Volume Title: Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition

Volume Author/Editor: Gary S. Becker

Volume Publisher: NBER

Volume ISBNs:

Volume URL: http://www.nber.org/books/beck-5

Conference Date: n/a

Publication Date: 1964

Chapter Title: Investment in Human Capital: Rates of Return

Chapter Author(s): Gary S. Becker

Chapter URL: http://www.nber.org/chapters/c14409

Chapter pages in book: (p. 37 – 66)

words "paternalism" may simply be a way of investing in the health and welfare of employees in underdeveloped countries.

An investment in human capital would usually steepen age-earnings profiles, lowering reported earnings during the investment period and raising them later on. But an investment in an increase in earnings may have precisely the opposite effect, raising reported earnings more during the investment period than later and thus flattening age-earning profiles. The cause of this difference is simply that reported earnings during the investment period tend to be net of the cost of general investments and gross of the cost of an increase in productive earnings.[37]

The productivity of employees depends not only on their ability and the amount invested in them both on and off the job but also on their motivation, or the intensity of their work. Economists have long recognized that motivation in turn partly depends on earnings because of the effect of an increase in earnings on morale and aspirations. Equation (17), which was developed to show the effect of investments outside the firm financed by an increase in earnings, can also show the effect of an increase in the intensity of work "financed" by an increase in earnings. Thus W and MP would show initial earnings and productivity, C the increase in earnings, and G the gain to firms from the increase in productivity caused by the "morale" effect of the increase in earnings. The incentive to grant a morale-boosting increase in earnings, therefore, would depend on the same factors as does the incentive to grant an increase used for outside investments. Many recent discussions of wages in underdeveloped countries have stressed the latter,[38] while earlier discussions often stressed the former.[39]

37 If E represents reported earnings during the investment period and MP the marginal product when there is no investment, $E = MP - C$ with a general investment, $E = MP$ with a specific investment paid by the firm, and $E = MP + C$ with an increase in productive earnings.

38 See Leibenstein, *Journal of Political Economy*, April 1957, and H. Oshima, "Underdevelopment in Backward Economies: An Empirical Comment," *Journal of Political Economy*, June 1958.

39 For example, Marshall stressed the effect of an increase in earnings on the character and habits of working people (*Principles of Economics*, pp. 529-532, 566-569).

CHAPTER III

Investment in Human Capital: Rates of Return

THE most important single determinant of the amount invested in human capital may well be its profitability or rate of return, but the effect on earnings of a change in the rate of return has been difficult to distinguish empirically from a change in the amount invested. For since investment in human capital usually extends over a long and variable period, the amount invested cannot be determined from a known "investment period." Moreover, the discussion of on-the-job training clearly indicated that the amount invested is often merged with gross earnings into a single net earnings concept (which is gross earnings minus the cost of or plus the return on investment).

1. Relation Between Earnings, Costs, and Rates of Return

In this section, some important relations between earnings, investment costs, and rates of return are derived. They permit one to distinguish, among other things, a change in the return from a change in the amount invested. The discussion proceeds in stages from simple to complicated situations. First, investment is restricted to a single period and returns to all remaining periods; then investment is distributed over a known group of periods called the investment period. Finally, it is shown how the rate of return, the amount invested, and the investment period can all be derived from information on net earnings alone.

The discussion is from the viewpoint of workers and is, therefore, restricted to general investments; since the analysis of specific investments and firms is very similar, its discussion is omitted.

Let Y be an activity providing a person entering at a particular age, called age zero, with a real net earnings stream of Y_0 during the first period, Y_1 during the next period, and so on until Y_n during the last period. The general term "activity" rather than occupation or another more concrete term is used in order to indicate that any kind of investment in human capital is permitted, not just on-the-job training but also schooling, information, health, and morale. As in the previous chapter, "net" earnings mean "gross" earnings during any period minus tuition costs during the same period. "Real" earnings are the sum of monetary earnings and the monetary equivalent of psychic earnings. Since many persons appear to believe that the term "investment in human capital" must be restricted to monetary costs and returns, let me emphasize that essentially the whole analysis applies independently of the division of real earnings into monetary and psychic components. Thus the analysis applies to health, which has a large psychic component, as well as to on-the-job training, which has a large monetary component. When psychic components dominate, the language associated with consumer durable goods might be considered more appropriate than that associated with investment goods; to simplify the presentation, investment language is used throughout.

The present value of the net earnings stream in Y would be

$$V(Y) = \sum_{j=0}^{n} \frac{Y_j}{(1+i)^{j+1}},^{1} \tag{18}$$

where i is the market discount rate, assumed for simplicity to be the same in each period. If X were another activity providing a net earning stream of $X_0, X_1, \ldots X_n$, with a present value of $V(X)$, the present value of the gain from choosing Y would be given by

$$d = V(Y) - V(X) = \sum_{j=0}^{n} \frac{Y_j - X_j}{(1+i)^{j+1}}. \tag{19}$$

Equation (19) can be reformulated to bring out explicitly the relation between costs and returns. The cost of investing in human

[1] The discussion assumes discrete income flows and compounding, even though a mathematically more elegant formulation would have continuous variables, with sums replaced by integrals and discount rates by continuous compounding. The discrete approach is, however, easier to follow and yields the same kind of results. Extensions to the continuous case are straightforward.

capital equals the net earnings foregone by choosing to invest rather than choosing an activity requiring no investment. If activity Y requires an investment only in the initial period and if X does not require any, the cost of choosing Y rather than X is simply the difference between their net earnings in the initial period, and the total return would be the present value of the differences between net earnings in later periods. If $C = X_0 - Y_0$, $k_j = Y_j - X_j$, $j = 1, \ldots n$, and if R measures the total return, the gain from Y could be written as

$$d = \sum_{j=1}^{n} \frac{k_j}{(1+i)^j} - C = R - C. \tag{20}$$

The relation between costs and returns can be derived in a different and, for our purposes, preferable way by defining the internal rate of return,[2] which is simply a rate of discount equating the present value of returns to the present value of costs. In other words, the internal rate, r, is defined implicitly by the equation

$$C = \sum_{1}^{n} \frac{k_j}{(1+r)^j}, \tag{21}$$

which clearly implies

$$\sum_{j=0}^{n} \frac{Y_j}{(1+r)^{j+1}} - \sum_{0}^{n} \frac{X_j}{(1+r)^{j+1}} = d = 0, \tag{22}$$

since $C = X_0 - Y_0$ and $k_j = Y_j - X_j$. So the internal rate is also a rate of discount equating the present values of net earnings. These equations would be considerably simplified if the return were the same in each period, or $Y_j = X_j + k$, $j = 1, \ldots n$. Thus equation (21) would become

$$C = \frac{k}{r}[1 - (1+r)^{-n}], \tag{23}$$

where $(1 + r)^{-n}$ is a correction for the finiteness of life that tends toward zero as people live longer.

[2] A substantial literature has developed on the difference between the income gain and internal return approaches. See, for example, Friedrich and Vera Lutz, *The Theory of Investment of the Firm*, Princeton, 1951, Chap. ii, and the articles in *The Management of Corporate Capital*, Ezra Solomon, ed., Glencoe, 1959.

If investment is restricted to a single known period, cost and rate of return are easily determined from information on net earnings alone. Since investment in human capital is distributed over many periods—formal schooling is usually more than ten years in the United States, and long periods of on-the-job training are also common—the analysis must, however, be generalized to cover distributed investment. The definition of an internal rate in terms of the present value of net earnings in different activities obviously applies regardless of the amount and duration of investment, but the definition in terms of costs and returns is not generalized so readily. If investment were known to occur in Y during each of the first m periods, a simple and superficially appealing approach would be to define the investment cost in each of these periods as the difference between net earnings in X and Y, total investment costs as the present value of these differences, and the internal rate would equate total costs and returns. In symbols,

$$\overset{1}{C_j} = X_j - Y_j, j = 0, \ldots m - 1,$$

$$C^1 = \sum_0^{m-1} \overset{1}{C_j} (1 + r)^{-j},$$

and

$$C^1 = \frac{k}{r} \frac{(1 - (1 + r)^{m-1-n})}{(1 + r)^{m-1}}. \tag{24}$$

If $m = 1$, this reduces to equation (23).

Two serious drawbacks mar this appealing straightforward approach. The estimate of total costs requires a priori knowledge and specification of the investment period. While the period covered by formal schooling is easily determined, the period covered by much on-the-job training and other investment is not, and a serious error might result from an incorrect specification: to take an extreme example, total costs would approach zero as the investment period is assumed to be longer and longer.[3]

[3] Since

$$C^1 = \sum_0^{m-1} (X_j - Y_j)(1 + r)^{-j}, \lim_{m \to n} C^1 = \sum_0^{n-1} (X_j - Y_j)(1 + r)^{-j} = 0,$$

by definition of the internal rate.

A second difficulty is that the differences between net earnings in X and Y do not correctly measure the cost of investing in Y since they do not correctly measure earnings foregone. A person who invested in the initial period could receive more than X_1 in period 1 as long as the initial investment yielded a positive return.[4] The true cost of an investment in period 1 would be the total earnings foregone, or the difference between what could have been received and what is received. The difference between X_1 and Y_1 could greatly underestimate true costs; indeed, Y_1 might be greater than X_1 even though a large investment was made in period 1.[5] In general, therefore, the amount invested in any period would be determined not only from net earnings in the same period but also from net earnings in earlier periods.

If the cost of an investment is consistently defined as the earnings foregone, quite different estimates of total costs emerge. Although superficially a less natural and straightforward approach, the generalization from a single period to distributed investment is actually greatly simplified. Therefore, let C_j be the foregone earnings in the j^{th} period, r_j the rate of return on C_j, and let the return per period on C_j be a constant k_j, with $k = \Sigma k_j$ being the total return on the whole investment. If the number of periods were indefinitely large, and if investment occured only in the first m periods, the equation relating costs, returns, and internal rates would have the strikingly simple form of[6]

[4] If C_0 was the initial investment, r_0 its internal rate, and if the return were the same in all years, the amount

$$\overset{1}{X_1} = X_1 + \frac{r_0 C_0}{1 - (1 + r_0)^{-n}}$$

could be received in period 1.

[5] Y_1 is greater than X_1 if

$$X_1 + \frac{r_0 C_0}{1 - (1 + r_0)^{-n}} - C_1 > X_1, \text{ or if } \frac{r_0 C_0}{1 - (1 + r_0)^{-n}} > C_1,$$

where C_1 is the investment in period 1.

[6] A proof is straightforward. An investment in period j would yield a return of the amount $k_j = r_j C_j$ in each succeeding period if the number of periods were infinite and the return were the same in each. Since the total return is the sum of individual returns,

$$k = \sum_0^{m-1} k_j = \sum_0^{m-1} r_j C_j = C \sum_0^{m-1} \frac{r_j C_j}{C} = rC.$$

I am indebted to Helen Raffel for important suggestions which led to this simple proof.

$$C = \sum_{0}^{m-1} C_j = \frac{k}{\bar{r}}, \tag{25}$$

where

$$\bar{r} = \sum_{0}^{m-1} w_j r_j, \quad w_j = \frac{C_j}{C},$$

and

$$\sum_{0}^{m-1} w_j = 1. \tag{26}$$

Total cost, defined simply as the sum of costs during each period, would equal the capitalized value of returns, the rate of capitalization being a weighted average of the rates of return on the individual investments. Any sequence of internal rates or investment costs is permitted, no matter what the pattern of rises and declines, or the form of investments, be they a college education, an apprenticeship, ballet lessons, or a medical examination. Different investment programs would have the same ultimate effect on earnings whenever the average rate of return and the sum of investment costs were the same.[7]

Equation (25) could be given an interesting interpretation if all rates of return were the same. The term k/r would then be the value at the beginning of the m^{th} period of all succeeding net earning differentials between Y and X discounted at the internal rate, r.[8] Total costs would equal the value also at the beginning of the m^{th} period—which is the end of the investment period—of the first m differentials between X and Y.[9] The value of the first m differentials between X

[7] Note that the rate of return equating the present values of net earnings in X and Y is not necessarily equal to $\bar{r}$, for it would weight the rates of return on earlier investments more heavily than $\bar{r}$ does. For example, if rates were higher on investments in earlier than in later periods, the over-all rate would be greater than r, and vice versa if rates were higher in later periods. Sample calculations indicate, however, that the difference between the over-all rate and $\bar{r}$ tends to be small as long as the investment period was not very long and the systematic difference between internal rates not very great.

[8] That is,

$$\sum_{j=m}^{\infty} (Y_j - X_j)(1+r)^{m-1-j} = k \sum_{m}^{\infty} (1+r)^{m-1-j} = \frac{k}{r}.$$

[9] Since, by definition,

$$X_0 - Y_0 = C_0, \quad X_1 - Y_1 = C_1 - rC_0,$$

and Y must equal the value of all succeeding differentials between Y and X, since r would be the rate of return equating the present values in X and Y.

The internal rate of return and the amount invested in each of the first m periods could be estimated from the net earnings streams in X and Y alone if the rate of return were the same on all investments. For the internal rate r could be determined from the condition that the present value of net earnings must be the same in X and Y, and the amount invested in each period seriatim from the relations[10]

$$C_0 = X_0 - Y_0, \quad C_1 = X_1 - Y_1 + rC_0$$
$$C_j = X_j - Y_j + r \sum_{k=0}^{j-1} C_k, \; 0 \leq j \leq m-1.\text{[11]} \tag{27}$$

and more generally

$$X_j - Y_j = C_j - r \sum_{k=0}^{j-1} C_k, \quad 0 \leq j < m,$$

then

$$\sum_{j=0}^{m-1} (X_j - Y_j)(1+r)^{m-1-j} = \sum_{j=0}^{m-1} \left(C_j - r \sum_{0}^{j-1} C_k\right)(1+r)^{m-1-j}$$
$$= \sum_{0}^{m-1} C_j \left\{(1+r)^{m-1-j} - r[1 + (1+r) + \dots + (1+r)^{m-2-j}]\right\}$$
$$= \sum_{0}^{m-1} C_j = C.$$

The analytical difference between the naive definition of costs advanced earlier and one in terms of foregone earnings is that the former measures total costs by the value of earning differentials at the beginning of the investment period and the latter by the value at the end of the period. Therefore, $C^1 = C(1+r)^{1-m}$, which follows from eq. (24) when $n = \infty$.

[10] If the rate of return were not the same on all investments there would be $2m$ unknowns—$C_0, \dots C_{m-1}$, and $r_0, \dots r_{m-1}$—and only $m+1$ equations—the m cost definitions and the equation

$$k = \sum_{0}^{m-1} r_i C_i.$$

An additional $m-1$ relation would be required to determine the $2m$ unknowns. The condition $r_0 = r_1 = \dots = r_{m-1}$ is only one form these $m-1$ relations can take; another is that costs decrease at certain known rates. If the latter were assumed, all the r_i could be determined from the earnings data.

[11] In econometric terminology this set of equations forms a "causal chain" because of the natural time ordering provided by the aging process. Consequently, there is no identification or "simultaneity" problem.

Thus costs and the rate of return can be estimated from information on net earnings. This is fortunate since the return on human capital is never empirically separated from other earnings and the cost of such capital is only sometimes and incompletely separated.

The investment period of education can be measured by years of schooling, but the periods of on-the-job training, of the search for information, and of other investments are not readily available. Happily, one need not know the investment period to estimate costs and returns, since all three can be simultaneously estimated from information on net earnings. If activity X were known to have no investment (a zero investment period), the amount invested in Y during any period would be defined by

$$C_j = X_j - Y_j + r \sum_{0}^{j-1} C_k, \text{ all } j, \tag{28}$$

and total costs by

$$C = \sum_{0}^{\infty} C_j. \tag{29}$$

The internal rate could be determined in the usual way from the equality between present values in X and Y, costs in each period from equation (28), and total costs from equation (29).

The definition of costs presented here simply extends to all periods the definition advanced earlier for the investment period.[12] The

[12] Therefore, since the value of the first m earning differentials has been shown to equal

$$\sum_{0}^{m-1} C_j$$

at period m (see footnote 9), total costs could be estimated from the value of all differentials at the end of the earning period. That is,

$$C = \sum_{0}^{\infty} C_j = \sum_{0}^{\infty} (X_j - Y_j)^{\infty-1-j}.$$

Thus the value of all differentials would equal zero at the beginning of the earning period—by definition of the internal rate—and C at the end. The apparent paradox results from the infinite horizon, as can be seen from the following equation relating the value of the first f differentials at the beginning of the g^{th} period to costs:

rationale for the general definition is the same: investment occurs in Y whenever earnings there are below the sum of those in X and the income accruing on prior investments. If costs were found to be greater than zero before some period m and equal to zero thereafter, the first m periods would be the empirically derived investment period. But costs and returns can be estimated from equation (28) even when there is no simple investment period.

A common objection to an earlier draft of this paper was that the general and rather formal definition of costs advanced here is all right when applied to on-the-job training, schooling, and other recognized investments, but goes too far by also including as investment costs many effects that should be treated otherwise.

Thus, so the protest might run, learning would automatically lead to a convex and relatively steep earnings profile not because of any associated investment in education or training, but because the well-known "learning curve" is usually convex and rather steep. Since the method presented here, however, depends only on the shape of age-earnings profiles, the effect of learning would be considered an effect of investment in human capital. I accept the argument fully; indeed, I believe that it points up the power rather than the weakness of my analysis and the implied concept of human capital.

To see this requires a fuller analysis of the effect of learning. Assume that Z permits learning and that another activity X does not and has a flat earnings profile: Z might have the profile labeled TT in Chart 1 (in Chapter II) and X that labeled UU. If TT were everywhere above UU—i.e., earnings in Z were greater than those in X at each age—there would be a clear incentive for some persons to leave X and enter Z. The result would be a lowering of TT and raising of UU; generally the process would continue until TT was no longer everywhere above UU, as in Chart 1. Earnings would now be lower in Z than in X at younger ages and higher only later on, and workers

$$V(f, g) = \sum_{j=0}^{f-1} (X_j - Y_j)(1 + r)^{g-1-j} = \sum_{j=0}^{f-1} C_j(1 + r)^{g-f}.$$

When $f = \infty$ and $g = 0$, $V = 0$, but whenever $f = g$,

$$V = \sum_{0}^{f-1} C_j.$$

In particular, if $f = g = \infty$, $V = C$.

would have to decide whether the later higher earnings compensated for the lower initial earnings.

They presumably would decide by comparing the present value of earnings in Z and X, or, what is equivalent, by comparing the rate of return that equates these present values with rates that could be obtained elsewhere. They would choose Z if the present value were greater there, or if the equalizing rate were greater than those elsewhere. Therefore, they would choose Z only if the rate of return on their learning were sufficiently great, i.e., only if the returns from learning—the higher earnings later on—offset the costs of learning—the lower earnings initially. Thus choosing between activities "with a future" and "dead-end" activities involves exactly the same considerations as choosing between continuing one's education and entering the labor force—whether returns in the form of higher subsequent earnings sufficiently offset costs in the form of lower initial ones. Although learning cannot be avoided once in activities like Z, it can be avoided beforehand because workers can enter activities like X that provide little or no learning. They or society would choose learning only if it were a sufficiently good investment in the same way that they or society would choose on-the-job training if it were sufficiently profitable.

Consequently, the conclusion must be that learning is a way to invest in human capital that is formally no different from education, on-the-job training, or other recognized investments. So it is a virtue rather than a defect of our formulation of costs and returns that learning is treated symmetrically with other investments. And there is no conflict between interpretations of the shape of earning profiles based on learning theory[13] and those based on investment in human capital because the former is a special case of the latter. Of course, the fact that the physical and psychological factors associated with learning theory[14] are capable of producing rather steep concave profiles, like TT and even $T'T'$ in Chart 1, should make one hesitate in relating them to education and other conventional investments. The converse is also true, however: the fact that many investments in human capital in a market economy would produce "the learning

[13] See, for example, J. Mincer, "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, August 1958, pp. 287-288.

[14] See, e.g., R. Bush and F. Mosteller, *Stochastic Models for Learning*, New York, 1955.

curve" should make one hesitate in relating it to the various factors associated with learning theory.

Another frequent criticism is that many on-the-job investments are really free in that earnings are not reduced at any age. Although this would be formally consistent with my analysis since the rate of return need only be considered infinite (in Chart 1, TT would be nowhere below UU), I suspect that a closer examination of the alleged "facts" would usually reveal a much more conventional situation. For example, if abler employees were put through executive training programs, as is probable, they might earn no less than employees outside the programs but they might earn less than if they had not been in training.[15] Again, the earnings of employees receiving specific training may not be reduced for the reasons presented in Chapter II. Finally, one must have a very poor opinion of the ability of firms to look out for their own interests to believe that infinite rates of return are of great importance.

So much in defense of the approach. To estimate costs empirically still requires a priori knowledge that nothing is invested in activity X. Without such knowledge, only the *difference* between the amounts invested in any two activities with known net earning streams could be estimated from the definitions in equation (28). Were this done for all available streams, the investment in any activity beyond that in the activity with the smallest investment could be determined.[16] The observed minimum investment would not be zero, however, if the rate of return on some initial investment were sufficiently high to attract everyone. A relevant question is, therefore: can the shape of the stream in an activity with zero investment be specified a priori so that the total investment in any activity can be determined?

The statement "nothing is invested in an activity" only means that nothing was invested after the age when information on earnings first became available; investment can have occurred before that age. If, for example, the data begin at age eighteen, some investment in schooling, health, or information surely must have occurred at younger ages. The earning stream of persons who do not invest after age eighteen would have to be considered, at least in part, as a return on the investment before eighteen. Indeed, in the developmental ap-

[15] Some indirect evidence is cited by J. Mincer "On-the-Job Training: Costs, Returns, and Some Implications," *Investment in Human Beings*, NBER Special Conference 15, supplement to *Journal of Political Economy*, October 1962, p. 53.

[16] The technique has been applied and developed further by Mincer (*ibid.*).

proach to child-rearing, most if not all of these earnings would be so considered.

The earning stream in an activity with no investment beyond the initial age (activity X) would be flat if the developmental approach were followed and earnings were said to result entirely from earlier investment.[17] The incorporation of learning into the concept of investment in human capital also suggests that earnings profiles would be flat were there no (additional) investment. Finally, the empirical evidence, for what it is worth (see comments in Chapter VII), suggests that earnings profiles in unskilled occupations are quite flat. If the earnings profile in X were flat, the unobserved investment could easily be determined in the usual way once an assumption were made about its rate of return.

The assumption that lifetimes are infinite, although descriptively unrealistic, often yields results that are a close approximation to the truth. For example, I show later (see Chapter VI, section 2) that the average rate of return on college education in the United States would be only slightly raised if people remained in the labor force indefinitely. A finite earning period has, however, a greater effect on the rate of return of investments made at later ages, say, after forty; indeed, it helps explain why schooling and other investments are primarily made at younger ages.

An analysis of finite earning streams can be approached in two ways. One simply applies the concepts developed for infinite streams and says there is disinvestment in human capital when net earnings are above the amount that could be maintained indefinitely. Investment at younger ages would give way to disinvestment at older ages until no human capital remained at death (or retirement). This approach has several important applications and is used in parts of the study (see especially Chapter VII). An alternative that is more useful for some purposes lets the earning period itself influence the definitions of accrued income and cost. The income resulting from an investment during period j would be defined as

$$k_j = \frac{r_j C_j}{1 - (1 + r_j)^{j-n}}, \tag{30}$$

[17] If C measured the cost of investment before the initial age and r its rate of return, $k = rC$ would measure the return per period. If earnings were attributed entirely to this investment, $X_i = k = rC$, where X_i represents earnings at the i^{th} period past the initial age.

where $n + 1$ is the earning period, and the amount invested during j would be defined by

$$C_j = X_j - Y_j + \sum_{k=0}^{k=j-1} \frac{r_k C_k}{1 - (1 + r_k)^{k-n}}. \tag{31}$$

2. *The Incentive to Invest*

NUMBER OF PERIODS

Economists have long believed that the incentive to expand and improve physical resources depends on the rate of return expected. They have been very reluctant, however, to interpret improvements in the effectiveness and amount of human resources in the same way, namely, as systematic responses or "investments" resulting in good part from the returns expected. In this section and the next one, I try to show that an investment approach to human resources is a powerful and simple tool capable of explaining a wide range of phenomena, including much that has been either ignored or given *ad hoc* interpretations. The discussion covers many topics, starting with the lifespan of activities and ending with a theory of the distribution of earnings.

An increase in the lifespan of an activity would, other things the same, increase the rate of return on the investment made in any period. The influence of lifespan on the rate of return and thus on the incentive to invest is important and takes many forms, a few of which will now be discussed.

The number of periods is clearly affected by mortality and morbidity rates; the lower they are, the longer is the expected lifespan and the larger is the fraction of a lifetime that can be spent at any activity. The major secular decline of these rates in the United States and elsewhere probably increased the rates of return on investment in human capital,[18] thereby encouraging such investment.[19] This con-

[18] I say *probably* because rates of return are adversely affected (via the effects on marginal productivity) by the increase in labor force that would result from a decline in death and sickness. If the adverse effect were sufficiently great, their decline would reduce rates of return on human capital. I am indebted to my wife for emphasizing this point.

[19] The relation between investment in training and length of life is apparently even found in the training of animals, as evidenced by this statement from a book I read to my children: "Working elephants go through a long period of schooling. Training requires about ten years and costs nearly five thousand dollars. In view of the animal's long life of usefulness [they usually live more than sixty years], this is not considered too great an investment" (M. H. Wilson, *Animals of the World,* New York, 1960).

clusion is independent of whether the secular improvement in health itself resulted from investment; if so, the secular increase in rates of return would be part of the return to the investment in health.

A relatively large fraction of younger persons are in school or on-the-job training, change jobs and locations, and add to their knowledge of economic, political, and social opportunities. The main explanation may not be that the young are relatively more interested in learning, able to absorb new ideas, less tied down by family responsibilities, more easily supported by parents, or more flexible about changing their routine and place of living. One need not rely only on life-cycle effects on capabilities, responsibilities, or attitudes as soon as one recognizes that schooling, training, mobility, and the like are ways to invest in human capital and that younger people have a greater incentive to invest because they can collect the return over more years. Indeed, there would be a greater incentive even if age had no effect on capabilities, responsibilities, and attitudes.

The ability to collect returns over more years would give young persons a much greater incentive to invest even if the internal rate of return did not decline much with age. The internal rate can be seriously misleading here, as the following example indicates. If $100 invested at any age yielded $10 a year additional income forever, the rate of return would be 10 per cent at every age, and there would be no special incentive to invest at younger ages if only the rate of return were taken into account. Consider, however, a cohort of persons aged eighteen deciding when to invest. If the rate of return elsewhere were 5 per cent and if they invested immediately, the present value of the gain would be $100. If they waited five years, the present value of the gain, i.e., as of age eighteen, would only be about $78, or 22 per cent less; if they waited ten years, the present value of the gain would be under $50, or less than half. Accordingly, a considerable incentive would exist for everyone to invest immediately rather than waiting. In less extreme examples some persons might wait until older ages, but the number investing would tend to decline rapidly with age even if the rate of return did not.[20]

[20] One clear application of these considerations can be found in studies of migration, where some writers have rejected the importance of the period of returns because migration rates decline strongly with age, at least initially, while rates of return (or some equivalent) decline slowly (see the otherwise fine paper by L. Sjaastad, "The Costs and Returns of Human Migration," *Investment in Human Beings*, pp. 89-90). My analysis suggests, however, that persons with a clear gain from migration

Although the unification of these different kinds of behavior by the investment approach is important evidence in its favor, other evidence is needed. A powerful test can be developed along the following lines.[21] Suppose that investment in human capital raised earnings for p periods only, where p varied between 0 and n. The size of p would be affected by many factors, including the rate of obsolescence since the more rapidly an investment became obsolete the smaller p would be. The advantage in being young would be less the smaller p was, since the effect of age on the rate of return would be positively related to p. For example, if p equaled two years, the rate would be the same at all ages except the two nearest the "retirement" age. If the investment approach were correct, the difference between the amount invested at different ages would be positively correlated with p, which is not surprising since an expenditure with a small p would be less of an "investment" than one with a large p, and arguments based on an investment framework would be less applicable. None of the life-cycle arguments seem to imply any correlation with p, so this provides a powerful test of the importance of the investment approach.

The time spent in any one activity is determined not only by age, mortality, and morbidity but also by the amount of switching between activities. Women spend less time in the labor force than men and, therefore, have less incentive to invest in market skills; tourists spend little time in any one area and have less incentive than residents of the area to invest in knowledge of specific consumption opportunities;[22] temporary migrants to urban areas have less incentive to invest in urban skills than permanent residents; and, as a final example, draftees have less incentive than professional soldiers to invest in purely military skills.

Women, tourists, and the like have to find investments that increase productivity in several activities. A woman wants her investment to be useful both as a housewife and as a participant in the labor force,

have a strong incentive to migrate early and not wait even a few years. Since the persons remaining presumably have either no incentive or little incentive to migrate, it is not surprising that their migration rates should be much lower than that of all persons.

[21] This test was suggested by George Stigler's discussion of the effect of different autocorrelation patterns on the incentive to invest in information (see "The Economics of Information," *Journal of Political Economy*, June 1961, and "Information in the Labor Market" in *Investment in Human Beings*, pp. 94-105).

[22] This example is from Stigler, "The Economics of Information," *Journal of Political Economy*, June 1961.

or a frequent traveler wants to be knowledgeable in many environments. Such investments would be less readily available than more specialized ones—after all, an investment increasing productivity in two activities also increases it in either one alone, extreme complementarity aside, while the converse does not hold; specialists, therefore, have greater incentive to invest in themselves than others do.

Specialization in an activity would be discouraged if the market were very limited; thus the incentive to specialize and to invest in oneself would increase as the extent of the market increased. Workers would be more skilled the larger the market, not only because "practice makes perfect," which is so often stressed in discussions of the division of labor,[23] but also because a larger market would *induce* a greater investment in skills.[24] Put differently, the usual analysis of the division of labor stresses that efficiency, and thus wage rates, would be greater the larger the market, and ignores the potential earnings period in any activity, while mine stresses that this period, and thus the incentive to *become* more "efficient," would be directly related to market size. Surprisingly little attention has been paid to the latter, that is, to the influence of market size on the incentive to invest in skills.

WAGE DIFFERENTIALS AND SECULAR CHANGES

According to equation (30), the internal rate of return depends on the ratio of the return per unit of time to investment costs. A change in the return and costs by the same percentage would not change the internal rate, while a greater percentage change in the return would change the internal rate in the same direction. The return is measured by the absolute income gain, or by the absolute income difference between persons differing only in the amount of their investment. Note that absolute, not relative, income differences determine the return and the internal rate.

Occupational and educational wage differentials are sometimes

23 See, for example, A. Marshall, *Principles of Economics,* New York, 1949, Bk. IV, Chap. ix.

24 If "practice makes perfect" means that age-earnings profiles slope upward, then according to my approach it must be treated along with other kinds of learning as a way of investing in human capital. The above distinction between the effect of an increase in the market on practice and on the incentive to invest would then simply be that the incentive to invest in human capital is increased even aside from the effect of practice on earnings.

measured by relative, sometimes by absolute, wage differences,[25] although no one has adequately discussed their relative merits. Since marginal productivity analysis relates the derived demand for any class of workers to the ratio of their wages to those of other inputs,[26] wage ratios are more appropriate in understanding forces determining demand. They are not, however, the best measure of forces determining supply, for the return on investment in skills and other knowledge is determined by absolute wage differences. Therefore neither wage ratios nor wage differences are uniformly the best measure, ratios being more appropriate in demand studies and differences in supply studies.

The importance of distinguishing between wage ratios and differences, and the confusion resulting from the practice of using ratios to measure supply as well as demand forces, can be illustrated by considering the effects of technological progress. If progress were uniform in all industries and neutral with respect to all factors, and if there were constant costs, initially all wages would rise by the same proportion and the prices of all goods, including the output of industries supplying the investment in human capital,[27] would be unchanged. Since wage ratios would be unchanged, firms would have no incentive initially to alter their factor proportions. Wage differences, on the other hand, would rise at the same rate as wages, and since investment costs would be unchanged, there would be an incentive to invest more in human capital, and thus to increase the relative supply of skilled persons. The increased supply would in turn reduce the rate of increase of wage differences and produce an absolute narrowing of wage ratios.

In the United States during much of the last eighty years, a narrow-

25 See A. M. Ross and W. Goldner, "Forces Affecting the Interindustry Wage Structure," *Quarterly Journal of Economics,* May 1950; P. H. Bell, "Cyclical Variations and Trend in Occupational Wage Differentials in American Industry since 1914," *Review of Economics and Statistics,* November 1951; F. Meyers and R. L. Bowlby, "The Interindustry Wage Structure and Productivity," *Industrial and Labor Relations Review,* October 1953; G. Stigler and D. Blank, *The Demand and Supply of Scientific Personnel,* New York, NBER, 1957, Table 11; P. Keat, "Long-Run Changes in Occupational Wage Structure, 1900-1956," *Journal of Political Economy,* December 1960.

26 Thus the elasticity of substitution is usually defined as the percentage change in the ratio of quantities employed per 1 per cent change in the ratio of wages.

27 Some persons have argued that only direct investment costs would be unchanged, indirect costs or foregone earnings rising along with wages. Neutral progress implies, however, the same increase in the productivity of a student's time as in his teacher's time or in the use of raw materials, so even foregone earnings would not change.

ing of wage ratios has gone hand in hand with an increasing relative supply of skill, an association that is usually said to result from the effect of an autonomous increase in the supply of skills—brought about by the spread of free education or the rise in incomes—on the return to skill, as measured by wage ratios. An alternative interpretation suggested by the analysis here is that the spread of education and the increased investment in other kinds of human capital were in large part *induced* by technological progress (and perhaps other changes) through the effect on the rate of return, as measured by wage differences and costs. Clearly a secular decline in wage ratios would not be inconsistent with a secular increase in real wage differences if average wages were rising, and, indeed, one important body of data on wages shows a decline in ratios and an even stronger rise in differences.[28]

The interpretation based on autonomous supply shifts has been favored partly because a decline in wage ratios has erroneously been taken as evidence of a decline in the return to skill. While a decision ultimately can be based only on a detailed re-examination of the evidence,[29] the induced approach can be made more plausible by considering trends in physical capital. Economists have been aware that the rate of return on capital could be rising or at least not falling while the ratio of the "rental" price of capital to wages was falling. Consequently, although the rental price of capital declined relative to wages over time, the large secular increase in the amount of physical capital per man-hour is not usually considered autonomous, but rather induced by technological and other developments that, at least temporarily, raised the return. A common explanation based on the

28 Keat's data for 1906 and 1953 in the United States show both an average annual decline of 0.8 per cent in the coefficient of variation of wages and an average annual rise of 1.2 per cent in the real standard deviation. The decline in the coefficient of variation was shown in his study (*ibid.*); I computed the change in the real standard deviation from data made available to me by Keat.

29 For those believing that the qualitative evidence overwhelmingly indicates a continuous secular decline in rates of return on human capital, I reproduce Adam Smith's statement on earnings in some professions. "The lottery of the law, therefore, is very far from being a perfectly fair lottery; and that, as well as many other liberal and honourable professions, is, in point of pecuniary gain, evidently under-recompensed" (*The Wealth of Nations*, Modern Library edition, New York, 1937, p. 106). Since economists tend to believe that law and most other liberal professions are now overcompensated relative to nonprofessional work "in point of pecuniary gain," the return to professional work could not have declined continuously if Smith's observations were accurate.

effects of economic progress may, then, account for the increase in both human and physical capital.[30]

RISK AND LIQUIDITY

An informed, rational person would invest only if the expected rate of return were greater than the sum of the interest rate on riskless assets and the liquidity and risk premiums associated with the investment. Not much need be said about the "pure" interest rate, but a few words are in order on risk and liquidity. Since human capital is a very illiquid asset—it cannot be sold and is rather poor collateral on loans—a positive liquidity premium, perhaps a sizable one, would be associated with such capital.

The actual return on human capital varies around the expected return because of uncertainty about several factors. There has always been considerable uncertainty about the length of life, one important determinant of the return. People are also uncertain about their ability, especially younger persons who do most of the investing. In addition, there is uncertainty about the return to a person of given age and ability because of numerous events that are not predictable. The long time required to collect the return on an investment in human capital reduces the knowledge available, for knowledge required is about the environment when the return is to be received, and the longer the average period between investment and return, the less such knowledge is available.

Informed observation as well as calculations I have made suggest that there is much uncertainty about the return to human capital.[31] The response to uncertainty is determined by its amount and nature and by tastes or attitudes. Many have argued that attitudes of investors in human capital are very different from those of investors in

30 Some quantitative evidence for the United States is discussed in Chapter VI, section 2.

31 For example, Marshall said: "Not much less than a generation elapses between the choice by parents of a skilled trade for one of their children, and his reaping the full results of their choice. And meanwhile the character of the trade may have been almost revolutionized by changes, on which some probably threw long shadows before them, but others were such as could not have been foreseen even by the shrewdest persons and those best acquainted with the circumstances of the trade" and "the circumstances by which the earnings are determined are less capable of being foreseen [than those for machinery]" (*Principles of Economics*, p. 571). In section 4 of Chapter IV some quantitative estimates of the uncertainty in the return to education are presented.

physical capital because the former tend to be younger,[32] and young persons are supposed to be especially prone to overestimate their ability and chance of good fortune.[33] Were this view correct, a human investment which promised a large return to exceptionally able or lucky persons would be more attractive than a similar physical investment. However, a "life-cycle" explanation of attitudes toward risk may be no more valid or necessary than life-cycle explanations of why investors in human capital are relatively young (discussed above). Indeed, an alternative explanation of reactions to large gains has already appeared.[34]

CAPITAL MARKETS AND KNOWLEDGE

If investment decisions responded only to earning prospects, adjusted for risk and liquidity, the adjusted marginal rate of return would be the same on all investments. The rate of return on education, training, migration, health, and other human capital is supposed to be higher than on nonhuman capital, however, because of financing difficulties and inadequate knowledge of opportunities. These will now be discussed briefly.

Economists have long emphasized that it is difficult to borrow funds to invest in human capital because such capital cannot be offered as collateral, and courts have frowned on contracts which even indirectly suggest involuntary servitude. This argument has been explicitly used to explain the "apparent" underinvestment in education and training and also, although somewhat less explicitly, underinvestment in health, migration, and other human capital. The importance attached to capital market difficulties can be determined not only from the discussions of investment but also from the discussions of consumption. Young persons would consume relatively little, productivity and wages might be related, and some other consumption patterns would follow only if it were difficult to capitalize future earning power. In-

32 Note that our argument above implied that investors in human capital would be younger.

33 Smith said: "The contempt of risk and the presumptuous hope of success, are in no period of life more active than at the age at which young people choose their professions" (*Wealth of Nations,* p. 109). Marshall said that "young men of an adventurous disposition are more attracted by the prospects of a great success than they are deterred by the fear of failure" (*Principles of Economics,* p. 554).

34 See M. Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risks," reprinted in *Readings in Price Theory,* G. J. Stigler and K. Boulding, eds., Chicago, 1952.

deed, unless capital limitations applied to consumption as well as investment, the latter could be indirectly financed with "consumption" loans.[35]

Some other implications of capital market difficulties can also be mentioned:

1. Since large expenditures would be more difficult to finance, investment in, say, a college education would be more affected than in, say, short-term migration.
2. Internal financing would be common, and consequently wealthier families would tend to invest more than poorer ones.
3. Since employees' specific skills are part of the intangible assets or good will of firms and can be offered as collateral along with tangible assets, capital would be more readily available for specific than for general investments.
4. Some persons have argued that opportunity costs (foregone earnings) are more readily financed than direct costs because they require only to do "without," while the latter require outlays. Although superficially plausible, this view can easily be shown to be wrong: opportunity and direct costs can be financed equally readily, given the state of the capital market. If total investment costs were $800, potential earnings $1,000, and if all costs were foregone earnings, investors would have $200 of earnings to spend; if all were direct costs, they would initially have $1,000 to spend, but just $200 would remain after paying "tuition," so their *net* position would be exactly the same as before. The example can be readily generalized and the obvious inference is that indirect and direct investment costs are equivalent in imperfect as well as perfect capital markets.

While it is undeniably difficult to use the capital market to finance investments in human capital, there is some reason to doubt whether otherwise equivalent investments in physical capital can be financed much more easily. Consider an eighteen-year-old who wants to invest a given amount in equipment for a firm he is starting rather than in a college education. What is his chance of borrowing the whole amount at a "moderate" interest rate? Very slight, I believe, since he would be untried and have a high debt-equity ratio; moreover, the

35 A person with an income of X and investment costs of Y ($Y < X$) could either use X for consumption and receive an *investment loan* of Y, or use $X - Y$ for consumption, Y for investment, and receive a *consumption loan* of Y. He ends up with the same consumption and investment in both cases, the only difference being in the names attached to the loans.

collateral provided by his equipment would probably be very imperfect. He, too, would either have to borrow at high interest rates or self-finance. Although the difficulties of financing investments in human capital have usually been related to special properties of human capital, in large measure they also seem to beset comparable investments in physical capital.

A recurring theme is that young persons are especially prone to be ignorant of their abilities and of the investment opportunities available. If so, investors in human capital, being younger, would be less aware of opportunities and thus more likely to err than investors in tangible capital. I suggested earlier that investors in human capital are younger partly because of the cost in postponing their investment to older ages. The desire to acquire additional knowledge about the return and about alternatives provides an incentive to postpone any risky investment, but since an investment in human capital is more costly to postpone, it would be made earlier and presumably with less knowledge than comparable nonhuman investments. Therefore, investors in human capital may not have less knowledge *because* of their age; rather both might be a *joint* product of the incentive not to delay investing.

The eighteen-year-old in our example who could not finance a purchase of machinery might, without too much cost, postpone the investment for a number of years until his reputation and equity were sufficient to provide the "personal" collateral required to borrow funds. Financing may prove a more formidable obstacle to investors in human capital because they cannot postpone their investment so readily. Perhaps this accounts for the tendency of economists to stress capital market imperfections when discussing investments in human capital.

3. *Some Effects of Human Capital*

EXAMPLES

Differences in earnings among persons, areas, or time periods are usually said to result from differences in physical capital, technological knowledge, ability, or institutions (such as unionization or socialized production). The previous discussion indicates, however, that investment in human capital also has an important effect on observed earnings because earnings tend to be net of investment costs and gross of investment returns. Indeed, an appreciation of the direct and in-

direct importance of human capital appears to resolve many otherwise puzzling empirical findings about earnings. Consider the following examples:

1. Almost all studies show that age-earnings profiles tend to be steeper among more skilled and educated persons. I argued earlier (Chapter II, section 1) that on-the-job training would steepen age-earnings profiles, and the analysis of section 1 of this chapter generalizes the argument to all human capital. For since observed earnings are gross of returns and net of costs, investment in human capital at younger ages would reduce observed earnings then and raise them at older ages, thus steepening the age-earnings profile.[36] Likewise, investment in human capital would make the profile more concave.[37]

2. In recent years students of international trade theory have been somewhat shaken by findings that the United States, said to have a relative scarcity of labor and an abundance of capital, apparently exports relatively labor-intensive commodities and imports relatively capital-intensive commodities. For example, one study found that

[36] According to eq. (28) earnings at age j can be approximated by

$$Y_j = X_j + \sum_{k=0}^{k=j-1} r_k C_k - C_j,$$

where X_j are earnings at j of persons who have not invested in themselves, C_k is the investment at age k, and r_k is its rate of return. The rate of increase in earnings would be at least as steep in Y as in X at each age and not only from "younger" to "older" ages if and only if

$$\frac{\Delta Y_j}{\Delta j} \geq \frac{\Delta X_j}{\Delta j},$$

or

$$r_j C_j \geq \frac{\Delta C_j}{\Delta j}.$$

This condition is usually satisfied since $r_j C_j \geq 0$ and the amount invested tends to decline with age.

[37] Following the notation of the previous footnote, Y would be more concave than X if and only if

$$\Delta\left(\frac{\Delta Y_j}{\Delta j}\right) - \Delta\left(\frac{\Delta X_j}{\Delta j}\right) = \Delta\left(\frac{r_j C_j}{\Delta j}\right) - \Delta\left(\frac{\Delta C_j}{\Delta j}\right) < 0.$$

The first term on the right is certain to be negative, at least eventually, because both r_j and C_j would eventually decline, while the second term would be positive because C_j would eventually decline at a decreasing rate. Consequently, the inequality would tend to hold and the earnings profile in Y would be more concave than that in X.

export industries pay higher wages than import-competing ones.[38]

An interpretation consistent with the Ohlin-Heckscher emphasis on the relative abundance of different factors argues that the United States has an even more (relatively) abundant supply of human than of physical capital. An increase in human capital would, however, show up as an apparent increase in labor intensity since earnings are gross of the return on such capital. Thus export industries might pay higher wages than import-competing ones primarily because they employ more skilled or healthier workers.[39]

3. Several recent studies have tried to estimate empirically the elasticity of substitution between capital and labor. Usually a ratio of the input of physical capital (or output) to the input of labor is regressed on the wage rate in different areas or time periods, the regression coefficient being an estimate of the elasticity of substitution.[40] Countries, states, or time periods that have relatively high wages and inputs of physical capital also tend to have much human capital. Just as a correlation between wages, physical capital, and human capital seems to obscure the relationship between relative factor supplies and commodity prices, so it obscures the relationship between relative factor supplies and factor prices. For if wages were high primarily because of human capital, a regression of the relative amount of physical capital on wages could give a seriously biased picture of the effect on wages of factor proportions.[41]

38 See I. Kravis, "Wages and Foreign Trade," *Review of Economics and Statistics*, February 1956.

39 This kind of interpretation has been put forward by many writers; see, for example, the discussion in W. Leontief, "Factor Proportions and the Structure of American Trade: Further Theoretical and Empirical Analysis," *Review of Economics and Statistics*, November 1956.

40 Interstate estimates for several industries can be found in J. Minasian, "Elasticities of Substitution and Constant-Output Demand Curves for Labor," *Journal of Political Economy*, June 1961, pp. 261-270; intercountry estimates in Kenneth Arrow, Hollis B. Chenery, Bagicha Minhas, and Robert M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Review of Economics and Statistics*, August 1961; as yet unpublished studies by Philip Nelson and Robert Solow contain both interstate and time-series estimates.

41 Minasian's argument (in his article cited above, p. 264) that interstate variations in skill level necessarily bias his estimates toward unity is actually correct only if skill is a perfect substitute for "labor." (In correspondence Minasian stated that he intended to make this condition explicit.) If, on the other hand, human and physical capital were perfect substitutes, I have shown (in an unpublished memorandum) that the estimates would always have a downward bias, regardless of the true substitution between labor and capital. Perhaps the most reasonable assumption would be that physical capital is more complementary with human capital than with labor; I have not, however, been able generally to determine the direction of bias in this case.

4. A secular increase in average earnings has usually been said to result from increases in technological knowledge and physical capital per earner. The average earner, in effect, is supposed to benefit indirectly from activities by entrepreneurs, investors, and others. Another explanation put forward in recent years argues that earnings can rise because of direct investment in earners.[42] Instead of only benefiting from activities by others, the average earner is made a prime mover of development through the investment in himself.[43]

ABILITY AND THE DISTRIBUTION OF EARNINGS

An emphasis on human capital not only helps explain differences in earnings over time and among areas but also among persons or families within an area. This application will be discussed in greater detail than the others because a link is provided between earnings, ability, and the incentive to invest in human capital.

Economists have long been aware that conventional measures of ability—intelligence tests or aptitude scores, school grades, and personality tests—while undoubtedly relevant at times, do not reliably measure the talents required to succeed in the economic sphere. The latter consists of particular kinds of personality, persistence, and intelligence. Accordingly, some writers have gone to the opposite extreme and argued that the only relevant way to measure economic talent is by results, or by earnings themselves.[44] Persons with higher earnings would simply have more ability than others, and a skewed distribution of earnings would imply a skewed distribution of abilities. This approach goes too far, however, in the opposite direction. The main reason for relating ability to earning is to distinguish its effects from

D. O'Neill is currently finishing a dissertation at Columbia University in which estimates of human capital are explicitly incorporated into the cross-sectional regressions.

42 The major figure here is T. W. Schultz. Of his many articles, see especially "Education and Economic Growth" in *Social Forces Influencing American Education*, Sixtieth Yearbook of the National Society for the Study of Education, Chicago, 1961, Part II, Chap. 3.

43 One caveat is called for, however. Since observed earnings are not only gross of the return from investments in human capital but also are net of some costs, an increased investment in human capital would both raise and reduce earnings. Although average earnings would tend to increase as long as the rate of return was positive, the increase would be less than if the cost of human capital, like that of physical capital, was not deducted from national income.

44 Let me state again that the word "earnings" stands for real earnings, or the sum of monetary earnings and the monetary equivalent of psychic earnings.

differences in education, training, health, and other such factors, and a definition equating ability and earnings *ipso facto* precludes such a distinction. Nevertheless, results are very relevant and should not be ignored.

A compromise might be reached through defining ability by earnings only when several variables have been held constant. Since the public is very concerned about separating ability from education, on-the-job training, health, and other human capital, the amount invested in such capital would have to be held constant. Although a full analysis would also hold discrimination, nepotism, luck, and several other factors constant, a reasonable first approximation would say that if two persons have the same investment in human capital, the one who earns more is demonstrating greater economic talent.

Since observed earnings are gross of the return on human capital, they are affected by changes in the amount and rate of return. Indeed, it has been shown that, after the investment period, earnings (Y) can be simply approximated by

$$Y = X + rC, \tag{32}$$

where C measures total investment costs, r the average rate of return, and X earnings when there is no investment in human capital. If the distribution of X is ignored for now, Y would depend only on r when C was held constant, so "ability" would be measured by the average rate of return on human capital.[45]

In most capital markets the amount invested is not the same for everyone nor rigidly fixed for any given person, but depends in part on the rate of return. Persons receiving a high marginal rate of return would have an incentive to invest more than others.[46] Since marginal and average rates are presumably positively correlated[47] and since

[45] Since r is a function of C, Y would indirectly as well as directly depend on C, and therefore the distribution of ability would depend on the amount of human capital. Some persons might rank high in earnings and thus high in ability if everyone were unskilled, and quite low if education and other training were widespread.

[46] In addition, they would find it easier to invest if the marginal return and the resources of parents and other relatives were positively correlated.

[47] According to a well-known formula,

$$r_m = r_a\left(1 + \frac{1}{e_a}\right),$$

where r_m is the marginal rate of return, r_a the average rate, and e_a the elasticity of the average rate with respect to the amount invested. The rates r_m and r_a would be positively correlated unless r_a and $1/e_a$ were sufficiently negatively correlated.

ability is measured by the average rate, one can say that abler persons would invest more than others. The end result would be a positive correlation between ability and the investment in human capital,[48] a correlation with several important implications.

One is that the tendency for abler persons to migrate, continue their education,[49] and generally invest more in themselves can be explained without recourse to an assumption that noneconomic forces or demand conditions favor them at higher investment levels. A second implication is that the separation of "nature from nurture" or ability from education and other environmental factors is apt to be difficult, for high earnings would tend to signify both more ability and a better environment. Thus the earnings differential between college and high-school graduates does not measure the effect of college alone since college graduates are abler and would earn more even without the additional education. Or reliable estimates of the income elasticity of demand for children have been difficult to obtain because higher-income families also invest more in contraceptive knowledge.[50]

The main implication, however, is in personal income distribution. At least ever since the time of Pigou economists have tried to reconcile the strong skewness in the distribution of earnings and other income with a presumed symmetrical distribution of abilities.[51] Pigou's main suggestion—that property income is not symmetrically distributed—does not directly help explain the skewness in earnings. Subsequent attempts have largely concentrated on developing *ad hoc* random and other probabilistic mechanisms that have little relation to the mainstream of economic thought.[52] The approach presented here, however,

[48] This kind of argument is not new; Marshall argued that business ability and the ownership of physical capital would be positively correlated: "[economic] forces . . . bring about the result that there is a far more close correspondence between the ability of business men and the size of the businesses which they own than at first sight would appear probable" (*Principles of Economics*, p. 312).

[49] The first is frequently alleged (see, for example, *ibid.*, p. 199). Evidence on the second is discussed in Chapter IV, section 2.

[50] See my "An Economic Analysis of Fertility" in *Demographic and Economic Change in Developed Countries*, Special Conference 11, Princeton for NBER, 1960.

[51] See A. C. Pigou, *The Economics of Welfare*, 4th ed., London, 1950, Part IV, Chap. ii.

[52] A sophisticated example can be found in B. Mandelbrot, "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review*, May 1960. In a recent paper, however, Mandelbrot has brought in maximizing behavior (see "Paretian Distributions and Income Maximization," *Quarterly Journal of Economics*, February 1962).

offers an explanation that is not only consistent with economic analysis but actually relies on one of its fundamental tenets, namely, that the amount invested is a function of the rate of return expected. In conjunction with the effect of human capital on earnings, this tenet can explain several well-known properties of earnings distributions.

By definition, the distribution of earnings would be exactly the same as the distribution of ability if everyone invested the same amount in human capital; in particular, if ability were symmetrically distributed, earnings would also be. Equation (32) shows that the distribution of earnings would be exactly the same as the distribution of investment if all persons were equally able; again, if investment were symmetrically distributed, earnings would also be.[53] If ability and investment both varied, earnings would tend to be skewed even when ability and investment were not, but the skewness would be small as long as the amount invested were statistically independent of ability.[54]

It has been shown, however, that abler persons would tend to invest

[53] Jacob Mincer ("Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy,* August 1958) concluded that a symmetrical distribution of investment in education implies a skewed distribution of earnings because he defines educational investment by school years rather than costs. If Mincer is followed in assuming that everyone was equally able, that schooling was the only investment, and that the cost of the n^{th} year of schooling equaled the earnings of persons with $n - 1$ years of schooling, then, say, a normal distribution of schooling can be shown to imply a long-normal distribution of school costs and thus a log-normal distribution of earnings.

The difference between the earnings of persons with $n - 1$ and n years of schooling would be $k_n = Y_n - Y_{n-1} = r_n C_n$. Since r_n is assumed to equal r for all n, and $C_n = Y_{n-1}$, this equation becomes $Y_n = (1 + r)\, Y_{n-1}$, and therefore

$$\begin{aligned} C_1 &= Y_0 \\ C_2 &= Y_1 = Y_0\,(1+r) \\ C_3 &= Y_2 = Y_1\,(1+r) = Y_0\,(1+r)^2 \\ C_n &= Y_{n-1} = \ldots = Y_0\,(1+r)^{n-1}\quad , \end{aligned}$$

or the cost of each additional year of schooling increases at a constant *rate*. Since total costs have the same distribution as $(1 + r)^n$, a symmetrical, say, a normal, distribution of school years, n, implies a log-normal distribution of costs and hence by eq. (32) a log-normal distribution of earnings. I am indebted to Mincer for a helpful discussion of the comparison and especially for the stimulation provided by his pioneering work. Incidentally, his article and the dissertation on which it is based cover a much broader area than has been indicated here.

[54] For example, C. C. Craig has shown that the product of two independent normal distributions is only slightly skewed (see his "On the Frequency Function of XY," *Annals of Mathematical Statistics,* March 1936, p. 3).

more than others, so ability and investment would be positively correlated, perhaps quite strongly. Now the product of two symmetrical distributions is more positively skewed the higher the positive correlation between them, and might be quite skewed.[55] The economic incentive given abler persons to invest relatively large amounts in themselves does seem capable, therefore, of reconciling a strong positive skewness in earnings with a presumed symmetrical distribution of abilities.

Variations in X help explain an important difference among skill categories in the degree of skewness. The smaller the fraction of total earnings resulting from investment in human capital—the smaller rC relative to X—the more the distribution of earnings would be dominated by the distribution of X. Higher-skill categories have a greater average investment in human capital and thus presumably a larger rC relative to X. The distribution of "unskilled ability," X, would, therefore, tend to dominate the distribution of earnings in relatively unskilled categories while the distribution of a product of ability and the amount invested, rC, would dominate in skilled categories. Hence if abilities were symmetrically distributed, earnings would tend to be more symmetrically distributed among the unskilled than among the skilled.[56]

Equation (32) holds only when investment costs are small, which tends to be true at later ages, say, after age 35. Net earnings at earlier ages would be given by

$$Y_j = X_j + \sum_{0}^{j-1} r_i C_i + (-C_j), \tag{33}$$

where j refers to the current year and i to previous years, C_i measures the investment cost of age i, C_j current costs, and r_i the rate of return on C_j. The distribution of $-C_j$ would be an important determinant

[55] Craig (*ibid.*, pp. 9-10) showed that the product of two normal distributions would be more positively skewed the higher the positive correlation between them, and that the skewness would be considerable with high correlations.

[56] As noted earlier, X does not really represent earnings when there is no investment in human capital, but only earnings when there is no investment after the initial age (be it 14, 25, or 6). Indeed, the developmental approach to child-rearing argues that earnings would be close to zero if there were no investment at all in human capital. The distribution of X, therefore, would be at least partly determined by the distribution of investment before the initial age, and if it and ability were positively correlated, X might be positively skewed, even though ability was not.

of the distribution of Y_j since investment is large at these ages. Hence the analysis would predict a smaller (positive) skewness at younger than at older ages partly because X would be more important relative to $\Sigma r_i C_i$ at younger ages and partly because the presumed negative correlation between $-C_j$ and $\sum_{0}^{j-1} r_i C_i$ would counteract the positive correlation between r_i and C_i.

A simple analysis of the incentive to invest in human capital seems capable of explaining, therefore, not only why the over-all distribution of earnings is more skewed than the distribution of abilities, but also why earnings are more skewed among older and skilled persons than among younger and less skilled ones. The renewed interest in investment in human capital may provide the means of bringing the theory of personal income distribution back into economics.

PART TWO
EMPIRICAL ANALYSIS

"An investment in knowledge
pays the best interest."

Benjamin Franklin, *Poor Richard's Almanack*

C /