THE PRODUCTION OF HUMAN CAPITAL AND THE LIFE CYCLE OF EARNINGS

YORAM BEN-PORATH*
University of Chicago

The application of capital theory to decisions on individual improvement, and in particular improvement of earning capacity, has provided a framework for the understanding of many aspects of observed behavior regarding education, health, occupational choice, mobility, etc., as rational investment of present resources for the purpose of enjoying future returns. The formulation by Friedman and Kuznets (1945) and the significant development of the theory by Becker (1962, 1964) and Mincer (1958, 1962) provided a novel view of the life cycle of earnings by linking it to the time profile of investment in human capital: People make most of their investments in themselves when they are young, and to a large extent by foregoing current earnings. Observed earnings are therefore relatively low at early years, and they rise as investment declines and as returns on past investments are realized. The main reason why investment is undertaken mostly by the young is that they have a longer period over which they can receive returns on their investment. The purpose of this paper is to combine that part of the argument concerning the demand for human capital with a more explicit treatment of the supply, or cost conditions, facing the individual.

It is hard to think of forms of human capital that the individual can acquire as final goods—he has to participate in the creation of his human capital. His own abilities, innate or acquired, the quality of co-operating inputs, the constraints and opportunities offered by the institutional setup—all determine the "technology," or the production function.[1] Together with the relevant factor prices, the properties of the production function determine the optimal way in which any quantity of human capital is to be produced and determine the cost of production. I shall show how the production function (through supply or cost conditions) enters into the determination of the optimal path of investment, analyze some of the implications for the individual's allocation of time, and demonstrate how the life cycle of earnings can be affected by various properties of the production function.

The basic model generates some of the qualitative characteristics of the observed life cycle of earnings—typically, an initial period of no earnings followed by a period in which earnings rise at a declining rate and, eventually, decline.

* This paper is based on part of my Ph.D. dissertation submitted to the Department of Economics, Harvard University, February, 1967. I benefited from discussing earlier drafts with my adviser, Professor Simon Kuznets, and with Professor Rodney A. Dobell, Gur Ofer, and Christopher A. Sims; I received many useful comments from Professors Gary S. Becker, Zvi Griliches, Theodore W. Schultz, Lester G. Telser, and other friends. Financial support was provided by a Ford Dissertation Fellowship and a grant from the Ford Foundation to the University of Chicago for Research on Investment in Human Capital.

[1] Such a production has been recently introduced also by Becker (1966).

Actual, or observed, earnings turn out to be always lower, to change faster, and to peak at a later age than the attainable maximum, the earning capacity of the individual. The reverse is true of the relation between observed earnings and earnings net of all investment costs: the former are always higher, change slower, and peak at an earlier date than the latter. The model thus specifies the nature of the bias that may exist when earnings are used, as they often have been, to infer changes in productive capacity with age.

I. THE MODEL

In addition to the assumptions incorporated in the production function, to be discussed later, I assume:

1. Individual utility is not a function of activities involving time as an input.
2. There is a fixed amount of time to be allocated every period to activities that produce earnings and additions to the stock of human capital.
3. The stock of human capital, K, of which every individual has some initial endowment, is homogeneous and subject to an exogenously given rate of deterioration, δ.
4. The stock of human capital is not an argument in the individual's utility function.
5. Unlimited borrowing and lending take place at a constant rate of interest, r.

The first two assumptions express the fact that leisure is ignored in this analysis. In conjunction with the other assumptions they allow the partition of the individual's decision-making into two stages:

a) The individual allocates the given periods of time between earning and producing human capital and finds the corresponding outlays on investment that maximize the discounted value of any time t of disposable earnings (defined below) from t to T, where T is assumed with certainty to be the end of life.

b) Given the optimal time path of disposable earnings, the individual decides on the timing of the consumption. This is the point of departure of the life-cycle theories of consumption, which take the stream of earnings as given and explain consumption as dependent upon it.

The stock of human capital is defined as a concept analogous to "machines" in the case of tangible capital. There is a market in which the services of human capital are traded, and a rental, a_0, is determined for the services of a unit of human capital, K, per unit of time. The sum of the services offered in the market by various individuals is an input into the production of other goods and services. It may well have a diminishing marginal productivity, which will cause a downward-sloping aggregate demand curve for the services of human capital. Each individual is assumed, however, to possess only a small fraction of the total homogeneous stock of human capital in the economy and is regarded as a perfect competitor facing a given rental a_0, which is independent of the volume of services that he offers in the market. Earning capacity at time t, Y_t, is therefore the maximum services of human capital the individual can offer in the market valued by the rental a_0.

$$Y_t = a_0 K_t . \quad (1)$$

Let E_t be disposable earnings in period t—the portion of current earnings disposable for purposes of consumption or the purchase of non-human assets. It may be smaller than earning capacity if the individual engages in production of human capital; other uses for time are excluded by assumption. The difference

$(Y_t - E_t)$ is the cost of investment I_t; it depends on the production function and on input prices.

Let (2) be the production function of human capital:

$$Q_t = \beta_0 (s_t K_t)^{\beta_1} D_t^{\beta_2}, \qquad (2)$$

where β_1, $\beta_2 > 0$ and $\beta_1 + \beta_2 < 1$; Q is the flow of human capital produced. D is the quantity of purchased inputs, the price of which is denoted by P_d; s_t is the fraction of the available stock of human capital allocated to the production of human capital, so that $s_t K_t$ is the quantity of human capital allocated for the purpose. If activities are not "mixed," that is, if there is no joint production of earnings and of human capital, then s_t is also the proportion of time devoted to the production of human capital. The fraction s_t is constrained by the condition

$$0 \leq s_t \leq 1 . \qquad (3)$$

(The properties of the production function including the assumption of decreasing returns to scale are discussed in a subsequent section.) The rate of change of the capital stock is given by (4):

$$\dot{K}_t = Q_t - \delta K_t \qquad (4)$$

(a dot above any variable indicates a derivative with respect to time), where δ is the rate by which the stock of human capital deteriorates. Equations (1) and (4) imply that a unit of capital can be used from the moment that it is produced.

Investment costs have two components:

$$I_t = a_0 s_t K_t + P_d D_t , \qquad (5)$$

that is, (*a*) opportunity costs, or "foregone earnings" (the value of the productive services withdrawn from the market), and (*b*) the direct costs of purchased goods and services. Minimizing I_t with respect to s_t and D_t, subject to (2) and ignoring (3), we get (6) as a condition for a minimum of (5):

$$\frac{a_0 s_t K_t}{P_d D_t} = \frac{\beta_1}{\beta_2}. \qquad (6)$$

From (2) and (6) substituted into (5) we get investment costs as function of output:

$$I_t = \frac{\beta_1 + \beta_2}{\beta_1} a_0 \left(\frac{\beta_1 P_d}{\beta_2 a_0}\right)^{\beta_2/(\beta_1+\beta_2)} \times \left(\frac{Q_t}{\beta_0}\right)^{1/(\beta_1+\beta_2)} . \qquad (7)$$

The objective of the individual at any time t is to maximize the present value of his disposable earnings:

$$W_t = \int_t^T e^{-r\nu} [a_0 K(\nu) - I(\nu)] d\nu . \qquad (8)$$

The objective expressed by (8) is treated as applicable to an individual from birth; up to a given age actual decisions are made by his parents, and if they were to take into account the full future life of their child, (8) still expresses the relevant maximand under the assumptions of the model. If the parents do not take into account the full economic life of their child, or in the extreme case where only the period of attachment to the present household is considered, less investment would be undertaken, and the age of entry into the labor force would be lower.

The problem posed here is suitable for treatment by the techniques of optimal control. Three phases are suggested by the constraints on s, the fraction of human capital, or time, allocated to the production of human capital: (i) The available stock of human capital K_t, even when fully allocated to produce human capital, is not large enough to provide the flow of services demanded given the relevant prices. The upper bound on s is thus an effective constraint. (ii) The available stock is large enough to supply the services demanded and more, so that $0 < s < 1$ and the services of human capital are truly a variable factor. (iii) The stock of capital is too big so that the optimal policy requires more disinvestment than is feasible through deterioration, that is, to produce negative quantities of human capital. Here s_t is constrained by its lower bound.[2]

The first phase is by definition a period in which no human capital is allocated to the market so that no earnings are realized. Using conventional tools, we shall first analyze the later phases in which $s < 1$ and then shall return to discuss the first.

Differentiating (7) with respect to Q we get (9), the marginal cost of producing human capital; it is a rising function of the quantity produced starting from the origin and is independent of the size of the existing stock of capital of the individual.

$$MC_t = \frac{\alpha_0}{\beta_0\beta_1}\left(\frac{\beta_1 P_d}{\beta_2\alpha_0}\right)^{\beta_2/(\beta_1+\beta_2)} \times\left(\frac{Q_t}{\beta_0}\right)^{[1/(\beta_1+\beta_2)]-1} . \quad (9)$$

The value at time t of acquiring an additional unit of human capital is the discounted value to that time of the additions to earnings that the undepreciated part of this unit will bring about. This is the "demand price" of human capital, given by (10). This price is

$$P_t = \alpha_0\int_t^T e^{-(r+\delta)\nu}d\nu = \frac{\alpha_0}{r+\delta}[1 - e^{-(r+\delta)(T-t)}], \quad (10)$$

independent of the number of units added or the existing stock. It is a declining function of time because of the presence of the finite horizon.[3] The optimal production of human capital is determined by equating the marginal cost to the price. Equating (9) and (10) we can solve for Q as shown in equation (11).

[2] In terms of the techniques developed by Pontryagin *et al.*, the maximization of (8) subject to (2), (3), and (4) involves the maximization of the Hamiltonian

$$H: e^{-rt}[(1-s)\alpha_0 K - P_d D] + q(Q - \delta K), \quad (1')$$

where q is the discounted shadow price of investment in human capital

$$\dot{q} = -\frac{\partial H}{\partial K} = -e^{-rt}(1-s)\alpha_0 - q\left(\beta_1\frac{Q}{K} - \delta\right).$$

In terms of current prices $\psi = qe^{rt}$

$$\dot{\psi} = -(1-s)\alpha_0 - \psi\left[\beta_1\frac{Q}{K} - (\delta + r)\right]; \quad (2')$$

Transversality condition:

$$\psi(T)K(T) = 0. \quad (3')$$

First-order conditions for the maximization of (1′):

$$\frac{\partial H}{\partial s} = -\alpha_0 K e^{-rt} + \psi e^{-rt}\frac{Q}{s}\beta_1 \geq 0. \quad (4')$$

$$\frac{\partial H}{\partial D} = -P_d e^{-rt} + \psi e^{-rt}\frac{Q}{D}\beta_2 = 0. \quad (5')$$

A systematic analysis of a variant of this problem, using the techniques of optimal control, is pursued in an unpublished note by Eytan Sheshinski.

[3] In terms of the framework of n. 2 we are operating in the phase where (4′) holds exactly as equality. Substituting (4′), (5′), and (2) into (2′), and given (3′), we get a differential equation to which (10) is a solution.

$$Q_t = \beta_0\left(\frac{\beta_0\beta_1}{r+\delta}\right)^{(\beta_1+\beta_2)/(1-\beta_1-\beta_2)}\left(\frac{\beta_2\alpha_0}{\beta_1 P_d}\right)^{\beta_2/(1-\beta_1-\beta_2)}[1 - e^{-(r+\delta)(T-t)}]^{(\beta_1+\beta_2)/(1-\beta_1-\beta_2)} = N[1 - e^{-(r+\delta)(T-t)}]^{(\beta_1+\beta_2)/(1-\beta_1-\beta_2)} \geq 0 . \quad (11)$$

The production of human capital, Q, the gross additions to the stock, are always positive except when $t = T$. At the date of compulsory retirement, T, the human capital loses its value, $P(T) = 0$ (see eq. [10]). No production is undertaken at period T, and the stock of human capital is reduced by sK (see eq. [3]). The point $t = T$ lies in the third phase. The reason why only this point lies in this phase is that for any earlier age, $t < T$, the demand price of human capital is positive, and the fact that the marginal cost curve starts from the origin insures that some positive quantity will always be produced. This would be true of any production function homogeneous of a degree less than 1 in these inputs.

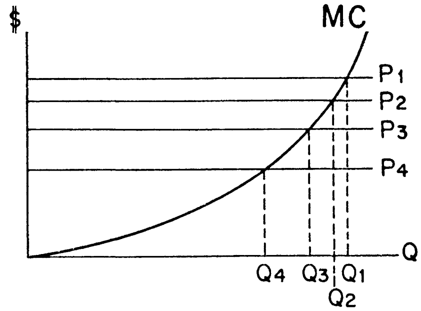

FIG. 1

The market prices that enter into the determination of the optimal production of human capital Q are: r, the rate of interest; a_0, the rental on human capital; and P_d, the price of purchased inputs. Differentiating (11) with respect to these shows that Q varies inversely with the rate of interest and with the price of purchase inputs and directly with the rental on human capital. It is, however, only the *ratio* between the last two that is important. The elasticity of Q with respect to the relative price a_0/P_a is $\beta_2/(1 - \beta_1 - \beta_2)$. When $\beta_2 = 0$, that is, when purchased inputs do not enter into the production of human capital, changes in a_0, the price of the services of human capital, do not affect the quantity produced. The obvious reason is that in this case, when the services of human capital are the only input, a rise in a_0 raises marginal cost by exactly the same amount that it raises the value of a unit of human capital. Needless to say, all this discussion refers to alternative stationary price levels when the current prices are expected to prevail throughout the individual's lifetime. Q_t is larger the larger the parameters of the production function β_0, β_1, β_2; the greater the length of economic life T; and the lower the rate of deterioration, which has exactly the same effect as the rate of interest.

As t rises, the flow of Q produced declines, as indicated by (12), the derivative of Q with respect to t:

$$\dot{Q} = \frac{\beta_1 + \beta_2}{1 - \beta_1 - \beta_2} N[1 - e^{-(r+\delta)(T-t)}]^{[(\beta_1-\beta_2)/(1-\beta_1-\beta_2)]-1} \times e^{(r+\delta)(T-t)}[-(r+\delta)] \leq 0 . \qquad (12)$$

What we have here is a marginal cost curve as a function of Q starting from the origin and remaining stationary through time, and a perfectly elastic demand curve which slides down with time, thus creating a pattern of positive but declining optimal quantities of Q to be produced (Fig. 1). If the horizon were infinite, the demand curve would not slide down, and there would be at this phase one stationary rate of production of human capital (given by N in eq. [11]). Note that we are saying something about the *gross* additions to stock of human capital but not about the *net* additions, which depend also on the rate of depreciation and the size of the existing stock. If the initial stock is very high, net additions may be negative from an early age. If we think of the normal case as one where the capital stock does rise over a period, eventually as gross additions become very small and the stock becomes large this must be reversed, and toward the end of life, T, the stock will decline, if there is any deterioration.

What are the implications of this pattern of investment for the life cycle of earnings? Disposable earnings were defined as the difference between earning capacity and investment outlays; their rate of change over time is given by (13).

$$\dot{E}_t = a_0 \dot{K}_t - \dot{I}_t \, . \qquad (13)$$

Let us define "observed earnings" $\hat{E}$ to be actual earnings realized in the labor market (14).

$$\hat{E}_t = a_0 K_t - \frac{\beta_1}{\beta_1 + \beta_2} I_t \, . \qquad (14)$$

They are larger than disposable earnings by the direct costs $P_d D$ and smaller than earning capacity by foregone earnings. Their rate of change over time is given by (15).

$$\dot{\hat{E}} = a_0 \dot{K}_t - \frac{\beta_1}{\beta_1 + \beta_2} \dot{I}_t = a_0 Q_t - \frac{\beta_1}{\beta_1 + \beta_2} (MC_t) \dot{Q}_t - a_0 \delta K_t \, . \qquad (15)$$

$\dot{I}$ is always negative (except when $t = T$), so that the change in observed earnings is always algebraically larger than the change in earning capacity and smaller than the change in disposable earnings

$$\left(\frac{\beta_1}{\beta_1 + \beta_2} < 1 \right) .$$

Thus, the curve of observed earnings exaggerates the rate of increase of earning capacity when the latter increases and understates its decline when it declines. In particular, when we observe an individual at the peak of his earnings he is already past the peak of his productive capacity. The existence of a downturn in observed earnings is here only a consequence of depreciation (recall that interactions with "leisure" are excluded by assumption). If depreciation is zero, there is always, except at T, an increase in the three types of earnings, and at each point in time their rank by rate of change will be the reverse of their rank by level.

From the second derivative of observed earnings with respect to time (16), we learn more about the shape of the life cycle of earnings:

$$\ddot{\hat{E}} = Q_t \left\{ a_0 - \frac{(r + \delta) MC_t}{1 - e^{-(r+\delta)(T-t)}} \frac{\beta_1}{\beta_1 + \beta_2} \times \left[1 - \frac{1}{1 - \beta_1 - \beta_2} e^{-(r+\delta)(T+t)} \right] \right\} - a_0 \delta Q_t + a_0 \delta^2 K_t \, . \qquad (16)$$

If there is no deterioration, that is, $\delta = 0$, the ever rising curve of observed earnings is always concave from below. This is also true of the rising portion of the curve when $\delta > 0$. If $\beta_1 + \beta_2 > \frac{1}{2}$, there will be a certain range in the vicinity of $t = T$ where the curve is convex from below.[4] All these qualitative results hold for any production function homogeneous of a degree smaller than 1 in sK and in D.

[4] $\dot{Q}$ is negative, and the term in braces is positive: from the equality between (11) and (12)

$$a_0 = \frac{r + \delta}{1 - e^{-(r+\delta)(T-t)}} MC \, .$$

The right-hand side of this equality is here multiplied by a term always smaller than +1 and deducted from the left side so that the difference is positive and the first term is negative. If $\delta = 0$ and there is no deterioration, the other two terms vanish, and the curve of observed earnings is always concave from below. This is also true of the curve of disposable earnings with 1 substituted for $\beta_1/(\beta_1 + \beta_2)$. When $\delta > 0$, concavity is assured for the range where $\dot{K} = Q - \delta K > 0$ If $\beta_1 + \beta_2 > \frac{1}{2}$, the first term tends to zero as $t \rightarrow T$; in the vicinity of $t = T$, $\dot{K} < 0$, so in this case the earnings curve will have there a shape that is convex from below.

The preceding discussion relates to the phase where the optimal s, the fraction of human capital or time allocated to the production of human capital, is smaller than 1. As indicated before, the maximization of (8) may, however, require that there will be a phase of complete specialization in the production of human capital in which $s = 1$ and the size of the stock is an effective constraint.

Movement along the marginal cost curve (7) by increasing Q implies an increase in the required sK. Once $s = 1$ is reached, larger rates of production can be achieved only by combining more purchased inputs with a fixed flow of services of human capital, thereby increasing costs at a higher rate than is implied by (7). The marginal cost curve described by (7) is a long-run envelope from which steeper marginal cost curves rise up, corresponding to alternative levels of the available stock of human capital. When people are young the value of a unit of human capital is only negligibly affected by the finite horizon T, so demand for human capital is relatively high. On the other hand, the available stock is still small; therefore the marginal cost curve rises up from the long-run curve, (7), at relatively small output, so that at a young age production is likely to occur at outputs where $s = 1$.

Note that the demand price given by (10) is now only the lower limit of the true shadow price of human capital. An increase in the stock of human capital implies a reduction in future costs of production of human capital. The precise value of this depends on the future quantities of Q that will be produced between t and t^*, the date of transfer from the first to the second phase. While in phase (ii), where $0 < s < 1$, we were able to determine the optimal Q_t by the intersection of demand and supply schedules independent of the future levels of Q, here only a complete, dynamic program in which the effects of present actions on future condition are explicitly taken into account will generate an optimal path of investment.

Phase (i) is identified by the absence of earnings. Another distinguishing characteristic of this phase is that the relaxation of the constraint K_t which lowers the marginal cost curve may induce a pattern of increasing investment with age, both through the allocation of the services of a growing stock of human capital for the purpose and through higher direct expenditure on D, justified by the higher marginal productivity of D that the increased K brings about. A phase of increasing direct costs in the early period when no earnings are realized is certainly not inconsistent with the real world.[5]

The age t^* when the individual enters the second phase and leaves the first is of great importance, because this is the date at which positive earnings begin to be realized. From the preceding discussion it should be clear that $K(t^*)$ is a declining function of t^*.[6] K_0 is the initial endowment of human capital, which, following the definition of K, is proportional to the maximum earnings the individual can realize when he "starts." All other parameters given, the larger this initial endowment, the earlier will be the date t^* at which specialization stops, noting that K_0 is an initial endowment only in terms of the capacity to earn, while we hold constant the capacity

[5] An explicit solution of our problem for the case in which $\beta_2 = 0$ is simple but not interesting. The interesting aspects of the problem derive from the possibility of using purchased inputs.

[6] This can be verified by substituting $s = 1$ into (4′), letting it hold as an equality, and substituting into it also the value of Q and ψ using (2), (5′), and the solution of (2′) referred to in n. 3.

to increase human capital thus defined, which is reflected in the β's of the production function.

II. THE PRODUCTION FUNCTION

ROLE AND IMPLICATIONS

The technology which the individual faces when he makes decisions about investing in himself is a complicated system of technical and institutional relationships covering a wide spectrum of activities including formal education, acquisition of skills on the job, child care, nutrition, health, etc. By writing down a simple production function of the sort used here we are attempting, not to reproduce this system, but only to provide a framework within which some of the possible characteristics of the technology can be considered and their implications studied.

A. PURCHASED INPUTS, TIME, AND THE SERVICES OF HUMAN CAPITAL

Both the theory and the measurement of investment in human capital (see Schultz, 1963) emphasize the importance of foregone earnings alongside direct cost. A composite surrogate for direct cost here is D, the index of purchased inputs which stands for anything from tuition to vitamin pills. The presence of opportunities for some substitution between purchased and own inputs has a "smoothing" effect on behavior and helps the individual overcome the constraints of his limited time and sometimes modify the effects of its increasing cost.

The nature of and the role played by own inputs require some further discussion. In the production function so far considered, own inputs are represented by the product sK. K is the total stock of human capital; s is described as a fraction that can take any value between 0 and 1. We shall first regard s as a measure of the allocation of time to the production of human capital, although this interpretation is not necessary.[7]

At any time t the stock of capital K_t is given. If we were to assume that any activity the individual engages in produces either human capital or earnings but not both, then the allocation of time between these two types of activities is also the allocation of the services of the existing stock of human capital. The larger the stock of human capital, the larger the earnings per unit of time that the individual could get in the market and therefore the higher the foregone earnings from diverting a unit of time away from the market (see Becker, 1965). Whether this should or should not affect the relative attractiveness of non-market activities hinges on whether the change that made an hour in the market more rewarding also made more productive an hour outside the market, in our case an hour of producing human capital.

The question is whether the real production relation involved is stable in terms of time or in terms of some other variable. The way we *defined* human capital was to make the production of earnings stable in terms of its services, rather than in terms of time as such. By making (sK) the relevant input in the production function of human capital we are also *asserting* that this other process is stable in terms of the services of human capital rather than in terms of time as such. The main implication of this formulation in terms of the alloca-

[7] Note that time enters in two completely different ways—t, which moves the individual along his life cycle and which is being treated as continuous, and s_t, which is a measure of the allocation of time at any point in t. If we were to treat t as an index of discrete periods, and if s were understood as referring to the allocation of each such period, this dual meaning probably would have posed no problem.

tion of time is that, when K is higher, the higher costs of an hour diverted away from the market (because of the higher foregone earnings) are exactly matched by the greater productivity of time in the production of human capital, so that the marginal cost curve of producing the latter is independent of the stock of human capital in the range where the constraint on s is not effective.

By arguing that in the production of human capital the services of human capital rather than time as such are relevant we gain some analytical simplicity, because beyond the period of complete specialization in the production of human capital the benefits associated with acquiring a unit of human capital, the addition to future disposable earnings, do not depend on the future allocation between the market and the production of human capital. This is the reason why the dynamic programing problem that was referred to in relation to phase (i) degenerated into a much simpler decision problem in phase (ii). Analytical convenience is, however, no substitute for relevance. In a general production function both time as such and human capital would appear as inputs, and if we consider again the Cobb-Douglas case, s and K may appear with different coefficients. Thus, consider (17):

$$Q = \beta_0 s^{\gamma_1} K^{\gamma_2} D^{\beta_2} = \beta_0 s^{\gamma_1 - \gamma_2} (sK)^{\gamma_2} D^{\beta_2}\,; \quad (17)$$

$$0 < \gamma_1,\ \gamma_2,\ \beta_2 < 1\ .$$

The corresponding marginal cost curve is (18):

$$MC_t = \frac{a_0}{\beta_0 \gamma_1} \left(\frac{\gamma_1}{\beta_2} \frac{P_d}{a_0} \right)^{\beta_2/(\gamma_1+\beta_2)} \times \left(\frac{Q_t}{\beta_0} \right)^{[1/(\gamma_1+\beta_2)]-1} K_t^{(\gamma_1-\gamma_2)/(\gamma_1+\beta_2)}\ . \quad (18)$$

If $\gamma_1 = \gamma_2$, we get an expression identical to (2), with γ_1 replacing β_1. If $\gamma_1 > \gamma_2$, the marginal cost curve is shifting upward as K increases. This can be viewed as a case where time plays an independent part in the production of human capital or, more directly, as a case where the larger productivity of time in the market indicated by an increase in K is not completely matched by a larger productivity in the production of human capital, so that the cost of the latter in terms of the former rises. The case where $\gamma_1 = \gamma_2$ implies, for example, that the more highly educated person is also better equipped for learning, so that his higher opportunity cost is matched by the greater amount of skills that he can acquire per hour. If this is not so, then the situation that we are describing now is relevant. A limiting case would be one in which human capital does not at all affect the ability to produce more human capital, $\gamma_2 = 0$.

In the basic model considered before, the decline over time in investment beyond the period of complete specialization is brought about by the downward drift, due to the approaching horizon T of the demand function, along a stationary, upward-sloping marginal cost curve. In the situation now described, as K increases the cost curve shifts upward. In young age, when $T - t$ is large, the quantitative effect of changes in t on the demand price is small, and the shifts in the cost function may be more important for changes in investment than in the downward drift of the demand curve. The role of purchased inputs here is also clear—the larger the β_2, the smaller the effect of K on the marginal cost.

The other case, in which $\gamma_2 > \gamma_1$, also cannot be ruled out. Here capital accumulation reduces the cost of producing human capital, and it is possible even in phase (ii) to have a stretch of time over which investment rises rather than declines (the downward shift in the costs being more important than the declines in demand). The corresponding life cycle may then have an early convex portion. Eventually the declining marginal productivity of K (due to $\gamma_2 < 1$) and the growing effect of the approaching horizon T will turn the rise in investment into a decline, and the life cycle will have the familiar S-shape.

B. THE ROLE OF RISING COSTS

With a homogeneous stock of human capital, the services of which are sold at a fixed price in a competitive market, both the determination of a finite desired stock and the speed in which the available stock is adjusted to the desired level depend on the cost of acquiring human capital. If human capital could be acquired at a fixed (or declining) cost without limitation, that is, if the supply schedule were perfectly elastic or declining, the desired stock would be either zero or infinite, and the optimal adjustment would be instantaneous. In the case of the demand for tangible capital, costs of adjustment are sometimes introduced to explain investment as a function of the interest rate. In the case of the aggregate economy, it is the rising cost of investment goods in terms of consumption goods that provides the negative slope of the aggregate demand function of investment (the marginal efficiency of investment curve) and a finite rate of investment.

The nature of human capital, the fact that it has to be produced by the individual, makes for some similarity in the considerations involved and provides a natural basis for dealing with the question of what makes the rate of investment and the attained stock of human capital finite. One form in which constraint on the rate of investment could come about is from individual capacity limitations solely. Thus, consider the case in which (2) $\beta_1 + \beta_2 = 1$, $\beta_1 > 0$, and $\beta_2 \geq 0$. In every period the available stock of human capital K_t is given. Costs would be constant up to that level of output which implies $s = 1$, that is,

$$Q = \beta_0 K \left(\frac{\alpha_0}{P_d}\frac{\beta_2}{\beta_1}\right)^{\beta_2}.$$

For larger outputs marginal costs would rise (if $\beta_2 > 0$) or become infinite (if $\beta_2 = 0$). In either case, given the nature of the demand for human capital, the optimum rate of investment would always imply a value of 1 or 0 for $s > n$, the allocation of all the services of the available stock of human capital either to the production of human capital or to the labor market, but not to both. Because of the declining demand over time for human capital, a period of complete specialization in the production of human capital (phase [i]), if it comes, must precede the period of complete specialization in market work (phase [iii]). The resulting life cycle of (observed) earnings would therefore have a portion of zero earnings and then a jump to some positive level, which would be stationary in the absence of depreciation and declining in its presence.

Neither this implied life cycle of earnings nor the behavior to which it is related is supported by what we know of the real world. The basic model allows for an initial period of complete specialization in the production of human capital and, correspondingly, a period of zero observed earnings. But if we assume that the sum of the production elasticities of the variable inputs is smaller than 1 ($\beta_1 + \beta_2$ in eq. [2] and $\gamma_1 + \beta_2$ in eq. [17], the more general formulation), marginal cost rises continuously even before the capacity constraint is reached. This is the source of the existence of a phase (ii), a period in which $0 < s < 1$, when the individual engages simultaneously, or alternately, in work in the market and in the production of his human capital. (A more complicated functional form could have allowed increasing returns to scale at small outputs and eventually decreasing returns. This may be more plausible and still would provide the eventual check on the rate of investment.)

In this framework, where there is not "automatic" growth in earnings, increase in earning can come only from the allocation of more services of human capital to the market. This can be either because of a reduction in what is allocated to investment (s_t) or because of an increase in the total stock available (K). In order for some positive earnings to be observed there must be some work in the market; in order for earnings to decline there must be some investment that either increases from period to period the total available stock or, by itself declining, raises the fraction allocated to the market. The closer the sum of the production elasticities of the variable factors to 1, that is, the closer we get to constant returns to scale in terms of the variable factors, the more concave is the life cycle of earnings; the case of constant marginal cost in which the life cycle of earnings is a one-step function is the limiting case.

Throughout the preceding discussions s_t has been interpreted as a parameter of the allocation of time. If there are, however, activities in which earnings and human capital can be jointly produced, then s_t loses this meaning. We have many examples of learning on the job in which the time spent on the job cannot be allocated in any meaningful sense between pure work and pure learning; jointness is too prevalent to be excluded. We can still think of the individual as being faced with a production frontier indicating the possible combination of flows of earnings and of human capital that he can produce. Movements along this frontier, however, would not necessarily represent shifts between pure activities but, rather, represent movement between activities, each offering a different mix of earnings and additions to productive capacity. Shifts along this frontier are represented by the control s_t, and they may involve a change in occupations, of jobs within the occupation, or of function with the job. The less numerous and close together are these alternative combinations, the smaller is the justification for the continuous differentiable frontier implied by our formulation. Discontinuities and kinks in this production frontier are translated into kinks and jumps in the life cycle of earnings. In either case, it is clear that s_t becomes an empty concept once it has lost its link with the observable phenomenon of the allocation of time.

III. THE MODEL AND AGGREGATE EARNING PROFILES

The preceding section clarifies the role played in this framework by increasing costs in the explanation of a gradually rising portion in the *individual* life cycle of earnings. The purpose of the present section is to show that even within this framework an explanation of cohort life cycles of earnings, or of the cross-section profile of earnings, can be provided without the assumption of rising costs. Thus, assume for simplicity that $\beta_2 = \delta = 0$ and $\beta_1 = 1$. By the argument presented in the preceding section, the individual life cycle will be a stepwise function.

The marginal (and average) cost of

producing human capital is (α_0/β_0), β_0 differing among people. The higher is β_0, the later comes the date t^* when the downward-drifting demand price crosses it from above—which is the date when people jump from observed earnings of zero to their earning capacity. By equating α_0/β_0 with P (10) we get (19):

$$t^* = T + \frac{1}{r}\ln\left(1 - \frac{r}{\beta_0}\right). \qquad (19)$$

The time path of the stock of human capital is given by

$$\begin{aligned} K_t &= K_0 e^{\beta_0 t} \quad \text{for} \quad 0 \le t \le t^* ; \\ K_t &= K_0 e^{\beta_0 t^*} \quad \text{for} \quad t^* \le t \le T . \end{aligned} \qquad (20)$$

Observed earnings are zero for $0 < t < t^*$ and are equal to earning capacity afterward.

$$\begin{aligned} E_t &= \alpha_0 K_0 e^{\beta_0 t^*} \\ &= \alpha_0 K_0 e^{\beta_0 \{T + (1/r)\ln[1-(r/\beta_0)]\}} . \end{aligned} \qquad (21)$$

Using (21) we can see that, given $\alpha_0 K_0$, earnings on the time of "emergence," t^*, and beyond it, are an increasing convex function of t^*. If in a given cohort K_0 and β_0 are not negatively correlated, we can expect to observe a rising curve as we follow the earnings history of the cohort through time (or as we examine the cross-section profile of earnings of a completely static population, with stationary distributions by β_0 and K_0, in a static economy). The exact form of the curve depends on the distribution of the cohort by β_0; reasonable distributions by β_0 can generate a curve with the familiar S-shape.

SUMMARY AND SOME OPEN QUESTIONS

To the theory of investment in man we have added here a production function of human capital and explored some of the implications of its properties for the optimal path of accumulation of human capital and the life cycle of earnings. The concepts and parameters introduced raise again some familiar problems and are used to focus attention on some others. In lieu of a summary we shall now refer to a few of these.

1. The particular definition of human capital K used is a measure of a quantity of a source of productive services. It is the stock that produces labor services in "standard units" and is thus the analogue of "machines" in the case of tangible capital. This should be sharply distinguished from W_t (defined in [8]), which is the *value* of the individual as a productive factor. It is affected by a broader view of the individual's productivity, by his durability, and by the rate of interest, and it can be misleading as a quantity measure of an input in any given point in time. Being a part of the individual's net worth (and the price that he could get for himself in a competitive slave market), it is probably better described as "human wealth."

2. The speed of adjustment, the rate at which the individual increases the stock of human capital, determines the ultimate size of this stock. The considerations affecting the speed of adjustment and the path of investment with age merit an explicit analysis. The mechanism provided here relies on the introduction of dependence between the marginal cost of producing human capital and the rate of production (a justification for this can come from the "learning curve" considerations). Beyond a certain point the saving in costs from postponing investment to the next period compensates for the loss of the returns that would have come from getting the unit of capital a period earlier.

Rising costs result here from a certain specification of the technology. The properties of the production function and the parameters introduced draw attention to certain distinctions that have to be made when the conglomerate of abilities and external conditions is considered and the implications of its properties are explored.

3. In discussing the "initial endowment" of individuals a different role is played by the initial endowment in terms of the ability to earn in the market (K_0) and the abilities to produce additions to earning capacity. These consist of the ability to make efficient use of inputs purchasable in the market, D (expressed by β_0 and β_2), and of one's own time (β_0, β_1, or γ_1 in [17]). A related issue is to what extent the individual, as he increases his human capital, K, also raises his efficiency in the production of K (in eq. [17] this is expressed by the distinction between γ_1 and γ_2). This question is related to the homogeneity of human capital.

4. The possibility of producing human capital in addition to earnings means that the "real output" of the individual consists of earnings plus the value of human capital produced, when the latter is evaluated by its shadow price. This will be a measure of by how much total inactivity at a given period would affect W.

5. Not much has been said of δ, the rate of deterioration. Consideration ought to be given to cases where δ is negative rather than positive as assumed here and also to the dependence of deterioration on the allocation of human capital to different activities.

6. The horizon T is treated here as exogenous. Opening the analysis to include leisure will, of course, make retirement endogenous, but even if T is the date of death there can be types of investments that will affect that date.

7. The three market prices—the rate of interest r, the rental on human capital a_0, and the price of purchased inputs P_d—affect behavior in the expected way. These are the parameters of the models that public policy can most directly attack. The framework presented here may prove convenient for mapping possible effects of public policy and their interaction with ability factors.

8. Various problems of measurement are implied but not directly discussed. One of these is the relation between the integral of investment costs, (7), over stretches of time and the integral of $\dot{K}$, or of Q. The problem of measurement in this model hinges on the measurability of s_t, which is a question about the prevalence of activities in which human capital and earnings are jointly produced. When s_t is an observable phenomenon of the allocation of time, the model becomes operational, but, without it, much is lost.

9. We have abstracted here from uncertainty and from capital rationing two important considerations in investment in human capital which should be incorporated in a complete analysis.

10. A natural extension of the two-way choice analyzed here between activities producing current earnings and investment in human capital would be to deal with the three-way choice involving also the allocation of time for activities of a consumptive nature, following the approach of Becker (1965).

The reader probably does not have to be told how simplifying many of the assumptions are and how many additional aspects of investment in human capital and of the determination of the life cycle of earnings are relevant. The main purpose of this paper was to raise some more questions rather than provide definite answers.

REFERENCES

Becker, G. S. "Investment in Human Capital: A Theoretic Analysis," *J.P.E.*, LXX, No. 5, Part 2 (October, 1962), 9–49.

———. *Human Capital*. New York: Columbia Univ. Press (for the National Bureau of Economic Research), 1964.

———. "A Theory of the Allocation of Time," *Econ. J.*, Vol. LXXV (September, 1965).

———. "Human Capital and the Personal Distribution of Income: An Analytical Approach." New York: National Bureau of Economic Research, October, 1966 (mimeographed).

Friedman, M., and Kuznets, S. *Income from Independent Professions*. New York: National Bureau of Economic Research, 1945.

Mincer, J. "Investment in Human Capital and Personal Income Distribution," *J.P.E.*, LXVI, No. 4 (August, 1958), 281–302.

———. "On-the-Job Training: Costs, Returns, and Some Implications," *ibid.*, LXX, No. 5, Part 2 (October, 1962), 50–79.

Schultz, T. W. *The Economic Value of Education*. New York: Columbia Univ. Press, 1963.