European Economic Review 6 (1975) 17–38. © North-Holland Publishing Company

ON THE RELATIONS BETWEEN DEMAND CREATION AND GROWTH IN A MONOPOLISTIC FIRM*

Eithan HOCHMAN** and Oded HOCHMAN

Department of Economics, Tel-Aviv University, Israel

Received August 1973, revised version received April 1974

This paper deals with the allocation of resources over time by a monopolistic firm between growth of the productive capacity and growth of the market-demand capacity. As the demand-creation relations follow an S-shaped curve, different phases in the behavior of the growing firm are conceived in which investment cycles occur both in productive and demand-creation activities.

The paper analyzes the case of homogeneous resources as well as nonhomogeneous resources. It is shown that in general the phases in the behavior of the growing firm are preserved in both cases. In the homogeneous case, when there are investment activities in both types of capital, it is shown that the firm will allocate its resources between the two activities in such a way that the ratio of the rate of growth of demand price (with respect to demand-creation capital) and the rate of growth of output (with respect to productive capital) will be equal to one plus the reciprocal of the elasticity of demand and will, therefore, be bounded between zero and one.

1. Introduction

A monopolistic firm makes decisions over time about the allocation of its resources between investments in the production process and investments in the selling process. Within a static framework, there is extensive literature on the subject; references can be found in Hahn (1959), Dorfman and Steiner (1954), Hieser and Soper (1966), and Ball (1968). Nerlove and Arrow (1962) formulated and analyzed a dynamic model for a monopolistic firm facing a demand law influenced by advertising. In their model they assume that there is a stock of goodwill measured in units having a price of $1.00 so that a dollar of advertising expenditure increases the stock of goodwill by a like amount. Even though they initially formulated the problem as a functional one in advertising and price, they then reduced it to a functional one in advertising alone. Dhrymes (1962) extended the same model to include investment in productive capital as well.

*Giannini Foundation Paper No. 369. We should like to acknowledge J. Frankel and Y. Weiss for their helpful comments on an earlier draft; we are greatly indebted to A. Jacquemin for his constructive suggestions and critical comments.

**Eithan Hochman is currently visiting in the Department of Agricultural Economics and the Giannini Foundation, University of California, Berkeley.

Thompson and Proctor (1969) analyzed the behavior of a monopolistic firm encompassing investments, output prices, informative advertising, and brand advertising; their model is basically linear in its structure with a linear demand function and a fixed coefficient production function.

A number of economists [Gould (1968), Treadway (1969), and Lucas (1967a, b), for example] recently contributed analyses using the 'cost of adjustment' argument to obtain an investment demand function for the competitive firm. Gould (1970) applied this approach to optimal advertising policy but retained the assumption of competitive conditions in the product market; he did not take into consideration investment in productive capital.

Jacquemin and Thisse (1972) show that it is not necessary to assume a non-linear cost function in order to have the price of goodwill change over time. They established that the Nerlove–Arrow theorem (a ratio of a stock and a flow) is not the direct dynamic counterpart of the Dorfman–Steiner theorem (a ratio of two flows) and is a special case of a more general expression where the price of goodwill is not assumed to be equal to unity.

In our present model we use an approach similar to the one adopted by Hochman et al. (1973) in analyzing the demand for investment in productive and financial capital and apply it to the relations between demand creation and the growth of a monopolistic firm.

As the demand-creation relations follow an S-shaped curve, different phases in the behavior of the growing firm are conceived.

In the early stages of growth, all resources are invested in the expansion of the firm's production capacity; there is no activity in demand creation. This phase is followed by a second one in which all investments are channeled to demand-creation capital. In this phase the firm takes advantage of the increasing marginal returns to demand-creation capital by diverting into demand creation some of the existing productive resources acquired during the first phase. In the last phase the firm chooses to invest in both types of capital. The steady state is reached in the last phase in regions of decreasing marginal returns to both types of capital.

Regarding the optimal dynamic path, it is shown that operation in a region where the schedule of demand creation follows an S-shaped curve will result in an *investment cycle* in productive capital: Positive investment in the first interval is followed by disinvestment in the second interval; then there is a renewal of investment in productive capital in the last interval. The cycle in demand-creation capital, on the other hand, is characterized by zero investments in the first interval followed by positive investment at an increasing rate through the following intervals although, during the last interval, the rate of investment starts to decrease. Investment in demand creation after it starts is always continuous, contrary to investment in productive capital.

When there are investment or disinvestment activities in both types of capital (phases II and III), it is shown that the Dorfman–Steiner theorem is replaced by the following: A firm which can influence the demand for its product through direct allocation to demand-creation capital will allocate its resources between this type of capital and productive capital in such a way that the ratio of the rate of growth of demand price (with respect to demand-creation capital) to the rate of growth of output (with respect to productive capital) will be equal to one plus the reciprocal of the elasticity of demand and will, therefore, be bounded between zero and one.

In the last section the assumption of homogeneous resources is relaxed, and different rates of depreciation as well as different costs of adjustments are assumed in the two types of stocks. It is shown that in general the three phases in the behavior of the growing firm are preserved.

2. The model

Let K denote the stock of resources utilized in producing the quantity sold q. The production function $q = q(K)$ is twice continuously differentiable where $q_K > 0$ and $q_{KK} < 0$. The assumption that production is a function of only one resource, which may be interpreted as a production function with fixed proportion between capital and labor, is adopted here since it simplifies the exposition considerably and allows us to concentrate on the main problem of allocation between production and demand creation.

The firm may divert part of its resources (human and non-human), such as skillful labor, research personnel, and equipment and buildings, to departments that either involve themselves directly with the promotion of sales [Hieser and Soper (1966)] or are involved in research and development (R&D) of the product, i.e., changes in the quality of the product holding the output constant. Let A denote this type of capital which is devoted directly to demand creation, and let the demand relations be defined by $p = p(q,A)$. The price function is twice continuously differentiable with $p_q < 0$, $p_{qq} < 0$, and $p_A > 0$. The second partial differential p_{AA} behaves as described in fig. 1; it is first positive and then changes to negative at inflection point ($\bar{A}$).

The assumption of S-shaped relations of demand creation has both theoretical and empirical bases [see the discussions in Rao (1970) and Hieser and Soper (1966)]. The state of the firm is described by the two variables K and A whose rates of change over time are given by:

$$\dot{K} = I - \sigma K, \tag{1}$$

$$\dot{A} = a - \sigma A, \tag{2}$$

where I denotes gross investment in productive capital and a denotes current outlays in advertising, R&D, and any other expenditures that directly influence the price of the product at a given output. We assume equal rates of depreciation of both stocks. This assumption may be justified by considering the total stock of resources available for the firm as pooled together under the heading of 'capital' while, on the optimal path, decisions are made as to what portion will be diverted into production and what portion will be diverted into demand creation. This may describe many real world situations, e.g., a firm deciding about the portion of its products used as samples for sales promotion, consulting

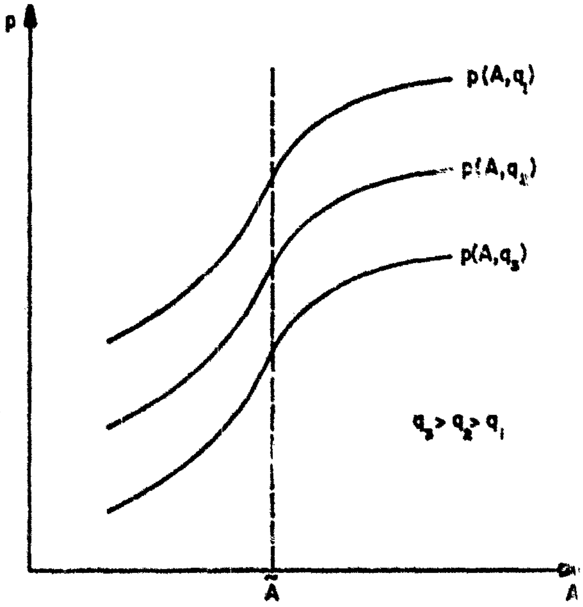


Fig. 1

firms allocating their human resources between the consultation services and drafting new clients, etc. The assumption of equal rates of depreciation simplifies the calculations considerably; however, in section 5, when this assumption is relaxed, it is shown that the phases in the behavior of the firm remain essentially the same.

The cash flow during each period of the firm is thus[1]

$$R = pq - w(c), \tag{3}$$

where $c = a+I$, the total gross investment at period t. The 'adjustment cost' function $w(c)$ includes the price of capital as well as the cost of adjustments and

[1]The independent variable t will be omitted whenever possible.

is defined by $w(c) \gtreqless 0$ for $c \gtreqless 0$ where $w_c > 0$ and $w_{cc} > 0$ for all values of c.[2,3] Thus, the maximization problem of such a firm can be stated as follows:

$$\max \int_0^\infty R\,e^{-rt}\,dt = \int_0^\infty [pq - w(c)]\,e^{-rt}\,dt \tag{4}$$

subject to

$$\dot{K} = I - \sigma K,$$

$$\dot{A} = a - \sigma A,$$

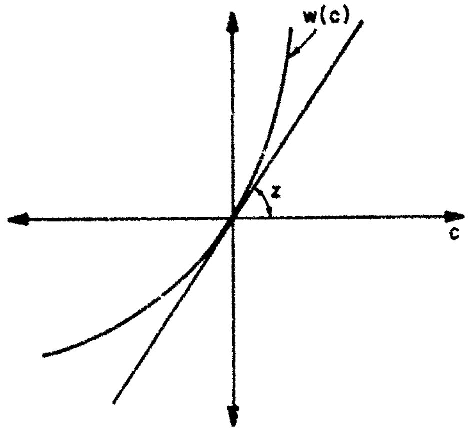
$$K + I \geqq 0,\ K(0) = K_0,$$

and

$$A + a \geqq 0,\ A(0) = A_0.$$

This is a problem of calculus of variations where the state variables are K and A and the controls are I and a.[4]

[2]This can be made more explicit by assuming the cost component in the function to be equal to $z \cdot c$ where z is the price of capital in the market. If the capital market is competitive and z has a fixed value, the function $w(c)$ has the following shape:



At the point $c = 0$, $w_c = z$, the adjustment costs divert the function $w(c)$ from its tangent at $c = 0$ as c increases or decreases. If the capital market is imperfect, the deviation from the tangent is increased.

[3]The same model may describe alternatively a firm which allocates its skilled labor between production and demand creation, other resources being fixed. Note that, if this approach is adopted, even though labor is hired, it is considered as a stock of human capital. This may be the case in a firm which supplies services only, and its employees are not fired as a matter of policy – e.g., the IBM Corporation. The prospective employee needs special training which is taken into account in the adjustment costs, and his 'price' is measured by the discounted value of his future salaries. In case of budget cutting, the firm gains the discounted value of all future salaries which the fired employees would have received after deduction of costs of adjustment caused by compensation payments and other frictional costs. We assume that the adjustment costs of recruiting new employees are the dominant factor so that the adjustment costs of reallocating them between the different departments may be ignored.

[4]See Arrow and Kurz (1970) and Pontryagin et al. (1964).

3. Description of resource allocation between demand creation and growth of the firm

Applying the Maximum Principle [Arrow–Kurz (1970) and Pontryagin et al. (1964)] by using the current value Lagrangian

$$L(A, K, \lambda_1, \lambda_2, a, I, \mu_1, \mu_2)\, e^{rt} = pq - w(c) + \lambda_1(a - \sigma A) + \lambda_2(I - \sigma K) + \mu_1(A + a) + \mu_2(K + I), \tag{5}$$

we obtain the necessary conditions for maximization as follows:

$$w_c \leq \lambda_1, \qquad (w_c - \lambda_1)(A + a) = 0, \tag{6a}$$

$$w_c \leq \lambda_2, \qquad (w_c - \lambda_2)(K + I) = 0, \tag{6b}$$

$$\dot{\lambda}_1 \leq \lambda_1(r + \sigma) - p_A q, \qquad [\dot{\lambda}_1 - \lambda_1(r + \sigma) + p_A q](A + a) = 0, \tag{6c}$$

$$\dot{\lambda}_2 \leq \lambda_2(r + \sigma) - q_K MRq, \qquad [\dot{\lambda}_2 - \lambda_2(r + \sigma) + q_K MRq](K + I) = 0, \tag{6d}$$

where $MRq = \mathrm{d}/\mathrm{d}q\,(pq)$.

Note that, if there is any production and sale activity by the firm, equality always holds in (6b) and (6d); the only alternative is exit from the industry.[5] Assuming the conventional negative-sloped marginal revenue curve (MRq), there is a level of output, say q_0, such that

$$MRq[q, p(q, 0)] \geq 0, \quad \text{for } q \leq q_0,$$

and

$$MRq[q, p(q, 0)] \leq 0, \quad \text{for } q \geq q_0.$$

There also exists a value of productive capital $\hat{K}$ such that

$$q_K(\hat{K}) MRq[q(\hat{K}), p(q(\hat{K}), 0)] = p_A[q(\hat{K}), 0] \cdot q(\hat{K}),$$

where $q(\hat{K}) < q_0$.

If the initial state is such that $K_0 < \hat{K}$ and $A_0 = 0$, the following system of equations holds:

$$A = 0, \quad a = 0, \tag{7a}$$

$$\lambda = w_c(I), \tag{7b}$$

$$\dot{\lambda} = \lambda(r + \sigma) - q_K(K) MRq[q(K), 0]. \tag{7c}$$

Conditions (7b) and (7c) have the usual interpretation: (7b) states that the shadow price $\lambda(t)$ must be equated to the marginal cost of investment in productive capital at time t; and (7c) – in integral form – states that $\lambda(t)$ is the discounted value at time t of later values of marginal products of productive capital which, in turn, equals – by (7b) – the immediate marginal cost of adjustment [Treadway (1969)].

[5] See discussion of the behavior in phase I in section 4 and also in Treadway (1969).

At $K = \hat{K}$, (6a) and (6c) become equalities and the firm starts to invest in demand creation as well. The following system of equations will replace (7),

$$A+a>0, \quad K+I>0, \tag{8a}$$

$$\lambda = w_c(c), \tag{8b}$$

$$\dot{\lambda} = \lambda(r+\sigma)-q_K(K)MRq[q(K), A] = \lambda(r+\sigma)-p_A[q(K), A]q(K), \tag{8c}$$

$$q_K(K)MRq[q(K), A] = p_A[q(K), A]q(K). \tag{8d}$$

Condition (8c) – in integral form – states that $\lambda(t)$ is, at the same time, the discounted value of later values of marginal products of demand creation capital. Condition (8d) describes the well-known equality of the values of the marginal products of the two types of capital. If we denote $\eta_{yx} = (\partial y/\partial x)(x/y)$, then (8d) can be rewritten

$$\frac{1}{K}\eta_{qK}(1+\eta_{pq}) = \frac{1}{A}\eta_{pA}. \tag{9}$$

From (9), we can verify that the ratio between the rate of growth in demand price resulting from investment in A and the rate of growth in output resulting from investment in $K = 1+\eta_{pq}$. From the fact that $\eta_{pq}<0$ and the rational behavior of the monopolist in choosing such outputs that $MRq \geqq 0$, we have on the optimal path:

$$0 \leqq \frac{\text{Rate of growth in demand price with respect to } A}{\text{Rate of growth in output with respect to } K} \leqq 1.$$

We now assume weak *separability* in the demand relations which imply:[6]

$$\frac{\partial}{\partial q}(\eta_{pA}) = 0, \tag{10a}$$

$$\frac{\partial}{\partial A}(\eta_{pq}) = 0. \tag{10b}$$

Thus, under (10), the left-hand side of (9) is a function $f(K)$ of K alone; and the right-hand side is a function $g(A)$ of A alone

In fig. 2 we draw $f(K)$ as a function of K under the assumption of diminishing marginal products of productive capital; and $g(A)$, as a function of A under the assumption that $p(A, q)$ for any given q, behaves as described in fig. 1. The relations between K and A on the optimal path can be derived directly from fig. 2 and are described in the (K, A) plane by the segmented curve (Q-curve) in fig. 3. Let

$$Q(K, A) = \frac{1}{K}\eta_{qK}(1+\eta_{pq})-\frac{1}{A}\eta_{pA}.$$

[6]The meaning of the assumptions of weak separability is that, in the plane (p, q), the tangents to the demand curves for different A's but the same q intersect at the same point; it is the same in the (p, A) plane for different q's but the same A.

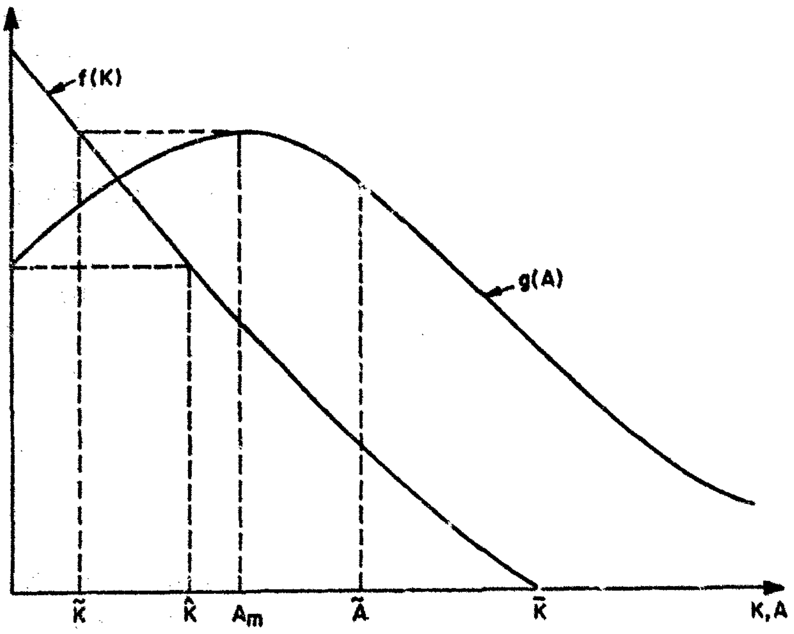

f(K)
g(A)
K̃
K̂
Am
Ã
K̄
K,A

Fig. 2

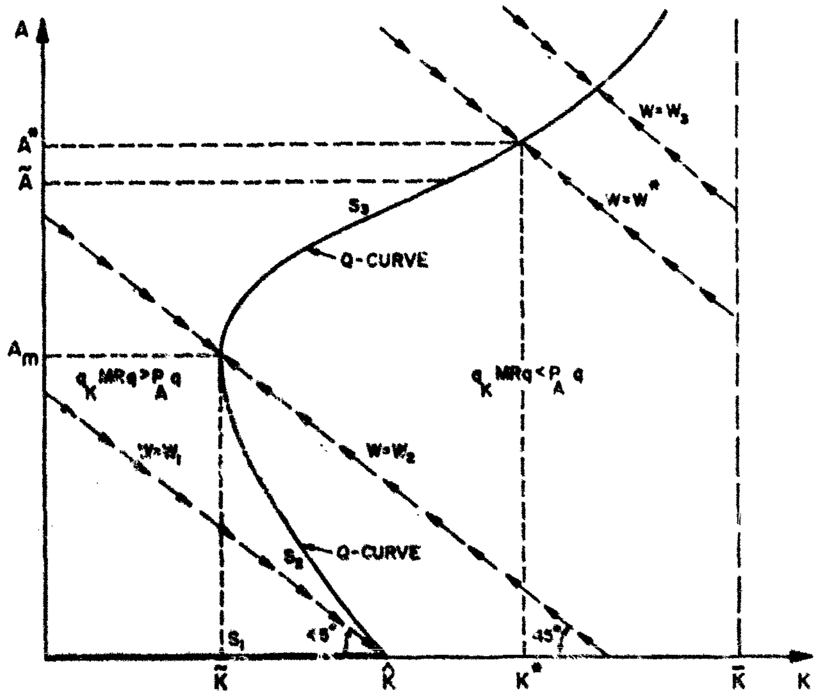

A
A*
Ã
Am
S3
Q-CURVE
W=W3
W=W*
qK MRq > PA q
qK MRq < PA q
W=W1
W=W2
S2
Q-CURVE
S1
45°
45°
K̃
K̂
K*
K̄
K

Fig. 3

Then the segmented Q-curve can be divided into three segments:

$$S_1 = \{(K, A): Q(K, A) > 0, 0 \leqq K \leqq \hat{K}, A = 0\},$$

$$S_2 = \{(K, A): Q(K, A) = 0, \hat{K} \leqq K \leqq \hat{K}, 0 \leqq A \leqq A_m\},$$

$$S_3 = \{(K, A): Q(K, A) = 0, \hat{K} \leqq K \leqq \hat{K}, A_m \leqq A \leqq \infty\}.$$

This locus divides the plane into two regions. If the initial resources of the firm ($W_0 = K_0 + A_0$) are such that the firm starts in the region to the left and above the Q-curve where $q_K MRq > p_A q$, the firm will move instantaneously to the right on a 45° budget line until it reaches one of the three segments of the Q-curve. This instantaneous movement is the result of the assumption we made that transfer of human and non-human capital *within the firm* does not involve costs of adjustment.[7] If initial resources are such that the firm starts in the region to the right and below, there will be an instantaneous movement on a 45° budget line in the opposite direction until one of the last two segments of the Q-curve is reached.

Note that it is only along segments S_2 and S_3 that eq. (8) holds and functional relations exist between A and K – the relations of one-to-one correspondence break down on segment S_1. On segment S_1 (which coincides with the abscissa), eq. (7) replaces eq. (8). The slope of the Q-curve along segments S_2 and S_3 is derived from the total differentiation of eq. (9), under the assumptions of weak separability in the demand, yielding

$$\frac{\mathrm{d}A}{\mathrm{d}K} = \frac{f_K(K)}{g_A(A)}, \tag{11}$$

where

$$f_K(K) = \left[-K^{-2}\eta_{qK} + K^{-1}\frac{\partial}{\partial K}(\eta_{qK})\right](1+\eta_{pq}) + K^{-1}\eta_{qK}q_K\frac{\partial}{\partial q}(\eta_{pq})$$

and

$$g_A(A) = -A^{-2}\eta_{pA} + A^{-1}\frac{\partial}{\partial A}(\eta_{pA}) = \frac{1}{p}\left(p_{AA} - \frac{p_A^2}{p}\right)$$

are correspondingly the slopes of the curves $f(K)$ and $g(A)$ in fig. 2.

Evaluating the sign of $f_K(K)$, we assume the following: (1) $\partial\eta_{pq}/\partial q < 0$ resulting from the assumption of a negatively sloped marginal revenue for all A, and (2) $q_K > 0$ and $q_{KK} < 0$ resulting from the assumptions on the sign of the first two derivatives of $q(K)$. These assumptions and the fact that $1 > 1 + \eta_{pq} > 0$ for $K < \bar{K}$ imply $f_K(K) < 0$ for all values of $K < \bar{K}$ where $\bar{K}$ satisfies $(1 + \eta_{pq}) = 0$. The

[7]Here, too, the assumption that rechanneling resources between the two types of capital does not involve costs of adjustment implicitly assumes the existence of a pooled stock of capital. Thus, we neglect costs of transferring existing resources from productive use to demand-creation use (the only case where such a transfer occurs in our model). Only the costs of acquiring capital goods outside the firm are taken into account here.

sign of $\mathrm{d}A/\mathrm{d}K$ will, therefore, be the opposite of $g_A(A)$. Note that, by the S-shaped curve in fig. 1, we assumed $p_{AA}>0$ for $A<\tilde{A}$.

At $A_m<A$ in figs. 2 and 3, the following equality holds: $p_{AA}/p_A = p_A/p$, i.e., the elasticities of p and p_A, both with respect to A, are equal; and we conclude that $g_A(A)\gtreqless 0 \Leftrightarrow A\lesseqgtr A_m$. Thus, on segment S_2 the slope of the Q-curve is negative and increases in its absolute value until it reaches infinity at A_m. Segment S_3 starts from A_m, with an infinitely positive slope, decreasing at first and then increasing. A increases to infinity while K approaches $\bar{K}$. Without loss of generality, we assume that $\mathrm{d}A/\mathrm{d}K<-1$ at $(\hat{K}, 0)$. Otherwise, there will be a subsegment where $0>\mathrm{d}A/\mathrm{d}K>-1$ which will represent a local minimum; the firm will not stay on this subsegment but will move instantaneously to the left along the 45° budget line until it reaches the 'right' part on segment S_2.

The optimal behavior of the firm is described by the movement along the Q-curve from any initial state (given by its intersection with a 45° budget line) toward a steady state which we will assume lies in segment 3.[8] The steady state may occur only in segments 1 and 3. If it occurs in segment 1, a steady state without demand-creating capital exists. The case in which the steady state is in segment 3 is far more interesting and, therefore, was chosen to be represented here. The necessary conditions for a steady-state solution are derived from the transversality conditions [eq. (12)], and the existence of the steady state ensures a unique solution. A steady-state solution exists if there is a solution in the positive plane (K, A) to the following set of equations:

$$w_c[\sigma(A+K)](r+\sigma) = q_K(K)MRq(K, A),$$

$$\{p_A q - w_c[\sigma(A+K)](r+\sigma\}\, A = 0.$$

Under our assumptions, such a solution exists.

If the firm starts from segment S_1, K increases up to $\hat{K}$, while $Q(K, 0)>0$. Along the segment S_2, K decreases; and A increases until the point A_m is reached. At this point, both K and A increase toward the steady state (K^*, A^*). Along S_2 and S_3, $Q(K, A) = 0$ holds; note that, though K decreases along S_2, the total resources of the firm are increased. This is demonstrated by the movement to higher equal wealth lines represented by the 45° budget lines ($W^*>W_2>W_1$). On the other hand, if the firm starts at initial wealth $W_3>W^*$, K and A decrease monotonically; and the firm will move along S_3 toward the steady state value (K^*, A^*).

The following transversality conditions are additional necessary conditions [Chetty (1972, sect. 4)]:

$$\lim_{t\to\infty} \lambda_t \geqq 0, \qquad \lim_{t\to\infty} K_t\lambda_t\,\mathrm{e}^{-rt} = \lim_{t\to\infty} A_t\lambda_t\,\mathrm{e}^{-rt} = 0. \tag{12}$$

Eqs. (6) and (12) constitute a set of necessary conditions for the firm's problem. (For discussion as to whether the necessary conditions are also sufficient, see appendix A.)

[8] The steady state will be analyzed later when phase diagrams are introduced.

4. The dynamic behavior of the firm

The functional correspondence between K and A makes it possible to construct alternatively phase diagrams in either the (K, λ) plane or the (A, λ) plane representing the patterns of optimal productive investment and optimal demand-creation investment, respectively. To construct phase diagrams, we use the following set of equations derived from conditions (8):[9]

$$\left.\frac{\mathrm{d}\lambda}{\mathrm{d}K}\right|_{\dot{K}=0} = \sigma w_{cc}\left(1+\frac{\mathrm{d}A}{\mathrm{d}K}\right), \tag{13a}$$

$$\left.\frac{\mathrm{d}\lambda}{\mathrm{d}A}\right|_{\dot{A}=0} = \sigma w_{cc}\left(1+\frac{\mathrm{d}K}{\mathrm{d}A}\right), \tag{13b}$$

$$\left.\frac{\mathrm{d}\lambda}{\mathrm{d}K}\right|_{\dot{\lambda}=0} = (r+\sigma)^{-1}\left[q_{KK}MRq+q_K^2(p_{qq}q+2p_q) + \frac{\mathrm{d}A}{\mathrm{d}K}\,p_A(1+\eta_{pq})\right], \tag{13c}$$

$$\left.\frac{\mathrm{d}\lambda}{\mathrm{d}A}\right|_{\dot{\lambda}=0} = (r+\sigma)^{-1}\left[q(K)p_{AA}+p_Aq_K\frac{\mathrm{d}K}{\mathrm{d}A}\right], \tag{13d}$$

$$\left.\frac{\partial\dot{K}}{\partial\lambda}\right|_{K=\text{const.}} = \left.\frac{\partial I}{\partial\lambda}\right|_{K=\text{const.}} = 1\Big/w_{cc}\left(1+\frac{\mathrm{d}A}{\mathrm{d}K}\right), \tag{13e}$$

$$\left.\frac{\partial\dot{A}}{\partial\lambda}\right|_{A=\text{const.}} = \left.\frac{\partial a}{\partial\lambda}\right|_{A=\text{const.}} = 1\Big/w_{cc}\left(1+\frac{\mathrm{d}K}{\mathrm{d}A}\right), \tag{13f}$$

$$\left.\frac{\partial\dot{\lambda}}{\partial\lambda}\right|_{\substack{K=\text{const.}\\A=\text{const.}}} = r+\sigma, \tag{13g}$$

where $\mathrm{d}A/\mathrm{d}K = f_K(K)/g_A(A)$.

The slope of the curve $\dot{K} = 0$ in the (K, λ) plane (fig. 4) and the slope of the curve $\dot{A} = 0$ in the (A, λ) plane (fig. 5) are determined by conditions (13a) and (13b), respectively. The slopes of the curves $\dot{\lambda} = 0$ (figs. 4 and 5) are determined by conditions (13c) and (13d), respectively. Since there is an overlapping in phases in fig. 4, fig. 6 is used for the exposition of the horizontal and vertical arrows in the (K, λ) plane. The direction of the horizontal arrows in figs. 5 and 6 can be verified from conditions (13e) and (13f), correspondingly, and the direction of the vertical arrows from condition (13g).

In the analysis that follows, we distinguish between three phases which correspond on the Q-curve to the three segments.

[9]This geometric method is generally used for problems characterized by only one state variable. In our problem the functional correspondence between K and A [eq. (11)] allows us to consider K and A in two separate phase diagrams.

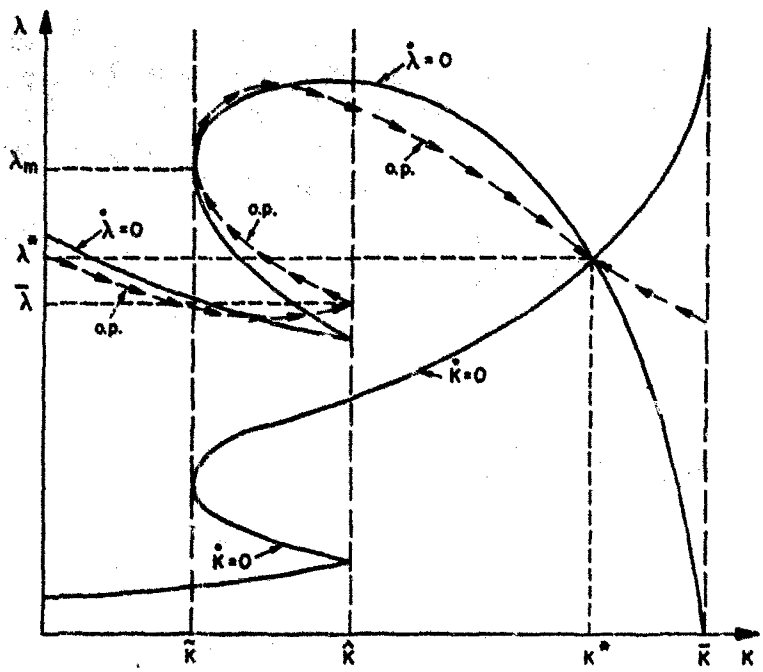

λ
λ̇=0
o.p.
λ_m
λ^*
$\bar{\lambda}$
K̇=0
$\tilde{K}$
$\hat{K}$
K^*
$\bar{K}$
K

Fig. 4

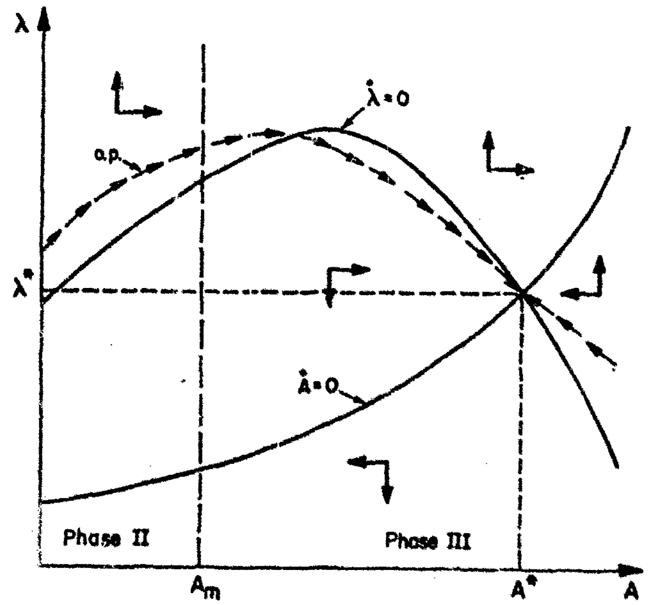

λ
λ̇=0
o.p.
λ^*
Ȧ=0
Phase II
Phase III
A_m
A^*
A

Fig. 5

Phase I: If the initial amount of resources is such that the firm is on segment 1, the firm starts at phase I where all investments are implemented into productive capital. This will characterize the optimal demand for investment as long as

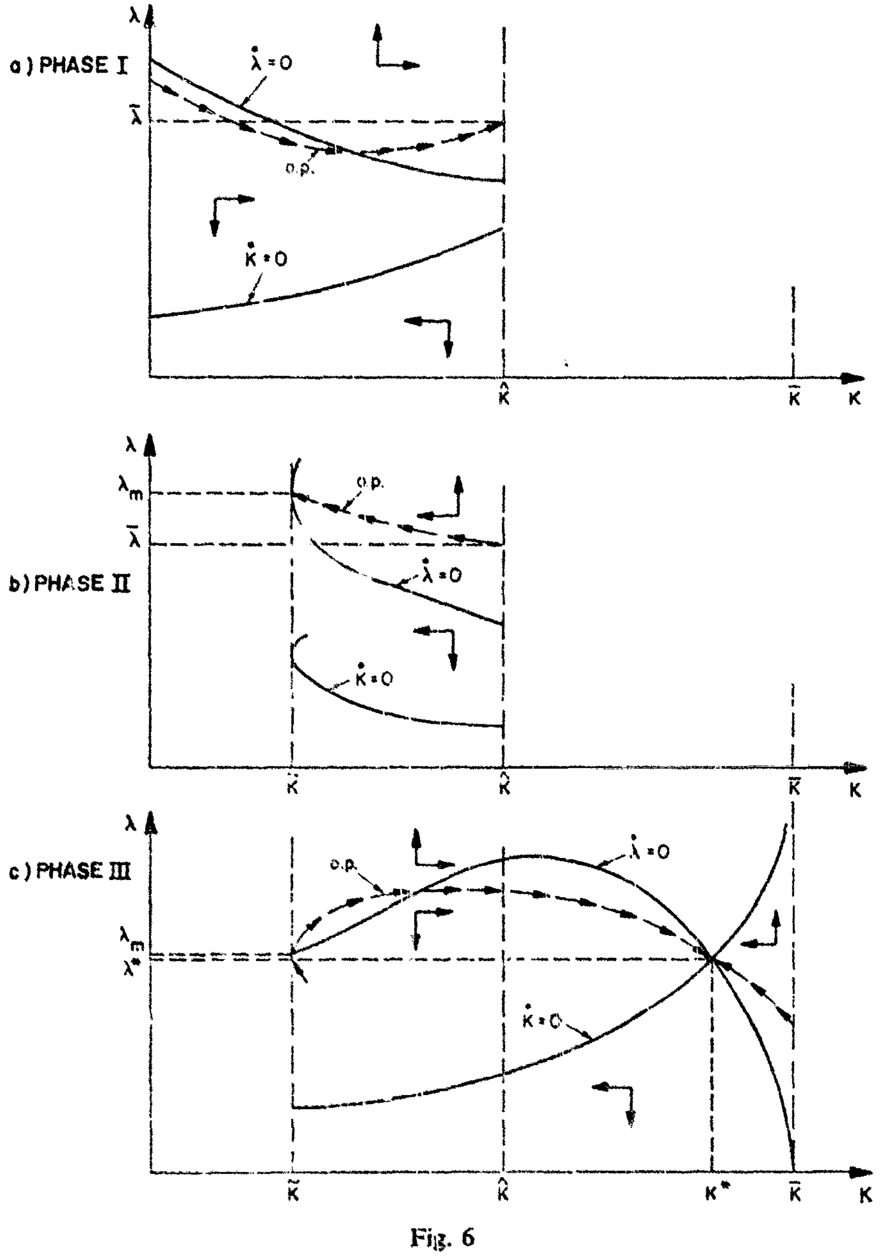


Fig. 6

$Q(K, 0) > 0$. From fig. 2 and fig. 6a, it is clear that the firm will expand first at a decreasing rate and then at an increasing one. However, the rate of investment in productive capital is accelerated in comparison with the case where demand creation is impossible, although at this phase no investment in demand creation has as yet been made. Note that *at this phase* some of the Treadway (1969) inferences about optimal demand for investment in productive capital hold even though we deal with a monopolistic firm, especially if we are willing to assume without loss of generality that $f(K)$ has a rising part at low values of K before obtaining the negative slope and thus allows for different production structures.[10] At the level of $\bar{K}$, the firm moves into phase II.

Phase II: At this phase (which coincides with segment S_2), the firm uses all of its new resources and parts of the existing resources (accumulated in the form of productive capital at phase I) to build its demand-creation capital. In doing so, the firm takes advantage of the increasing marginal returns to demand creation, $p_{AA}>0$, described by the lower part of fig. 1. Along the optimal trajectory, conditions (8) hold; and the values of the marginal products of both types of capital are equal. The dynamic behavior of the firm is described by the phase diagrams. The point $(\bar{K}, \bar{\lambda})$, where phase I ends and phase II begins, is a discontinuous point of the controls, i.e., I becomes negative from positive and a becomes positive from zero. It is not a differential point of $\lambda(t)$, $K(t)$, and $A(t)$. A cycle in $K(t)$ begins at this stage where K decreases instead of increasing, and it goes on decreasing until the end of phase II is reached at point $(\bar{K}, \lambda m)$. At this point, I acquires a zero value. The direction of the optimal trajectory in the (K, λ) plane is explained by the horizontal and vertical arrows in fig. 6b, and the direction of optimal investment in A is explained by the optimal path within phase II in fig. 5.

Phase III: In this phase both K and A increase toward their steady-state values (K^*, A^*). At early stages, both rates of investment are increasing though both $g_A(A)$ and $f_K(K)$ are negative; the monopolist firm still has the advantages of $p_{AA}>0$ for $A<\bar{A}$ and the relatively high elasticities of demand $(1/\eta_{pq})$. At later stages, as p_{AA} changes to negative and the elasticities of demand continue to diminish, K and A increase at a decreasing rate until a steady state is reached.

If the initial amount of resources is such that the firm starts on segment 3, say at $W_3 > W^*$, both K and A decrease until steady state (K^*, A^*) is reached. These processes can be verified from figs. 4, 5, and 6c.

5. Extension of the model

In the following section, while holding to our previous assumptions about the forms of production function and the price–demand function as well as the separability assumption, we relax the assumptions of the homogeneity of the resources allocated between production and demand. It is of great interest to consider the results of assuming (1) unequal rates of depreciation for the stock of

[10] Thus, for example, under increasing returns to scale in production, conditions may arise [see the discussion in Treadway (1969, pp. 236/7)] that the firm should leave the industry.

productive capital and the stock of demand-creation capital and (2) separate adjustment costs for the two stocks of capital.

Thus, the rates of change of the variables K and A over time are given instead by

$$\dot{K} = I-\sigma_2 K, \tag{1a}$$

$$\dot{A} = a-\sigma_1 A, \tag{2a}$$

where σ_1 and σ_2 are the corresponding rates of depreciation of A and K.

The two adjustment cost functions are $f(a)$ and $g(I)$ where $f(0) = g(0) = 0$, $f_a(0) =$ price of $A = s_1 =$ constant, and $g_I(0) =$ price of $K = s_2 =$ constant. An additional assumption that $f_{aa} = f_0$ and $g_{II} = g_0$ is applied mainly to simplify the computations. (This assumption is not an essential one and merely assumes that a quadratic approximation of the functions is enough.)

Under the above assumptions, $f(a)$ and $g(I)$ will be correspondingly

$$f(a) = s_1 a+\frac{f_0}{2}a^2, \tag{14}$$

$$g(I) = s_2 I+\frac{g_0}{2}I^2. \tag{15}$$

Now, applying the Maximum Principle by using the current value Lagrangian,

$$L(A, K, \lambda_1, \lambda_2, a, I, \beta)\,e^{rt} = pq-f(a)-g(I)+\lambda_1(a-\sigma_1 A) + \lambda_2(I-\sigma_2 K)+\beta(A+a), \tag{16}$$

(assuming that $K+I>0$; section 3), the following set of equations is obtained:

$$\lambda_1 \leq f_a, \tag{17a}$$

$$\lambda_2 = g_I, \tag{17b}$$

$$\dot{\lambda}_1 \leq \lambda_1(r+\sigma_1)-p_A q, \tag{17c}$$

$$\dot{\lambda}_2 = \lambda_2(r+\sigma_2)-q_K MRq. \tag{17d}$$

By differentiating (17a) and (17b) with respect to time and introducing the results in (17c) and (17d), correspondingly, the following equations are obtained (assuming that the equalities hold):

$$f_0\dot{a} = f_a(a)(r+\sigma_1)-p_A q, \tag{18}$$

$$g_0\dot{I} = g_I(I)(r+\sigma_2)-q_K MRq. \tag{19}$$

Eqs. (18) and (19), together with (1a) and (2a), furnish the basic equations for describing the optimal trajectories in the (K, A) plane.[11] But in trying to solve the

[11] For the model presented in section 5, additional work on sufficiency proof is in order. Since this would entail a substantial increase in the amount of mathematical analysis in what is already an overly long paper, we have decided to leave this consideration for a later work.

problem graphically, we encounter the difficulty of reducing the four-dimensional plane of a two-state variable problem $(K, A, \dot{I}, \dot{A})$ into a two-dimensional plane (K, A). This is done by considering different cross-sections of this plane; i.e., in the (K, A) plane we identify the singular curves $\dot{K} = 0$ and $\dot{A} = 0$ given $\dot{I}$ and $\dot{a}$. Nevertheless, when the optimal path is described, note that the movement is not necessarily in a plane where $\dot{I}$ and $\dot{a}$ are constant but they may also change. Thus, the singular curves $\dot{K} = 0$ and $\dot{A} = 0$ shift in the plane (K, A) and, for each point on the optimal path in the plane (K, A), there correspond two different curves $(\dot{K} = 0, \dot{A} = 0)$ representing the given $\dot{a}$ and $\dot{I}$ of this point. Note that the singular curves move continuously without jumps. Jumps are possible if and only if $\dot{I}$ and $\dot{a}$ jump, and these points of discontinuity should be located in advance. In general, locating the singular curves is not enough to characterize the optimal behavior of the system, and specific points or regions should be selected in order to characterize the optimal behavior. For example, if $\{K^*, A^*\}$ are the steady-state coordinates, checking the optimal path on the lines $K = K^*$ and $A = A^*$ supply us with important information.

Thus, checking the existence of a steady state should be a starting point of the analysis. At this state, $\dot{K} = \dot{A} = \dot{I} = \dot{a} = 0$ and, introduced in the basic eqs. (1a), (2a), (18) and (19), result in the conditions:

$$\frac{q_{K^*}MRq}{s_2+g_0\sigma_2K^*}-\sigma_2 = \frac{p_{A^*}q}{s_1+f_0\sigma_1A}-\sigma_1 = r, \tag{20}$$

$$\begin{bmatrix}\text{Net marginal}\\ \text{value of produc-}\\ \text{tive capital}\end{bmatrix} \begin{bmatrix}\text{Net marginal}\\ \text{value of demand-}\\ \text{creation capital}\end{bmatrix} \begin{bmatrix}\text{Opportunity}\\ \text{cost of}\\ \text{capital}\end{bmatrix}.$$

A necessary condition for the existence of a steady state is that there exist K^* and A^* for which the equalities in (20) hold and also that, at K^* and A^*, $MRq>0$ and $p_A>0$. We assume that steady state exists and proceed with the analysis of the dynamic path.

The mapping of the conditional $\dot{K} = 0$ curves, given $\dot{I}$, is derived from eq. (21) and described in fig. 7,

$$\dot{K} = 0\,|_{d\dot{I}=0}: g_0\dot{I}-(s_2+g_0\sigma_2K)(r+\sigma_2)+q_KMRq = 0, \tag{21a}$$

$$\left.\frac{dA}{dK}\right|_{\dot{K}=d\dot{I}=0} = \frac{g_0\sigma_2(r+\sigma_2)-q_{KK}MRq-q_K^2(\partial MRq/\partial q)}{q_K(\partial MRq/\partial A)}>0, \tag{21b}$$

$$\left.\frac{dA}{d\dot{I}}\right|_{\dot{K}=dK=0} = \frac{g_0}{q_K(\partial MRq/\partial A)}<0, \tag{21c}$$

$$\left.\frac{d\dot{K}}{dA}\right|_{d\dot{I}=dK=0} = \frac{q_K(\partial MRq/\partial A)}{(r+\sigma_2)g_0}>0. \tag{21d}$$

Eq. (21a) describes the conditional singular curve $\dot{K} = 0$ (given $\mathrm{d}\dot{I} = 0$), (21b) its slope, (21c) the shift of the curve caused by a change in $\dot{I}$, and (21d) the direction of the movement described by the horizontal arrows (in fig. 7 the horizontal arrows correspond to $\dot{K} = 0/\dot{I} = 0$).

As $\dot{I}$ is increased, the curves shift to the right until the curve $\dot{K} = 0/\dot{I} = \dot{I}_{max}$ coincides with the vertical line at $K = \bar{K}$ where, at $\bar{K}$, $MRq(A, \bar{K}) = 0$ (note that, by the separability assumption, $\bar{K}$ is unique for all A).

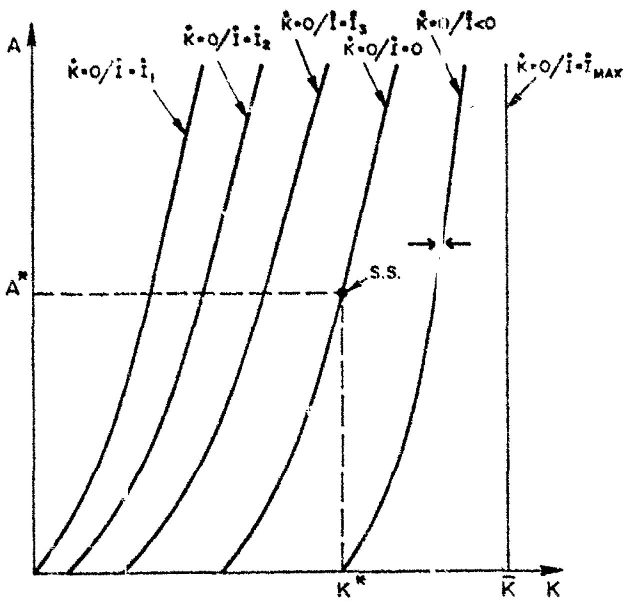


Fig. 7

Similarly, for $\dot{A} = 0$ curves given $\dot{a}$, we obtain

$$\dot{A} = 0\,|_{\mathrm{d}\dot{a}=0}: f_0\dot{a} - (s_1 + f_0\sigma_1 A)(r+\sigma_1) + p_A q = 0 \tag{22a}$$

$$\left.\frac{\mathrm{d}A}{\mathrm{d}K}\right|_{\dot{A}=\mathrm{d}\dot{a}=0} = \frac{q_K(\partial MRq/\partial A)}{(r+\sigma_1)(f_0\sigma_1) - p_{AA}q}, \tag{22b}$$

$$\left.\frac{\mathrm{d}K}{\mathrm{d}\dot{a}}\right|_{\dot{A}=\mathrm{d}A=0} = \frac{f_0}{q_K(\partial MRq/\partial A)} < 0, \tag{22c}$$

$$\left.\frac{\mathrm{d}\dot{A}}{\mathrm{d}A}\right|_{\mathrm{d}\dot{a}=\mathrm{d}K=0} = \frac{p_{AA}q}{f_0(r+\sigma_1)} - \sigma_1. \tag{22d}$$

In the case where $p_{AA} < 0$, the signs on the right-hand side of (22b) and (22d) are positive and negative, correspondingly; but if $p_{AA} \geqslant 0$ is large enough, the signs will interchange. The corresponding conditional singular curves $\dot{A} = 0$, given $\dot{a}$, are described in fig. 8 ($\dot{a}_1 > \dot{a}_2 > \dot{a}_3 > 0$).

Figs. 7 and 8 supply us with information about the optimal paths the firm will select. Starting from an initial point (K_0, A_0), the decision variables are $\dot{a}$ and $\dot{I}$ which determine uniquely the curves $\dot{A} = 0$ and $\dot{K} = 0$; and, therefore, the initial values of a and I, assuming continuity,[12] will, in fact, determine the behavior along the optimal path. Thus, for example, from the curvature of the $\dot{A} = 0$ map, it is clear that, if we start close to $K = 0$, the movement will be diagonally in the direction of the K axis and then either along the K axis until a certain level of K is reached where A starts to increase, or until an upward optimal path is

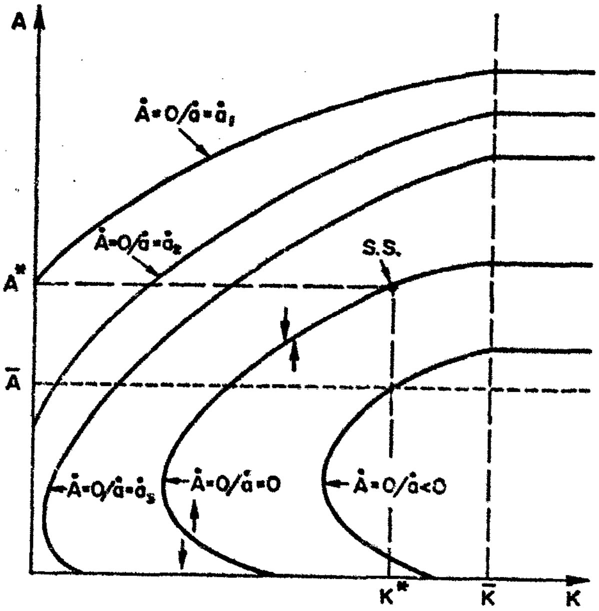


Fig. 8

encountered, which leads toward the steady state. Before we draw the map of the optimal path toward the steady state, let us consider fig. 9 where the (K, A) plane is divided by the two lines vertical at A^* and K^* correspondingly. Consider, first, the direction of the movement from any point along the vertical line at $A = A^*$ but $K < K^*$. Movement toward the steady state along this line (i.e., $\dot{A} = 0$) is impossible since, then, $a = \sigma_1 A$ and also $\dot{a} = 0$ so that eq. (18) cannot hold. If $a > a^* = \sigma_1 A$, it is immediate that the right-hand side of eq. (18) will increase with respect to the steady state since $K < K^*$ and $a > \sigma_1 A$. But at the steady state, equality holds with $\dot{a} = 0$; hence, for the equality to continue to hold, $\dot{a}$ has to be positive. In fact, since we started initially with $a > a^*$, a will

[12]Continuity of λ_i, A, and K [see Arrow–Kurz (1970) and Pontryagin (1964)] ensures continuity and differentiability of the controls by eq. (17) when equality holds, and then eqs. (18) and (19) ensure continuity of $\dot{a}$ and $\dot{I}$ in the same range (note that, when $A + a = 0$, a and $\dot{a}$ are not necessarily continuous).

continue to increase and will never reach a^* and enter a diverging process. We thus have a contradiction, and $a<\sigma_1 A$ must hold as the vertical arrow shows in fig. 9. For $K>K^*$ on the same line using the same reasoning, the vertical arrow will show the opposite direction. The horizontal arrows can be derived by applying similar arguments to eq. (19). Thus, all the information can be summarized in fig. 10 where the optimal paths from different initial points are described.

We note that either quadrant II or quadrant IV must be used in order to reach the steady state (K^*, A^*). The broken and solid lines trace the *unconditional* curves $\dot{A} = 0$ and $\dot{K} = 0$, respectively, which are, correspondingly, the locus of

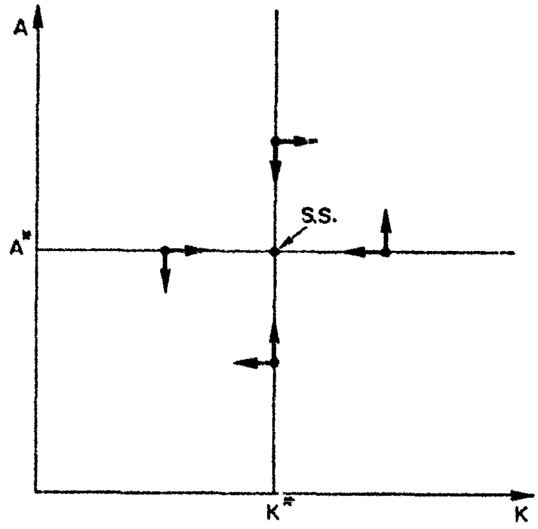


Fig. 9

all points on the optimal trajectory where $\dot{A}$ and $\dot{K}$ change their signs. Comparing the present case to the previous one, we note that, the cycle of an initial increase in the productive capital (K) – in order to transform part of it later into demand-creation capital – does not necessarily always occur when there are high adjustment costs involved. However, the influence of increasing marginal returns to demand-creation capital prevents growth in A up to a certain point. Moreover, if the firm starts with an initial positive level of A, it will depreciate it toward zero; then, from a certain point, A will increase at a very fast rate of growth, while K will either increase at a slower rate or, exactly as in the previous case, depreciate. Thus, the cycles obtained in the previous model appear in the extended model as well. These different phases in the behavior of the growing firm seem to be in close agreement with observed facts. The economic reasoning is quite simple. In order to keep the stock at its existing level, the firm should invert the amount of the depreciated stock; and a reduction in the stock is achieved by zero level of investment (negative net investment). If there is an excess in one of the stocks, negative net investment occurs and, in spite of the cost of adjustments, differences in the corresponding net marginal values of the two resources [eq. (20)] will cause 'consumption' of one resource by the other more remunerating resource.

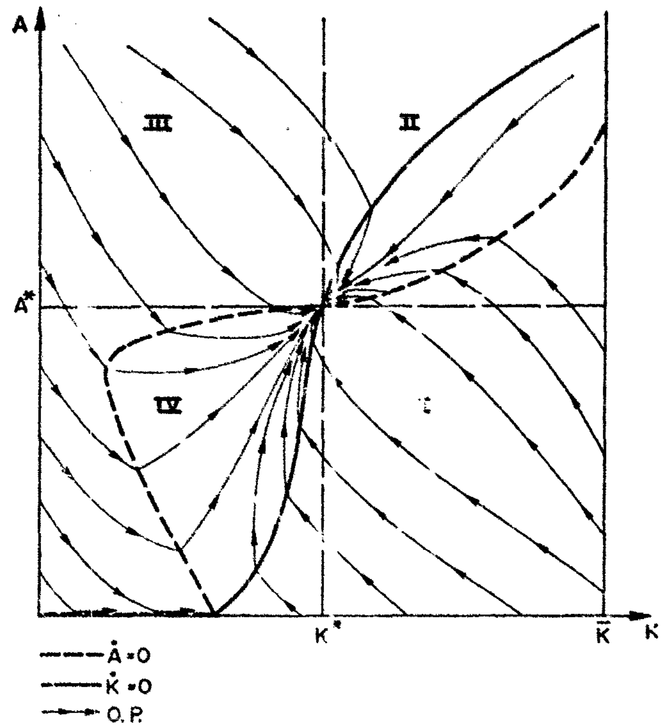


Fig. 10

Appendix A

The question of the sufficiency of the necessary conditions stated in eqs. (6) and (12) will be dealt with here. Note first that, if the initial conditions are such that the firm starts either in segment 1 or in segment 3 and the steady state is in the same segment the firm started from, we certainly have a concave integrand; and the necessary conditions are also sufficient [Arrow and Kurz (1970, pp. 45/6)].

In the general case, sufficiency can be shown considering an alternative formulation of the problem stated in (4):

$$\max \int_0^\infty [pq - w(c)]\, e^{-rt}\, dt$$

subject to:

$$G = A + K,$$

$$G \geqq A \geqq 0,$$

$$\dot{G} = C - \sigma G,$$

and

$$G + C \geqq 0.$$

This problem can be solved as a two-stage maximization. At the first stage, a decision is made on the allocation of K and A provided G is given; at the second stage, G and C are determined. The first stage is a simple static Kuhn–Tucker problem, while the second stage is the actual control problem with one state variable and one control.

The Kuhn–Tucker solution to the problem, max pq, subject to $A+K = G$ and $A \geqq 0$, will be:

$$p_A q \leqq q_K MRq, \tag{A.1}$$

which is a version of eq. (9) in the text. Equality will hold in (A.1) if and only if $A \neq 0$ and $K \neq 0$.

The equality condition will be necessary and sufficient (i.e., concavity of the objective function) if and only if

$$p_{AA}q - 2p_A q_K(1+\eta_{pq}) + q_K^2(\partial MRq/\partial q) + q_{KK}MRq < 0.$$

By the assumptions that the production function is concave and $1+\eta_{pq} > 0$, it is easy to verify that all terms except $p_{AA}q$ are always negative. The sign of p_{AA} depends on A. If A is such that, if equality (A.1) holds and $p_{AA}q > 2p_A q_K(1+\eta_{pq}) - q_K^2(\partial MRq/\partial q) - q_{KK}MRq$, then equality (A.1) is not a sufficient condition; by Kuhn–Tucker theorem, inequality will hold, i.e., $A = 0$, and the firm will be in phase I.

If G is such that equality (A.1) holds, though $p_{AA} > 0$ since

$$p_{AA}q < 2p_A q_K(1+\eta_{pq}) - q_K^2(\partial MRq/\partial q) - q_{KK}MRq,$$

then (A.1) is a necessary and sufficient condition, and both $K > 0$ and $A > 0$. This condition holds in phases II and III.

The necessary conditions for the control problem are then

$$w_c = \lambda, \tag{A.2}$$

$$\dot{\lambda} = \lambda(r+\sigma) - [\partial(pq)/\partial G], \tag{A.3}$$

where pq is a function of G and, in phase I, G satisfies $A = 0$ and $G = K$. In phases II and III, A and K are related through eq. (9), and their sum is equal to G. Now it can be verified that H^0 [the value of the Hamiltonian when c satisfies (A.2) in phase I and (A.2) and (9) in phases II and III] is a concave function of G.

In conclusion, there is an alternative direct way to calculate and compare all possible solutions as described in fig. 3. The argumentation which has a strong intuitive appeal runs as follows: A movement from any initial starting point toward the steady state involves moving from one 45° line to the next one in a way that will maximize the cash flow function. Now compare this by calculating the cash flow values of all points along a given 45° line. It is easy to show that the point of intersection between this line and the Q-curve maximizes the integrand for a given 45° budget line which proves our assertion. Note, too, that, if part of segment 2 has a negative slope of less than 45°, an immediate adjustment occurs in the level of A from zero to a certain positive amount. This is a clear example of jumps in the state variable [Arrow and Kurz (1970, p. 51)].

References

Arrow, K.J. and M. Kurz, 1970, Public investment, the rate of return, and optimal fiscal policy (Johns Hopkins Press, Baltimore, Md.) ch. 2.

Ball, R.J., 1968, Classical demand curves and the optimal relationship between selling costs and output, The Economic Record, 342–348.

Chetty, V.K., 1972, Necessary and sufficient conditions for a class of dynamic models with infinite time horizon, CORE Discussion Paper 7209, April.

Dhrymes, P.J., 1962, On optimal advertising capital and research expenditures under dynamic conditions, Economica.

Dorfman, R. and P.O. Steiner, 1954, Optimal advertising and optimal quality, American Economic Review, 826–836.

Gould, J.P., 1968, Adjustment costs in the theory of investment of the firm, Review of Economic Studies, 47–56.

Gould, J.P., 1970, Diffusion processes and optimal advertising policy, in: A. Phelps et al., Microeconomic foundations of employment and inflation theory (W.W. Norton and Co., New York).

Hahn, F.H., 1959, The theory of selling costs, Economic Journal, 293–312.

Hieser, R.O. and C.S. Soper, 1966, Demand creation: A radical approach to the theory of selling costs, Economic Record, 384–396.

Hochman, E., O. Hochman and A. Razin, 1973, Demand for investment in productive and financial capital, European Economic Review, 67–83.

Jacquemine A.P., and J. Thisse, 1972, Strategy of the firm and market structure: An application of optimal control theory, in: A. Cowling, Market structure and corporate behavior (Gray-Mills, London).

Lucas, R.E., 1967a, Optimal investment policy and the flexible accelerator, International Economic Review, 78–85.

Lucas, R.E., 1967b, Adjustment costs and the theory of supply, Journal of Political Economy, 321–334.

Nerlove, M. and K.J. Arrow, 1962, Optimal advertising policy under dynamic conditions, Economica, 129–142.

Pontryagin, L.S., V.G. Boltyanskii, R.V. Gambreldzi and E.F. Mishenker, 1964, The mathematical theory of optimal processes (Pergamon Press, Oxford).

Rao, A.G., 1970, Quantitative theories in advertising (John Wiley and Sons, New York).

Thompson, R.G. and M.S. Proctor, 1969, Optimal production, investment, advertising, and price controls for the dynamic monopoly firm, Management Science, 211–220.

Treadway, A.B., 1969, On rational entrepreneurial behavior and the demand for investment, Review of Economic Studies, 227–239.