders and Product Cycles." Quarterly Journal of Economics vol. 106 no. 2 pp. 557-586.

__________. (1990) "Trade, Innovation, and Growth." American Economic Review vol. 80 no. 2 pp. 86-91.

Phillips, Kerk L. (1990) "A Model of Development using Quality Ladders and Product Cycles." University of Michigan, manuscript.

Rivera-Batiz, Luis A., and Paul M. Romer. "Economic Integration and Endogenous Growth." Quarterly Journal of Economics vol. 106 no. 2 pp. 531-556.

__________. (1991b) "International Trade with Endogenous Technical Change." European Economic Review forthcoming.

Romer, Paul M. (1990) "Endogenous Technological Change." Journal of Political Economy vol. 98 no. 5 pp. S71-S102.

Segerstrom, Paul S. (1991) "Innovation, Imitation and Economic Growth." Journal of Political Economy vol. 99 no. 4, forthcoming.

Segerstrom, Paul S., T. C. A. Anant, and Elias Dinopoulos. (1987) "A Schumpeterian Model of the Product Life Cycle." American Economic Review .

Figure 1