more moderate differences, similar regressions were calculated for all fifty states. To avoid going into details at this time let us simply report that the results strongly confirm those found at the extremes: there is a very sizable positive correlation across states between inequality in adult male incomes, adjusted rates of return, inequality in schooling, coefficients of determination, and residual inequality in incomes, while they are all negatively related to the average level of schooling and income. Whereas only about 18 per cent of the inequality in income within a state is explained, on the average, by schooling, the remaining 82 per cent explained by the residual, about one-third of the differences in inequality between states is directly explained by schooling, one-third directly by the residual, and the remaining one-third by both together through the positive correlation between them.

Similar calculations have also been made for several countries having readily available data: United States, Canada, Mexico, Israel, and Puerto Rico (treated as a country). Again there is a strong tendency for areas with greater income inequality to have higher rates of return, greater schooling inequality, higher coefficients of determination, and greater residual inequality. While there is also a tendency for poorer countries to have lower average years of schooling, greater inequality in income, etc., there are a couple of notable exceptions. For example, Israel, for reasons rather clearly related to the immigration of educated Europeans during the 1920s and 1930s, had unusually high schooling levels and low inequality in earnings until the immigration of uneducated Africans and Asians after 1948 began to lower average education levels and raise the inequality in earnings.

Addendum: Human Capital and the Personal Distribution of Income: An Analytical Approach*

1. Introduction

Interest among economists in the distribution of income has as long a history as modern economics itself. Smith, Mill, Ricardo, and others recognized that many problems of considerable economic importance partly turned on various aspects of income distribution. Although they defined poverty, for example, in absolute terms, they also recognized that each generation's "poor" are mainly those significantly below the average income level. In addition to poverty, the degree of opportu-

* Originally published as Woytinsky Lecture, University of Michigan, 1967.
HUMAN CAPITAL AND THE PERSONAL DISTRIBUTION OF INCOME

nity, aggregate savings, and investment, the distribution of family sizes and the concentration of private economic power were believed to be related to income distribution.

How does one explain then that in spite of the rapid accumulation of empirical information and the persisting and even increasing interest in some of these questions, such as poverty, economists have somewhat neglected the study of personal income distribution during the past generation? In my judgment the fundamental reason is the absence, despite ingenious and valiant efforts, of a theory that both articulates with general economic theory and is useful in explaining actual differences among regions, countries, and time periods. By emphasizing investment in human capital one can develop a theory of income distribution that satisfies both desiderata. This essay focuses on the relation between investment in human capital and the distribution of earnings and other income. The discussion is theoretical and no systematic effort is made to test the theory empirically. I expect to report on some quantitative tests in a future publication.

The next section sets out the basic theory determining the amount invested in human capital by a "representative" person, and shows the relation between earnings, investments, and rates of return. Essentially all that is involved is the application to human capital of a framework traditionally used to analyze investment in other capital, although several modifications are introduced. Section 3 of the essay shifts the attention from a single person to differences among persons, and shows how the distribution of earnings and investments are determined by the distributions of ability, tastes, subsidies, wealth, and other variables.

Section 4 uses the framework developed in sections 2 and 3 to analyze the effects on the distribution of earnings of an increase in the equality of opportunity, of a more efficient market for human capital, of the use of tests and other "objective" criteria to ration investments in human capital, and of legislation requiring a minimum investment in human capital. Section 5 extends the discussion to the distribution of property income, and suggests why such income, both inherited and self-accumulated, is more unequally distributed than earnings. Section 6 summarizes the discussion and adds a few conclusions, and 7 is a mathematical appendix.

2. Optimal Investment in Human Capital

a. The model. I have shown elsewhere that what I call the "net" earnings of a person at any age \( t \) (\( E_t \)) approximately equals the earn-
Investment in Human Capital: Rates of Return

Earnings he would have at time $t$ if no human capital had been invested in him ($X_t$) plus the total returns to him at time $t$ on investments made in him earlier ($k_t$) minus the cost to him of investments at time $t$ ($C_t$), as in

$$E_t = X_t + k_t - C_t. \quad (78)$$

Total returns depend on the amounts invested and their rates of return; for example, if returns on each investment were the same at all ages during the labor force period, total returns would be the sum of the products of the amounts invested and their rates of return, adjusted for the finiteness of the labor force period. Equation (78) could then be written as

$$E_t = X_t + \sum_{j=1}^{n} r_{t-j} f_{t-j} C_{t-j} - C_t. \quad (79)$$

where $r_{t-j}$ is the rate of return on capital invested at time $t-j$ and $f_{t-j}$ is the finite life adjustment. I applied this analysis to various problems, including the shapes of age-earnings and age-wealth profiles, the relation between unemployment and on-the-job training, the so-called Leontief paradox, and several others.

I suggested that differences in the total amounts invested by different persons are related to differences in the rates of return obtainable, a suggestion that can explain why white urban males with high IQs acquire more education than others, or why the division of labor is limited by the extent of the market. I did not, however, systematically develop a framework to explain why rates of return and investments differ so greatly among persons. This essay tries to develop such a framework. This not only provides a rigorous justification for these suggestions in Human Capital, but also begins to provide an explanation of the personal distribution of earnings.

The term $X_t$ in equations (78) and (79) represents the earnings of a person that are unrelated to human capital invested in him, and are presumably, therefore, largely independent of his current choices. Particularly in developed economies but perhaps in most, there is sufficient investment in education, training, informal learning, health, and just plain child rearing that the earnings unrelated to investment in human capital are a small part of the total. Indeed, in the develop-

76 This is the “one-hoss shay” assumption applied to human capital.
77 See this volume, Chapters II–III.
mental approaches to child rearing, all the earnings of a person are ultimately attributed to different kinds of investments made in him. Consequently, there is considerable justification for the assumption that $X$, is small and can be neglected, an assumption we make in this paper. In any case a significant $X$, only slightly complicates the analysis and can be readily incorporated.

Another assumption made throughout most of the paper is that human capital is homogeneous in the sense that all units are perfect substitutes in production for each other and thus add the same amount to earnings. Of course, this assumption does not deny that some units may have been produced at considerably greater costs than others. The assumption of homogeneous human capital clearly differs in detail rather drastically from the usual emphasis on qualitative differences in education, training, and skills. I hope to demonstrate that these differences, while descriptively realistic and useful, are not required to understand the basic forces determining the distribution of earnings; indeed, they sometimes even distract attention from these determinants. Section 5g does, however, generalize the analysis to cover many kinds of human capital.

Chart 4 plots along the horizontal axis the amount invested in human capital measured for convenience by its cost rather than in physical units. Equal distances along the axis, therefore, do not necessarily measure equal numbers of physical units.

The curve $D$ shows the marginal benefit, for simplicity measured by the rate of return, to a particular person on each additional dollar of investment, and is supposed to represent his demand curve for human capital. The curve $S$ shows the effective marginal financing cost to him, measured for simplicity by the rate of interest, of each additional dollar invested, and represents in essence his supply curve of capital. If $D$ exceeded $S$, the marginal rate of return would exceed the marginal rate of interest, and income would be increased by additional investment, while the opposite would be true if $S$ exceeded $D$. Consequently, income is maximized by investing up to the point where $D = S$, given by $p$ in the figure, and implying a total capital investment of $OC_p$.

**b. Demand curves.** The marginal rate of return depends on the time series of marginal returns and the marginal production cost of investment: if returns are constant for a long labor force period, it essentially equals the ratio of returns to these costs. Since all human

---

capital is assumed to be homogeneous, even an extremely large percentage change in the capital invested by any one person would have a negligible effect on the total quantity of capital available. Consequently, in order to explain why the demand curves for human capital in Chart 4 are negatively inclined and not horizontal, other effects of capital accumulation must be analyzed.

The principal characteristic that distinguishes human from other kinds of capital is that, by definition, the former is embedded or embodied in the person investing. This embodiment of human capital is the most important reason why marginal benefits decline as additional capital is accumulated. One obvious implication of embodiment is that since the memory capacity, physical size, etc. of each investor is limited, eventually diminishing returns set in from produc-
ing additional capital. The result is increasing marginal costs of producing a dollar of returns.

Closely dependent on the embodiment of human capital is the importance of an investor's own time in the production of his own human capital. Own time is so important that an increase in the amount invested in good part corresponds to an increase in the time spent investing; in fact the only commonly used measures of schooling and training are years of schooling and training, measures entirely based on the input of own time. The cost of this time has been measured for several kinds of human capital, shown to be generally important, and given the name "foregone earnings." If the elasticities of substitution between own time and teachers, books, and other inputs were infinite, the use of own time and the deferral of investments could be avoided, without cost, aside from the limitations imposed by B, by an accumulation of all the desired capital instantaneously through complete substitution of other inputs for own time. If substitution were significantly imperfect (which is the more likely situation), the elimination of own time would cause the marginal costs of producing human capital to be higher and rise faster as capital was accumulated than if it was combined optimally with other inputs. In the latter case, however, the accumulation of capital is necessarily spread out over a period of calendar time called the

If

\[ h = f(I, B), \]

where \( h \) is the number of units of capital produced by a person per unit time, \( f \) is his production function, \( I \) is his capital investment in dollars per unit time, and \( B \) represents his physical and mental powers, then eventually

\[ \frac{\partial h}{\partial I} < 0. \]

The production function in the previous footnote can be expanded to

\[ h = f(R, T, B), \]

where \( R \) is the rate of input of other resources, and \( T \) is the rate of input of the investor's time per unit calendar time.

If the horizontal axis in Chart 4 were replaced by one measuring investment time, the chart would be almost identical to those used in the "Austrian" theory of capital to explain optimal aging of trees or wine. Indeed, the main relevance of the Austrian approach in modern economics is to the study of investment in human capital.

INVESTMENT IN HUMAN CAPITAL: RATES OF RETURN

"investment period." Presumably there are optimal combinations of inputs over an optimal investment period that maximizes the present value of benefits from a given capital investment. The spreading out of capital accumulation forced by the importance of own time can, however, only reduce but not eliminate the decline in marginal benefits as more is accumulated.84

In the first place, with finite lifetimes, later investments cannot produce returns for as long as earlier ones and, therefore, usually have smaller total benefits. This effect is important in societies with heavy adult mortality, but probably is not in the low mortality environment of modern Western societies. For unless fewer than approximately twenty years of working life remained, a reduction of, say, a year in the number of years remaining does not have much effect on the present value of benefits.85 In the second place, later investments are less profitable than earlier ones because the present value of net benefits (or profits) is reduced merely by postponing them (and the reduction can be sizable, even for postponements of a few years).86

A third consideration is probably of great importance, although one cannot yet measure its quantitative significance. Since nobody can use his time at any activity without taking with him all of his human capital, the latter enters as an input along with his time in the production of additional capital. Initially, at young ages, the value of the time is small and probably even negative because parents or other baby-sitting services must be employed if he is not in school, or otherwise investing.87 As he continues to invest, however, the capital accumulated becomes increasingly valuable, and so does his time.

Other things being the same, an increase in the value of time raises the marginal cost of later investments compared to earlier ones since the former use more expensive time. For any given rate of increase in its value as he ages, the costs of later investments are relatively greater, the larger the share of foregone earnings in costs and the smaller the

84 The fact that a person's optimal stock of human capital is not immediately reached is often used in explaining the shape of his demand curve for human capital. On the problems in explaining why his optimal stock of nonhuman capital is not immediately reached, see D. W. Jorgenson, "The Theory of Investment Behavior," in Determinants of Investment Behavior, Universities-National Bureau Conference Series No. 18, Columbia University Press for NBER, 1967.
85 For a demonstration of this, see pp. 47–48.
86 See ibid., pp. 72–73.
elasticity of substitution between own time and other inputs. One other thing that may not remain the same is the productivity of time: just as a greater amount of human capital is more productive than a lesser amount of capital in the rest of the economy, so too it may be more productive when used to produce additional human capital itself. Marginal costs of later investments would not be greater if the increased productivity of own time was at least as great as its increased value. Because own human capital is carried along with own time, more productive or not, I am inclined to believe that its effect on productivity would be less, at least eventually, than its effect on the cost of own time. If so, the accumulation of human capital would on balance eventually increase later investment costs, and thus decrease the present value of later benefits.

To digress a moment, the presumption that the marginal costs of typical firms are rising is usually rationalized in terms of a limited "entrepreneurial capacity," an input that can only be imperfectly replaced by managers and other hired inputs. "Entrepreneurial capacity" is a construct developed to reconcile competition, linear homogeneous production functions, and determinate firm sizes, and most writers agree that there are no obvious empirical counterparts. Indeed, the extremely large size achieved by many firms suggests that, frequently at least, entrepreneurial capacity is not very limiting. Per-

88 This elasticity is relevant because investors may try to economize on their more costly time by substituting other inputs for time. Rough evidence of such substitution in education is found in the tendency for more valuable resources to be used per hour of the time of more advanced than less advanced students. The elasticity probably does not exceed unity, however, since the share of foregone earnings in total costs appears to rise with the level of education (see Schultz, op. cit.).

89 If $H$ measures the stock of human capital embodied in an investor, then the production function in footnote 81 can be expanded to include $H$, as in

$$h = f(R, T, H, B).$$

The productivity of greater human capital means a positive sign to $\partial h/\partial H$.

90 Even if the effect on productivity continued to exceed that on the cost of own time, diminishing returns would cause the decrease in investment costs to become smaller and smaller over time. (For an illustration of this in a model that is quite similar to, although more rigorously developed than, the one presented here, see Y. Ben-Porath, "The Production of Human Capital and the Life Cycle of Earnings," Journal of Political Economy, August 1967). On the other hand, the decrease in the present value of benefits that results from a decrease in the number of years remaining would become larger and larger over time.

91 This presumption can be justified by the observation that usually only firms producing a limited share of the output of an industry manage to survive.

sons investing in human capital can be considered "firms" that combine such capital perhaps with other resources to produce earning power. Since "entrepreneurial" time is required to produce human capital, and since the latter is embodied in the entrepreneur, teachers, managers, and other hired resources can only imperfectly substitute for him. Therefore, in this case, "entrepreneurial capacity" is a definite concept, has a clear empirical counterpart, and, as has been indicated, can lead to significantly rising costs, which in turn limits the size of these "firms."

It is the sum of monetary benefits and the monetary equivalent of psychic benefits (which may be negative) from human capital, not just the former alone, that determines the demand curve for capital investment. If one makes the usual assumption of diminishing monetary equivalents, marginal psychic as well as monetary benefits would decline as capital is accumulated. The considerable uncertainty about future benefits also contributes to a negatively inclined demand curve if there is increasing marginal aversion to risk as more capital is accumulated.

c. Supply curves. The supply curves in Chart 4 show the marginal cost of financing, as opposed to producing, an additional unit of capital. The marginal cost of financing can be measured, for simplicity, by the rate of interest that must be paid to finance an additional dollar of capital. If the annual repayment required on a "loan" was constant for the remaining period of labor force participation, the marginal rate of interest would simply equal the annual repayment on an additional dollar of funds, adjusted upward for the finiteness of the labor force period.

If the capital market were homogeneous, with no segmentation due to special subsidies or taxes, transaction costs, legal restrictions on lending or borrowing, etc., and if risk were constant, even a large change in the amount of capital used by any person would have a negligible effect on his marginal cost of funds since it would have a negligible effect on the funds available to others. In the actual world, however, the market for human capital is extremely segmented: there are local subsidies to public elementary and high schools, state and federal subsidies to certain undergraduate and graduate students, transaction costs that often make own funds considerably cheaper than borrowed funds, and significant legal limitations on the kind of borrowing that is permitted. The result is that although certain sources of funds are cheaper than others, the amounts available to any person from the cheaper sources are usually rationed since the total demand
for the funds tends to exceed their supply. This means that a person accumulating capital must shift from the cheapest to the second cheapest and on eventually to expensive sources. This shift from less to more expensive sources is primarily responsible for the positive inclination of the supply curve of funds even to one person. The rate of increase in each curve tends to be greater the greater the segmentation, since there is then greater diversity in the cost of different sources, with smaller amounts available from each.

The cheapest sources usually are gifts from parents, relatives, foundations, and governments that can be used only for investment in human capital. Their cost to investors is nil, and is represented in Chart 4 by the Og segment of the supply curve S that lies along the horizontal axis.93

Highly subsidized but not free loans from governments and universities, for example, that also can be used only for investment in human capital are somewhat more expensive: they are represented by the g'u segment of S. Then come the resources of investors themselves, including inheritances and other outright gifts, that could be used elsewhere. Their cost is measured by the foregone opportunities represented by the u'h segment of S. After these funds are exhausted, investors must turn either to commercial loans in the marketplace or to reductions in their own consumption during the investment period. These funds are usually available only at considerably higher, and somewhat rapidly rising costs: they are represented by the upward sloped segment h'S.

As emphasized earlier, the accumulation of human capital is not instantaneous, but is usually spread over a lengthy investment period. The rate of increase in financing costs, like that in production costs, would generally be less, the more slowly capital is accumulated because, for example, the accumulation of own resources could reduce the need to rely on more expensive sources.94 The rate of increase in

93 Conceptual separation of production costs from financial conditions suggests that direct government and private subsidies to educational institutions and other "firms" producing human capital might be included in the Og segment. When so separated, demand curves incorporate all production costs, not only those borne by investors themselves, supply curves incorporate all subsidies, and the rates of return relate "private" returns to "social" costs (for definitions of "private" and "social" see this volume, Chapter V).

94 Superficially, there are many actual examples of the cost of funds depending on the period or stage of accumulation, such as the special subsidies to students of medicine or advanced physics. Many of these are best treated, however, as examples of a segmented capital market for different kinds of human capital, and are more appropriately discussed in section 3g, where the interaction among different kinds is analyzed.
each supply curve also depends, therefore, on the accumulation pattern that is chosen.

d. Equilibrium. Since both the stream of benefits and of financing costs depend on the path of capital accumulation, the latter cannot be chosen with respect to either alone. The rational decision is to select a path that maximizes the present value of "profits"; that is, the present value of the difference between these benefits and costs. With a model as general as the one presented so far, the supply and demand curves shown in Chart 4 would not be uniquely determined nor independent of each other. In order to justify, therefore, uniqueness and independence and to permit a relatively simple analysis of income distribution, it is sufficient to assume that own time and hired inputs are used in fixed proportions to produce human capital, that a unit of hired inputs is available at a given price, and that a unit of own time is also available at a given price (foregone earnings) up to a certain maximum amount, beyond which no time is available at any price. If the analysis of income distribution presented in this essay turns out to be useful, the implications of more general assumptions about the production of human capital should be explored.95

With these assumptions, the value of benefits is given by the area under the unique demand curve shown in Chart 4, the value of financing costs by that under the unique supply curve,96 and the maximum difference is found by investing up to their point of intersection. At that point, marginal benefits equal marginal financing costs, which can be taken to mean that the marginal rate of return equals the marginal rate of interest.

Corresponding to the optimal accumulation path is an optimal investment period. If both the returns on each dollar invested and the repayments on each dollar borrowed were constant for the remaining labor force period, the current value of total profits, which is the difference between total returns and total repayments, would rise throughout the optimal investment period. A peak would be reached at the end, remain constant at that level throughout the labor force period, and then drop to zero.

95 A start is made by Ben-Porath, op. cit., section 4.
96 For simplicity, the figures in this essay plot along the vertical axis marginal rates of return and interest on each additional dollar of investment rather than the present or current values of marginal benefits and financing costs. If returns and repayment costs were constant for indefinitely long periods, marginal rates of return and interest would exactly equal the current values of the flow of benefits and financing costs respectively on an additional dollar of investment.
The earnings actually measured in national income accounts do not purport to represent the profits on human capital. For one thing, the costs of funds are not deducted from returns, regardless of whether they consist of direct interest payments, foregone income, or undesired reductions in consumption. During the investment period, moreover, some and often all the costs of producing human capital are implicitly deducted before reporting earnings. Consequently, measured earnings after the investment period only represent total returns, while during the period it is a hybrid of returns and production costs. I discuss first and most extensively the factors determining the distribution of measured earnings after the investment period, and only briefly consider the distribution of profits or of measured earnings during this period.

A major assumption of the remainder of this essay is that actual accumulation paths are always the same as optimal paths. Sufficient conditions for this assumption are that all persons are rational and that neither uncertainty nor ignorance prevents them from achieving their aims. Of course, these are strong conditions, and a fuller model would make room for irrationality, uncertainty, discrepancies between actual and "desired" capital stocks, etc. Given, however, our rudimentary knowledge of the forces generating income distributions, it is instructive to determine how far even a simple model takes us. What impresses me about this model are the many insights it appears to provide into the forces generating inequality and skewness in the distribution of earnings and other income. In any case, it can be easily generalized to incorporate many of the considerations neglected, such as uncertainty, or discrepancies between actual and "desired" capital stocks.

3. The Distribution of Earnings

This model implies that the total amount invested in human capital differs among persons because of differences in either demand or supply conditions: those with higher demand or lower supply curves

---

97 This intermingling of stocks and flows has many implications for age-earnings and age-wealth profiles that have been discussed elsewhere (see this volume, Chapters II, III, and VII).

98 Since all persons are very young during much of their investment period, it may seem highly unrealistic to assume that their decisions are rational. Children have their decisions guided, however, as well as partly financed, by their parents, and as long as parents receive some monetary or psychic benefits from an increase in their children's economic well-being, parents have an incentive to help children make wise decisions.
invest more than others. There is some evidence that in the United States, persons with urban employment or high IQ and grades tend to invest more in formal education than those with rural employment or low IQ and grades partly because the former receive higher rates of return.99 If the model is empirically correct, as assumed in the remainder of the essay, the sizable observed differences in education,100 on-the-job training, and other kinds of human capital would suggest sizable differences in either one or both sets of curves.

Persons who invest relatively large amounts in themselves tend to receive relatively high profits and measured earnings after the investment period. If they invest more because of higher demand curves, as $D'$ is higher than $D$ in Chart 4, both the area under the demand curve and the difference between it and the area under a given supply curve is greater (compare point $p'$ with $p$). If they invest more because of lower supply curves, the area under the supply curve for a given capital investment is smaller, and the difference between it and the area under a given demand curve, therefore, is greater.

a. "Egalitarian" approach. Instead of starting immediately with variations in both supply and demand conditions, I first treat a couple of important special cases. One of them assumes that demand conditions are the same for everyone, and that the only cause of inequality is differences in supply conditions. This can be considered an approximate representation of the "egalitarian" approach to the distributions of investments in human capital and earnings, which assumes that everyone more or less has the same capacity to benefit from investment in human capital. Investment and earnings differ because of differences in the environment; in luck, family wealth, subsidies, etc. which give some the opportunity to invest more than others. Eliminating environmental differences would eliminate these differences in opportunities, and thereby eliminate the important differences in earnings and investments.

Adam Smith took this view in his The Wealth of Nations when he said "The difference between the most dissimilar characters, between a philosopher and a common street porter, for example, seems to arise not so much from nature, as from habit, custom, and education." 101

100 For example, the standard deviation of years of schooling exceeds three years in more countries.
101 Modern Library edition, New York, 1937, p. 15. E. Cannan, ed., remarks that Smith was following David Hume, who said "consider how nearly equal all men are in their bodily force, and even in their mental powers and faculties, are cultivated by education" (quoted ibid.).
Currently, many persons in the United States argue that most persons are intrinsically equally capable of benefiting from a college education; only poverty, ignorance, and prejudice prevent some from acquiring one.

Generally, the most important cause of differences in opportunities is differences in the availability of funds. These in turn are derived from the same segmentation in the capital market which implies that cheaper funds are rationed, and that supply curves of funds are positively inclined even to individual investors. For a variety of reasons cheaper funds are more accessible to some persons than to others, and the former then have more favorable supply conditions. Some may live in areas providing liberal government and other subsidies to investment in human capital, or receive special scholarships because of luck or political contacts. Others may be born into wealthy families, have generous parents, borrow on favorable terms, or willingly forego consumption while investing. For all these reasons and more, supply curves of funds could differ considerably, and Chart 5 shows a few that differ in level or elasticity. For simplicity they are assumed to rise more continuously than the supply curve depicted in Chart 4.

If supply conditions alone varied, the equilibrium positions of different persons would be given by the intersections of the common demand curve with the different supply curves; the points $p_1$, $p_2$, $p_3$, and $p_4$ in Chart 5 represent a few such positions. Full knowledge of these positions, of the marginal rate of return associated with each amount of capital investment, would permit the common demand curve to be “identified.” Moreover, the marginal rates could themselves be “identified” from the earnings received by persons with different capital investments.

Persons with favorable supply conditions would invest relatively large amounts in themselves: the equilibrium positions in Chart 5 are further to the right, the lower the supply curves are. The distribution of the total capital invested obviously would be more unequal and skewed, the more unequal and skewed was the distribution of supply curves.

102 Of course, it is not the only cause; for example, discrimination and nepotism are often important, and yet usually affect the benefits from rather than the financial costs of investing in human capital.

103 Using the assumption that white males have the same demand curve for formal education, G. Hanoch first estimated the marginal rates of return to education from earning differentials between persons at different education levels, and then “identified” their common demand curve. See his Personal Earnings and Investment in Schooling, Ph.D. dissertation, University of Chicago, 1965, Chapter II.
If the labor force period was long, earnings would be related to the amount of capital invested by

\[ E = \tau C, \]  

(80)

where \( E \) is earnings, \( C \) the total capital invested, and \( \tau \) the average rate of return on \( C \). The distribution of \( E \) clearly depends on the distribution of \( C \); indeed, if the demand curve for capital was completely elastic, \( \tau \) would be the same for everyone, and the distributions of
earnings and investments would be identical except for a difference in units (\( \hat{r} \)) that depended on the aggregate supply of and demand for human capital. Since it is shown above that \( C \) is more unequally distributed and skewed the more unequal and skewed is the distribution of supply curves, the same applies to the distribution of \( E \).

As we have seen, the demand curve for capital investment is usually negatively inclined rather than infinitely elastic primarily because human capital is embodied in investors. \( E \) will, therefore, usually be more equally distributed than \( C \) because a given percentage change in \( C \) will change \( E \) by a smaller percentage since \( \hat{r} \) will decline as \( C \) increases and increase as \( C \) declines. Moreover, both \( E \) and \( C \) will be more unequally distributed and skewed the more elastic the demand curve is; for the greater the latter, the more that persons with favorable supply conditions would be encouraged to invest still more by a higher \( \hat{r} \); and the more that those with unfavorable supply conditions would be encouraged to invest still less by a lower \( \hat{r} \).

Similarly, an increase in the elasticities of supply curves that held constant their locations at the average value of \( C \) would also increase the inequality and skewness in \( E \) and \( C \). Persons with unfavorable supply conditions would be encouraged to cut back their investments at the same time that those with favorable conditions were encouraged to expand theirs.

In the Mathematical Appendix exact relations between the distributions of \( E \) and \( C \) and the parameters of supply and demand curves are derived under the special assumption that all supply curves have the same constant elasticity, and that the demand curve also has a constant elasticity. Among the results of this more special model is that earnings are likely to be less unequally distributed and less skewed than supply curves (that is, than opportunities).\(^{104}\)

b. "Elite" approach. At the other end of the spectrum is the assumption that supply conditions are identical and that demand conditions alone vary among persons. This can be considered an approximate representation of the "elite" approach to the distributions of investment in human capital and earnings, which assumes that everyone more or less has effectively equal opportunities. Actual investments and earnings differ primarily because of differences in the capacity to benefit from investment in human capital: some persons are abler and form an elite. In spite of the position taken by Smith and Hume, educational policy in England and some other parts of Europe has

\(^{104}\) See section 6 of the Appendix.
been predicated on a version of the elite view: "There is a tendency of long historical standing in English educational thought (it is not nearly so visible in some other countries) to concentrate too much on the interests of the abler persons in any group that is being considered and to forget about the rest." 105

Just as opportunities have been measured primarily by supply curves, so capacities are measured primarily by demand curves.106 For a given (dollar) amount invested, persons with higher demand curves receive higher rates of return than others; or looked at differently, they have to invest more than others to lower the marginal rate to a given level. Since all human capital is assumed to be identical, demand curves can be higher only if more units of capital are produced by a given expenditure. It is natural to say that persons who produce more human capital from a given expenditure have more capacity or "ability." 107

Since a higher demand curve means greater earnings from a given investment, in effect, ability is being measured indirectly; namely, by the earnings received when the investment in human capital is held constant.108 This approach is an appealing compromise between defi-


106 Let me repeat, however, that some differences in opportunities, such as those resulting from discrimination and nepotism, affect demand curves. Similarly, some differences in capacities affect supply curves.

107 If the production function notation of footnote 81 is used, the ith and jth persons have the functions

\[ h_i = f(R_i, T_i, B_i) \]
\[ h_j = f(R_j, T_j, B_j). \]

The ith person has more ability if \( f_i > f_j \) when \( R \) and \( T \), the inputs of market resources and own time, respectively, are held constant. If sometimes \( f_i > f_j \) and sometimes \( f_i < f_j \), there is no unique ranking of their abilities.

Note, however, that since demand curves incorporate psychic benefits and costs from human capital as well as monetary ones, \( i \) could have a higher demand curve than \( j \), and thus be considered to have more capacity, simply because he receives more psychic benefits than \( j \) does.

108 Note the similar definition by R. H. Tawney: "In so far as the individuals between whom comparison is made belong to a homogeneous group, whose members have equal opportunities of health and education, of entering remunerative occupations, and of obtaining access to profitable financial knowledge, it is plausible, no doubt, if all questions of chance and fortune are excluded, to treat the varying positions which they ultimately occupy as the expression of differences in their personal qualities" (Equality, Capricorn Books edition, New York, 1961, p. 121).
nitions of ability in terms of scores on IQ, personality, or motivation tests without regard to the effect on earnings, and definitions in terms of earnings without regard to opportunities. The former pay excessive attention to form and not enough to results, while the latter hopelessly confound “nature” and “nurture,” or ability and environment. Our approach directly relates ability to results, and at the same time recognizes the impact that environment has on results.

If demand curves alone varied, the capital investments and marginal rates of return of different persons would be found at the intersections of the different demand curves with the common supply curve. In Chart 6 there clearly is a positive relation between the height of a demand curve, the amount of capital invested and the marginal rate. Knowledge of the latter two quantities for many different persons would permit an “identification” of the common supply curve, just as such information earlier permitted an “identification” of a common demand curve.

An important difference, however, is that the marginal rates themselves could not now be “identified” from information on the earnings and investments of different persons alone. In Chart 6 the marginal rate of return to investing $OC_3$ rather than $OC_2$ would be proportional to the area $p_2C_2C_3q_2$ for persons with the demand $D_2$ and to the larger area $q_3C_2C_3p_3$ for those with $D_3$. If a marginal rate was simply estimated from the difference in earnings between persons investing $OC_2$ and $OC_3$, the estimate would be proportional to $D_2p_2C_2C_3p_3D_3$, which clearly greatly exceeds both true rates. To arrive at correct estimates, either the earnings of persons investing $OC_2$ would be adjusted upward by the area $D_3D_2p_2q_3$, or the earnings of those investing $OC_3$ adjusted downward by the area $D_3D_2q_2p_3$.

Note, incidentally, that those arguing that most of the differences in earnings between persons at different levels of education or training result from differences in ability are essentially assuming a common supply curve and steeply inclined demand curves.

Aside from chance, Tawney mainly stresses the importance of holding constant health, education, and financial knowledge, which are simply different kinds of human capital.


I have not tried to explain why some people are “abler” than others; this might ultimately be traced back to differences in numerous basic ability “factors.” For a model of this kind, see Mandelbroth, op. cit.

Some adjustments along these lines to estimated rates of return on formal education can be found in this volume, pp. 202–204.
Equilibrium Levels of Investment in Human Capital Resulting from Differences in "Abilities"
Earnings and capital investments are clearly more unequally distributed and skewed the more unequally distributed and skewed are demand curves. The same kind of arguments as those used in the previous section should make it apparent that both distributions are also more unequal and skewed, the greater the elasticities of the supply and demand curves. If the supply curve was positively inclined, the average rate of return would tend to be greater, the larger the amount invested. Therefore, earnings would tend to be more unequally distributed and skewed than investments.

In the Mathematical Appendix exact relations between the distributions of earnings and capital investment and the parameters of the supply and demand curves are derived under the special assumption that all demand curves have the same constant elasticity, and that the supply curve also has a constant elasticity. One of the more interesting results is that earnings and investments would necessarily be more unequally distributed and skewed than demand curves. If, for example, demand curves (i.e., capacities) were symmetrically distributed, both earnings and investments would be skewed to the right.

c. A comparison of these approaches. Before moving on to the general case that incorporates variations in both supply and demand conditions, it is illuminating to contrast the more important implications of these special cases. For under the guise of the “egalitarian” and “elite” approaches, they are frequently explicitly advanced and are still more widely implicitly assumed.

The “egalitarian” approach implies that the marginal rate of return is lower, the larger the amount invested in human capital, while the “elite” approach implies the opposite relation. Marginal rates of return appear to decline in the United States as years of schooling increase, which supports the “egalitarian” approach. However, in Canada, a country in many economic respects quite similar to the United States, estimated marginal rates do not decline consistently as schooling increases.

\[\text{112 See section 4 of the mathematical appendix on p. 138.}\]

\[\text{113 I say “appear” because these rates have not been fully corrected for differences in the average level of “ability” at different education levels; such a correction might eliminate the apparent decline (see Hanoch, op. cit., or this volume, p. 202).}\]

\[\text{114 See J. R. Podoluk, Earnings and Education, Dominion Bureau of Statistics, 1965. Note that since different years of schooling are not perfect substitutes, the pattern of rates are also affected by the relative demand for and supply of different years. Thus the relatively small number of college-educated persons in Canada might explain the relatively high rates of return to college education there.}\]
The inequality in earnings tends to be less than that in supply conditions in the "egalitarian" approach, and greater than that in demand conditions in the "elite" approach because the former implies a negative, and the latter a positive, correlation between rates of return and amounts invested. Put differently and perhaps more interestingly, to understand the observed inequality in earnings, the "egalitarian" approach has to presume greater inequality in opportunities than the "elite" one has to presume about capacities. Inequality in earnings is a more serious problem to the former, therefore, in the sense that a given observed amount implies greater underlying "inequities" or "noncompeting groups" than it does to the latter.

For a similar reason, the positive skewness in earnings is probably less than that in opportunities under the "egalitarian" approach and greater than that in capacities under the "elite" approach. Indeed, as pointed out in the last section, it is shown in the Mathematical Appendix using the assumptions of constant and identical elasticities of demand, and a constant elasticity of supply, that a symmetrical distribution of capacities necessarily results in a positively skewed distribution of earnings. Therefore, an age-old problem of economists —how to reconcile a skewed distribution of income with a presumed symmetrical normal distribution of abilities—turns out to be no problem at all. In the "egalitarian" approach, on the other hand, observed skewness is more difficult to explain because it implies still greater skewness in the distribution of opportunities, a skewness that may be associated with a skewed distribution of gifts and inheritance, etc.

The interpretation of income inequality in terms of noncompeting groups was popular among nineteenth and early twentieth century writers. For a review see H. Dalton, Some Aspects of the Inequality of Incomes in Modern Communities, London, 1920, Part II. "Groups" may be noncompeting either because of differences in opportunities, as assumed in the "egalitarian" approach, or because of differences in capacities, as assumed in the "elite" approach.

For example, A. C. Pigou said "Now, on the face of things, we should expect that, if as there is reason to think, people's capacities are distributed on a plan of this kind [i.e., according to a symmetrical normal distribution], their incomes will be distributed in the same way. Why is not this expectation realized?" The Economics of Welfare, 4th edition, New York, 1950, p. 650. See also P. A. Samuelson, Economics, 6th edition, New York, 1964, pp. 120-121.

It is not possible, however, to reconcile extremely large skewness in earnings with a symmetrical distribution of capacities.

Pigou's principal answer to the question he sets out in footnote 116 is largely based on a presumed skewed distribution of inheritances, which affects, among other things, the distribution of investments in training (ibid., pp. 651-654). Or, as Allyn Young said, "The worst thing in the present situation is undoubtedly the
d. A more general approach. If either all demand or all supply curves were identical, the supply and demand curves of persons investing the same amount would also be identical if different demand or different supply curves did not touch in the relevant region. This, in turn, means that all persons investing the same amount would have identical earnings. Yet if the amount invested is measured by years of schooling, there is abundant evidence of considerable variability in the earnings of persons with the same investment. Possibly improved measures of investment or the introduction of transitory earnings would eliminate most of the variability; I suspect, however, that a significant portion would remain. If so, neither special case—that is neither variations in demand nor in supply curves alone—is sufficient, although one set of curves might vary much more than the other.

If both supply and demand curves varied, different persons could invest the same amount, and yet some could earn more than others because they had higher demand (and supply) curves; in Chart 7, the same amount would be invested by persons with $D_3$ and $S_1$, $D_2$ and $S_2$, and $D_1$ and $S_3$. As this example indicates, knowledge of the various equilibrium marginal rates of return and investments would no longer be sufficient to identify either a supply or a demand curve because the equilibrium positions would be on different curves. Moreover, again the marginal rates themselves could not be identified from information on earnings and investments alone because persons with different investments would generally have different demand curves.

The distributions of earnings and investments would partly depend on the same parameters already discussed: both would be more unequal and skewed, the greater the elasticities of supply and demand curves, and the more unequal and skewed their distributions. The distributions of earnings and investments also depend, how-

---

extreme skewness of the income frequency curve . . . reflecting as it undoubtedly does the presence of a high degree of inequality in the distribution of opportunity" ("Do the Statistics of the Concentration of Wealth in the United States Mean What they are Commonly Assumed to Mean?" Journal of the American Statistical Association, 15, 1917, pp. 481-482). One should point out, however, that even "a high degree of inequality in the distribution of opportunity" is not sufficient to produce skewness in earnings, and that skewed distribution of opportunities is necessary, at least in the "egalitarian" approach.

119 For example, the coefficient of variation in the incomes of white males aged 35-44 in 1949 was 0.60 for high school graduates and 0.75 for college graduates (see this volume, Table 9, p. 182). Or in 1959, years of schooling explained less than 20 per cent of the variance in the earnings of white males aged 25-64 in both the South and non-South (see Table I, p. 92).
Ever, on a new parameter: namely, the correlation between different curves.

There are several reasons why supply conditions do not vary independently of demand conditions. Abler persons are more likely
to receive public and private scholarships, and thus have their supply curves shifted downward. Or children from higher-income families probably, on the average, are more intelligent and receive greater psychic benefits from human capital. On the other hand, private and public "wars" on poverty can significantly lower the supply curves of some poor persons. Since the first two considerations have, unquestionably, been stronger than the third, it is reasonable to presume a positive \(^{120}\) correlation between supply and demand conditions, perhaps a sizable one.

If supply and demand curves were uncorrelated, one might have the equilibrium positions given by \(p_{31}, p_{32}, \text{ and } p_{33}\) in Chart 7; if they were negatively correlated, by \(p_{21}, p_{22}, \text{ and } p_{23}\); and if they were positively correlated, by \(p_{11}, p_{22}, \text{ and } p_{33}\). The chart clearly shows that a positive correlation increases the inequality in both investments and earnings; it also increases skewness by increasing the earnings and investments of persons who would have relatively high earnings and investments anyway.

An impression of a negative correlation between supply and demand conditions—that is, between opportunities and capacities—is sometimes obtained from persons investing the same amount. As the curves \(D_3, S_1, D_2, S_2, D_1, \text{ and } S_3\) in Chart 7 clearly show, however, the supply and demand curves of persons investing the same amount must be negatively correlated, regardless of the true overall correlation between them. Valid evidence of this latter correlation is provided by information on the amount of variation in earnings "explained" (in the analysis of variance sense) by the variation in investments. For example, if the correlation between supply and demand curves was perfect and positive, all the variation in earnings would be "explained" by investments. Moreover, the smaller the algebraic value of this correlation, the less the variation in earnings is "explained" by investments, and the more that earnings vary among persons making the same investment.

Supplement: Estimating the Effect of Family Background on Earnings

One important implication of the above analysis on the interaction between opportunities and capacities (i.e., supply and demand condi-

\(^{120}\) By "positive" is meant that more favorable demand conditions are associated with more favorable supply conditions.

\(^{121}\) The issues considered in this addendum were already briefly considered by J. Mincer in "The Distribution of Labor Incomes: A Survey," *Journal of Economic Literature*, March 1970, p. 20.