Chapter 10

WAGE DETERMINANTS: A SURVEY AND REINTERPRETATION OF HUMAN CAPITAL EARNINGS FUNCTIONS

ROBERT J. WILLIS*

University of Chicago and Economic Research Center, NORC

1. Introduction

This chapter provides a survey and exposition of the development of the earnings function as an empirical tool for the analysis of the determinants of wage rates. Generically, the term "earnings function" has come to mean any regression of individual wage rates or earnings on a vector of personal, market, and environmental variables thought to influence the wage. As such, it has been applied to a wide variety of problems such as, for example, studies of discrimination by race or sex (see Chapter by Cain in this Handbook), the estimation of the "value of life" from data on job safety [Thaler and Rosen (1975)], or compensation for increased unemployment probabilities [Abowd and Ashenfelter (1981)].

The premier application, of course, is to the study of the effects of investment in schooling and on-the-job training on the level, pattern, and interpersonal distribution of life cycle earnings associated with the pioneering work on human capital by Becker (1964, 1975), Becker and Chiswick (1966), and, especially, by Mincer (1958, 1962, 1974). The bulk of this chapter is devoted to the theoretical and empirical development of the human capital earnings function during the past twenty-five years. In part, this restricted focus is justified by the importance accorded to investment in human capital as an explanation of wage differentials in the vast literature spawned by human capital theory. In addition, many of the analytical and statistical issues that arise in the estimation and interpretation of generic earnings functions also pertain to the study of other wage determinants or to tests of rivals to the human capital theory of wage determination.

* This chapter was written while the author was a member of the Economics Department, State University of New York at Stony Brook.
The standard human capital earnings function developed by Mincer (1974) is of the form

\[ \ln y = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + u. \]

The schooling coefficient, \( \beta_1 \), provides an estimate of the rate of return to education which is assumed to be constant in this specification. The concavity of the observed earnings profile is captured by the quadratic experience terms, \( x \) and \( x^2 \), whose coefficients, \( \beta_2 \) and \( \beta_3 \), are respectively positive and negative. Since early data sources such as Census data did not record a worker's actual labor force experience, a transformation of the worker's age was used as a proxy for his experience. Mincer uses the transformation \( x = a - s - 6 \), which assumes that a worker begins full-time work immediately after completing his education and that the age of school completion is \( s + 6 \). As an empirical tool, the Mincer earnings function has been one of the great success stories of modern labor economics. It has been used in hundreds of studies using data from virtually every historical period and country for which suitable data exist. The results of these studies reveal important empirical regularities in educational wage differentials and the life cycle pattern of earnings which are described later in this chapter.

To me, perhaps the most fascinating question concerning the human capital earnings function is why it should work so well. In a lucid survey of econometric problems that arise in estimating the returns to education, Griliches (1977, p. 1) presents a list of seven questions concerning the specification of an econometric model of earnings of this type. The fifth question is:

Why should there be a relation like this in the first place? In other words: (a) what interpretation can be given to such an equation? (b) What interpretation can be given to the estimated [schooling] coefficient? (b) Can one expect it to be “stable” across different samples and different time periods?

He goes on to say that he will “skip lightly” over several of the questions including the fifth question which he characterizes as the “...one really hard one on this list”.

In this survey, I will go into considerable detail in an attempt to deal with the set of issues raised by the “hard question” on Griliches’ list. In particular, I will argue that some of the issues he raises in this question and treats later in his paper can be both clarified and simplified by a reinterpretation of the theoretical underpinnings of Mincer’s earnings function within a framework which goes back to Adam Smith’s theory of equalizing differences and more recently to the theory used by Friedman and Kuznets (1945) in their explanation of income differences among independent professionals. It can also be regarded as a reinterpretation of Becker’s justly famous Woytinsky Lecture [Becker (1967, 1975)] which views
investment in human capital as the outcome of interaction between the supply of finance and the demand for investment. Unlike Becker, who assumes that human capital is homogeneous, I assume that each job or occupation entails a particular set of skills which a worker can acquire by combining his own innate talents with an appropriate duration and curricular content of schooling. The resulting theory tends to correct an imbalance in the human capital literature which has emphasized the supply far more than the demand for human capital.

Under certain conditions labeled "equality of opportunity" and "equality of comparative advantage", the earnings function is remarkably stable in the sense that, as long as the rate of interest remains constant, the structure of educational wage differentials tends to remain constant in the long run even in the face of substantial variations in the pattern of occupational demand arising from shifts in income, product prices, and production technology. After treating this special case, I show how variations in opportunity and comparative advantage influence the empirical form of the earnings function and use the framework to interpret some of empirical literature on "ability bias".

The chapter begins by surveying the empirical estimates of the rate of return to education and the pattern of life cycle earnings in Section 2. Section 3 discusses the derivation of human capital earnings functions under the assumption of homogeneous human capital and Section 4 introduces the model of heterogeneous human capital described above. Section 5 considers theoretical and econometric issues which arise when there is inequality of opportunity and ability and closes with a discussion of empirical findings concerning ability bias. In Section 6, I briefly describe some recent literature on several topics such as signalling, implicit contracts, and specific human capital which extend or modify certain aspects of the human capital model used in the rest of the chapter. This section is followed by some concluding remarks on topics for future research.

2. Statistical earnings functions

2.1. The theory in a nutshell

Additional schooling entails opportunity costs in the form of forgone earnings plus direct expenses such as tuition. To induce a worker to undertake additional schooling, he must be compensated by sufficiently higher lifetime earnings. To command higher earnings, more schooled workers must be sufficiently more productive than their less schooled fellow workers. In long-run competitive equilibrium, the relationship between lifetime earnings and schooling is such that (a) the supply and demand for workers of each schooling level are equated and (b) no worker wishes to alter his schooling level.
The preceding paragraph provides a nutshell summary of the human capital theory of educational choice. In order to extend the theory to explain educational wage differentials, it is necessary to specify how variations in earnings are divided between hours of work and hourly wage rates and how wages and hours are distributed over the life cycle. The essentials of the extended theory can be stated by replacing the word "schooling" with the term "on-the-job training" in the preceding paragraph.

In fact, the development of the human capital literature has not always followed the theoretical structure just outlined. In particular, the literature often emphasizes the supply side of the theory by focusing on individual decisions to invest in human capital but neglects the demand for human capital by firms and the implications of labor market equilibrium. Even the supply-oriented studies often treat schooling and patterns of post-school investment as exogenous rather than as the outcome of optimizing decisions.

For the most part, the failure of the literature always to meet standards of full theoretical purity is explained (and to a considerable extent justified) by the pragmatic trade-offs any applied economist must make between theoretical rigor, analytic tractability, and limitations of available data and econometric methodology. In this section of the chapter I will follow the historical development of the empirical literature on the returns to education and life cycle earnings functions without attempting to interpret it within the hedonic theory outlined above. In later sections, I provide a critique of the theoretical underpinnings of some of the empirical work and then offer a fairly detailed reinterpretation of the theory which, on the one hand, is consistent with the hedonic view of labor market equilibrium and, on the other hand, provides a justification in an important special case for the major empirical formulation of earnings functions pioneered by Mincer (1974).

In this section and in most of this chapter, I will follow a convention of the earnings function literature by assuming that the life cycle pattern of hours is fixed exogenously and will treat the life cycle patterns of hourly wages and annual earnings as essentially synonymous. The neglect of labor supply considerations provides considerable analytic simplification because it (along with certain additional assumptions) enables human capital investment decisions to be treated within a wealth rather than utility maximizing framework. An unfortunate consequence of this convention is that it has led to a bifurcation of the human capital and labor supply literatures which is only slowly being bridged. (See Chapter 11 by Weiss and Chapter 1 by Pencavel, respectively, in this Handbook for surveys of the life cycle earnings and labor supply literatures.) In particular, the omission of labor supply considerations is untenable when considering the returns to human capital investments for women because of their substantial commitment to non-market household activities and the high degree of variability of market labor over the life cycle. (See, for example, Mincer and Polacheck
Ch. 10: Wage Determinants

(1974) and Chapter 4 by Gronau in this Handbook. Consequently, the discussion in this chapter will be largely confined to male earnings.

2.2. Statistical earnings functions

Consider a hypothetical economy made up of workers who differ by years of schooling, s (which is assumed to begin at age 6); differ in age, t; and differ in the length of labor force experience, \( x = t - s - 6 \), but who are otherwise observationally identical. In this economy, data on annual earnings, \( y \), and years of schooling may be described by a statistical earnings function

\[
y = \varphi(s, x) + u,
\]

where \( \varphi(s, x) \) is the functional form that best fits the data and \( u \) is a residual with zero mean. Note that \( u \) captures the effect of any unobserved variables such as ability which influence individual productivity. For the time being, assume that \( u \) is statistically independent of \( s \) and \( x \).

In actual data, education and earnings are positively correlated. Assume that this is true when (1) is estimated so that \( \varphi(s, x) \) is positive monotonic in \( s \). Typically, the experience profile of earnings is positively sloped through most or all of working life and concave with the growth rate of earnings being highest at early ages and slowing or even turning negative at the later stages of life. Some illustrative profiles for different schooling groups based on cross-section data from the 1960 U.S. census are presented in Figure 10.1.

In terms of the theory outlined above, the function \( \varphi(s, x) \) in (1) may be interpreted as a hedonic price function in the sense of Rosen (1974) which reflects the equilibrium of the supply and demand for workers at each level of schooling and experience. In most of the following discussion I shall also assume that the economy and population are in long run, steady state equilibrium such that \( \varphi(s, x) \) holds cross-sectionally in each period and, hence, also describes the longitudinal earnings path of representative individuals in each cohort, conditional on their schooling. Underlying this assumption are assumptions of zero aggregate productivity change and a constant rate of population growth with an associated stable age distribution.

2.3. Internal rate of return

Beginning with the early studies of investment in education by Becker (1962, 1964), Hanoch (1967), Hansen (1963) and others, statistical earnings functions like \( \varphi(s, x) + u \) in (1) have been used to estimate the internal rate of return to
education. By definition, the marginal internal rate of return is that rate of
discount, \( p(s_1, s_2) \), such that the present value of the earnings streams net of
direct costs of education which are associated with two different schooling levels,
\( s_1 \) and \( s_2 \), are equated. In this section I will describe how this is done.

Ideally, the data used to estimate statistical earnings functions and the internal
rate of return to education would consist of complete longitudinal life histories of
the earnings of individuals beginning with their age of entry into the labor force
and ending with their retirement and would also provide information about the
direct costs of education such as tuition payments. Unfortunately, such ideal data
are seldom available. The early studies of investment in education typically used
cross-sectional census data to estimate the rate of return to education. Such data
contain information on current earnings of those in the labor force, age, and
years of education but no information on tuition paid, age of entry into the labor
force or age of retirement. Even the more recent longitudinal data sets such as the
Panel Study of Income Dynamics or the National Longitudinal Studies contain
only partial life histories of selected cohorts and very limited information on
Because of these data limitations, a more or less conventional set of simplifying assumptions have been made to permit estimates of the rate of return to education with available data. Since these assumptions also simplify the exposition, I will adopt them now and continue to use them throughout this survey unless otherwise noted. Specifically, assume that the only cost of schooling is forgone earnings, that individuals enter the labor force immediately upon the completion of schooling at age \( t = 6 + s \), and that each individual's working life of \( n \) years is independent of his years of education.\(^1\)

Given the additional assumption of a steady state with no productivity growth, the present value of the lifetime earnings of a "representative" individual with \( s \) years of education, evaluated at the age of school entry, is

\[
V(s, r) = \int_0^n \varphi(s, x)e^{-r(s+x)} \, dx, \tag{2}
\]

where \( \varphi(s, x) \) is based on the estimated statistical earnings function and \( r \) is a discount rate.

Let \( s < s + d \) be two levels of schooling, where \( d > 0 \), and let \( \hat{\rho}(s, s + d) \) be an estimate of the marginal internal rate of return to an individual with \( s \) years of schooling who invests in an additional \( d \) years. By definition \( \hat{\rho}(s, s + d) \) is the rate of discount that solves \( V(s, r) = V(s + d, r) \). Using (2), it is straightforward to show that this definition implies that

\[
\hat{\rho}(s, s + d) = \frac{1}{d} \left\{ \ln \left( \int_0^n \varphi(s + d, x)e^{-r(s+d+x)} \, dx \right) 
\right. 
- \left. \ln \left( \int_0^n \varphi(s, x)e^{-r(s+x)} \, dx \right) \right\}. \tag{3}
\]

In practice, \( \hat{\rho}(s, s + d) \) is usually unique because the age-earnings profiles of two schooling groups typically only cross once when \( \varphi(s, x) \) is chosen to be a smooth functional form which eliminates erratic sampling fluctuations in age-earnings profiles.

\(^1\)As will be discussed below, a distinction is sometimes made between "private" and "social" rates of return to take into account differences between the private and social costs of schooling under public education and between the private and social benefits of schooling due to the taxation of earnings. To the extent that the tax system is proportional, the use of after-tax or before-tax earnings do not affect the rate of return if there are no fixed costs (e.g. tuition). It has also been argued by Schultz (1960) and Becker (1964) that part-time earnings of college students in the United States tend to offset the bulk of the direct costs of college so that direct costs can be ignored without seriously affecting the estimated rate of return. [However, see Parsons (1974) for a critique of this assumption.] In this case, the "conventional" set of simplifying assumptions in the text yield estimates of both the private and social rates of return.
In general, the rate of return must be calculated using numerical methods. However, there are two simpler approaches which are of interest. First, suppose that the rate of growth of earnings at any given experience level is independent of the level of experience. In this case, the earnings function in (1) can be written in the weakly separable form:

\[ y = f(s)g(x) + u, \quad (4) \]

and the present value of lifetime earnings is

\[ V(s, r) = f(s)e^{-rs}\int_0^n g(x)e^{-rx}dx. \quad (5) \]

In this case it is easy to show that the estimated marginal internal rate of return to education is given by the logarithmic derivative of the statistical earnings function with respect to \( s \). Thus, using (3)–(5), it follows that

\[ \hat{\rho}(s, s + d) = \frac{\ln(f(s + d)) - \ln(f(s))}{d}. \quad (6) \]

Letting \( d \) become arbitrarily small, it is clear from (6) that the estimated return to a small increase in schooling above a given level of \( s \) is equal to the logarithmic derivative of the statistical earnings function in (4) evaluated at \( s \). That is,

\[ \frac{d\ln y}{ds} = \frac{\phi_s(s, x)}{\phi(s, x)} = \frac{f'(s)}{f(s)} = \hat{\rho}(s), \quad (7) \]

where \( \hat{\rho}(s) \) is the estimated marginal internal rate of return to schooling and \( \phi_s(s, x) \) is the partial derivative of the earnings function.

If the profile of log earnings with respect to experience of different schooling groups are approximately parallel, this result provides a rationale for utilizing regression methods to estimate the rate of return to education. For example, let

\[ \ln y = \ln(f(s)) + \ln(g(x)) + \epsilon \]

\[ = b_0 + b_1 s + b_2 s^2 + b_3 x + b_4 x^2 + \epsilon \quad (8) \]

be a regression function which is a quadratic approximation to the logarithm of the weakly separable earnings function in (4), where \( \epsilon \) is an error term. The estimated marginal rate of return to education is then \( \hat{\rho}(s) = b_1 + 2b_2 s \). Some empirical examples of this approach will be discussed below.

The logarithmic derivative of the statistical earnings function provides an estimate of \( \rho(s) \) only if it is assumed, as in (4), that a given increment of schooling has the same proportional effect on earnings at all levels of experience. If this assumption is not true, Mincer (1974) has suggested a "short cut"
approximate method of estimating the rate of return to schooling which avoids
the need for using numerical methods. It has the added advantage that the rate of
return can be estimated from data on the first ten years or so of a cohort’s
earnings history.

Mincer’s short cut method involves the use of an “overtaking” concept.
Specifically, assume that average earnings evolve according to the earnings
function \( y = \varphi(s, x) \) and let \( V_s \) be the present value of this earnings profile. Let \( \bar{y}(s) \) be a constant level of earnings which has the same present value. Now define
the overtaking experience level as \( x^*(s) \) such that \( \bar{y}(s) = \varphi(s, x^*(s)) \).

Given these definitions, it follows that

\[
V_s = \int_0^\infty \bar{y}(s)e^{-rx}dx = \alpha \bar{y}(s)/r = \alpha \varphi(s, x^*(s))/r, \tag{9}
\]

where \( \alpha = (1-e^{-r}) \). By analogy, let \( x^*(s+d) \) be the overtaking level of
experience for the earnings profile \( \varphi(s + d, x) \) which is associated with a higher
level of schooling, \( s + d \). Substituting (9) into (3), the marginal internal rate of
return is

\[
\rho(s, s + d) = \frac{\ln(\varphi(s + d, x^*(s + d)) - \ln(\varphi(s, x^*(s)))}{d}. \tag{10}
\]

This expression provides an empirically useful short cut method for estimating
the internal rate of return if the two overtaking levels of experience, \( x^*(s) \) and
\( x^*(s + d) \), are known. In this case, \( \rho(s, s + d) \) can be evaluated by simply
plugging the average log earnings levels of the two schooling groups at their
overtaking experience levels into (10). Mincer (1974) argues that the overtaking
level of experience will be less than or equal to the reciprocal of the internal rate
of return. For example, if the rate of return is about 10 percent, then the short
cut method may be applied by evaluating (10) using the average earnings of
individuals with about 8–10 years of experience.

Mincer develops the overtaking argument for a special case in which the
overtaking experience level is exactly \( 1/p \) regardless of the level of schooling. The
argument is as follows. Assume that individuals enter the labor force with an
earnings capacity of \( \bar{y}(s) \) dollars and that they invest \( C \) dollars in on-the-job
training in each year after leaving school for which they pay \( C \) dollars of forgone
earnings during the period of investment. The investments have a constant own
rate of return of \( \rho \) percent in perpetuity. Given these assumptions, the earnings
of an individual with \( s \) years of schooling and \( x \) years of experience is

\[
y(s, x) = \bar{y}(s) + \rho \int_0^x C dt - C = \bar{y}(s) + C(\rho x - 1). \tag{11}
\]

If the worker is assumed to have an infinitely long working life and the discount
rate is \( \rho \), then the present value of the earnings stream in (11) is \( \bar{y}(s)/\rho \) which is
also the present values of a constant earnings stream of \( \bar{y}(s) \). From (11), it is easy to see that \( x^*(s) = 1/p \) is the value of \( x \) which solves \( y(s, x) = \bar{y}(s) \).

Note that the growth path of earnings in (11) implies constant dollar growth (but decreasing percentage growth) as experience increases. Empirically, dollar growth in earnings tends to decrease as \( x \) increases. In this case, the constant level of earnings with the same present value as \( y(s, x) \) would tend to be lower than \( \bar{y}(s) \) and the overtaking point would tend to occur earlier. Thus, Mincer argues that the overtaking experience level will tend to be somewhat less than \( 1/p \).

The overtaking concept has an important implication for the distribution of individual earnings paths about the population average for individuals. In the special case described above, the earnings of all individuals with the same "earnings capacity" at school leaving [i.e. \( \bar{y}(s) \)] will be equal when \( x = 1/p \) but will differ at earlier and later values of \( x \) if individuals differ in their levels of post-school investment (i.e. have different values of \( C \)). In particular, individuals with high rates of investment in on-the-job training will tend to have lower initial earnings and higher earnings growth than comparable individuals who invest at a lower rate. Thus, the variance of earnings across individuals will tend to be U-shaped with the minimum occurring at \( x = 1/p \), assuming that initial earnings capacity and the rate of post-school investment are uncorrelated. At the minimum point, the variance of earnings is entirely a consequence variance in initial earning capacities due to differences in schooling or ability, but is independent of post-school investment. Evidence for U-shaped patterns of variance in life cycle earnings has been found by Mincer (1974), Hause (1980), and Dooley and Gottschalk (1984) among others.

2.4. The self-selection problem

A key assumption underlying the use of a statistical earnings function to estimate the rate of return to schooling is that it accurately represents the opportunity set faced by a typical individual (after controlling for observable exogenous characteristics such as race or sex). If it does, it is capable of answering counterfactual questions of the sort: "What would a given individual's (expected) life cycle earnings path be if he chose \( s_2 \) rather than \( s_1 \) years of school?"

From its inception, one of the major concerns of the literature on investment in human capital is the possibility that statistical earnings functions do not, in fact, correctly measure individual opportunity sets. For example, a large literature addresses the issue of the extent to which the estimated rate of return to education is upward biased because ability is unobserved and "high ability" individuals, on average, have higher schooling attainment than "low ability" individuals. [See Griliches (1977, 1979) for recent surveys of this literature.] If so, the residual, \( u \), in (1) will be positively correlated with \( s \) and the estimated
earnings function will be subject to an "ability bias" which overstates the earnings gain a person of given ability would achieve through increased schooling.

The fundamental problems are (a) that it is impossible to observe the life cycle earnings paths of the same individual who has made alternative schooling (or post-school) investments and (b) that it is impossible to observe all the variables (e.g. ability) which determine his earnings opportunities. At best, we observe the earnings path of a given individual who has chosen (or been assigned) a given level of schooling. Hence, any measure of the returns to investment must be based on the comparison of the earnings of different individuals who differ in levels of schooling.

If schooling levels (and post-school investments) were assigned at random for each ability group according to an experimental design, a statistical earnings function estimated from interpersonal differences in earnings, schooling, and experience would provide an unbiased estimate of the opportunity set of a typical individual in that group (i.e. it would provide the best estimate of the difference in life cycle earnings an individual could expect given alternative levels of schooling) because, by design, the error term, \( u \), in (1) would be independent of \( s \) and \( x \).

However, the basic behavioral hypothesis of economics is the hypothesis that economic agents select the most preferred alternative from their opportunity set. If the full opportunity set cannot be observed and opportunities vary across agents, then the act of optimal choice implies that market data are systematically censored and there is no guarantee that estimates based on interpersonal differences in earnings and schooling will accurately estimate the opportunity set of any individual in the population.

In the context of the literature on investment in schooling, this has come to be known as the "self-selection problem" [see Rosen (1977a), Willis and Rosen (1979), and Kenny, Lee, Maddala and Trost (1979)]. Clearly, however, the self-selection issue is ubiquitous in economics and will present difficult econometric problems in any situation in which the full opportunity set of each agent is not observed. Since many of the empirical issues, including the question of ability bias, that have arisen in the earnings function literature can be interpreted in terms of the self-selection problem various aspects of it will be discussed in detail as the survey proceeds. Before turning to these issues, it is useful first to describe some of the major empirical findings of this literature as it developed.

2.5. **Empirical internal rate of return studies**

At a gross level, the observed positive correlation between schooling and earnings provides support for (and indeed prompted) the hypothesis that education is an investment which receives a pecuniary return in the labor market [Schultz (1960,
Table 10.1
Estimates of private internal rates of return to successive levels of schooling: United States 1959.

<table>
<thead>
<tr>
<th>Schooling level</th>
<th>5-7</th>
<th>8</th>
<th>9-11</th>
<th>12</th>
<th>13-15</th>
<th>16</th>
<th>17+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites/north</td>
<td>0.218</td>
<td>0.163</td>
<td>0.160</td>
<td>0.071</td>
<td>0.122</td>
<td>0.070</td>
<td>–</td>
</tr>
<tr>
<td>Whites/south</td>
<td>0.144</td>
<td>0.182</td>
<td>0.188</td>
<td>0.093</td>
<td>0.110</td>
<td>0.073</td>
<td>–</td>
</tr>
</tbody>
</table>

Source: Hanoch (1967, Table 3).

This interpretation was strengthened in early studies of investment in education by Becker (1964), Hanoch (1967), Hansen (1963) and others which made calculations of the internal rate of return to education based on statistical earnings functions like \( \varphi(s, x) \) in (1).

In perhaps the most thorough of the early studies, Hanoch (1967) estimated a set of internal rates of return between pairs of schooling level by race (white and non-white) and region (north and south) using cross-sectional data from the 1960 U.S. Census one-in-one thousand sample. His estimates of marginal internal rates of return, \( \hat{\beta}(s_1, s_2) \), for whites by region are reproduced in Table 10.1, where \( s_1 \) is the indicated level of schooling and \( s_2 \) is the next level.

Hanoch's estimates show a clear pattern of decreasing marginal rates of return to schooling. This pattern is also evident in a number of other studies such as Hansen (1963), Becker (1964), and Mincer (1974). If correct, this pattern suggests the possibility that a redistribution of educational investment which reduced educational differentials would be efficient. However, there are a number of caveats to such a conclusion. For instance, low rates of return to graduate study may reflect the existence of substantial fellowships and scholarships which reduce the cost of schooling, but are not included in the estimation procedure. On the other hand, the high estimated rates of return to elementary school may result from ability bias. In addition, it may be noted that Mincer (1974) finds that the rate of return tends to be constant when he controls for weeks worked.

As Hanoch notes, the magnitude of the estimated rates of return in Table 10.1 appear to be similar to estimates of rates of return to physical capital estimated by Stigler (1963), but higher than the real interest rate. In rough fashion, therefore, these estimates tend to support the hypothesis that education is an investment for which individuals require compensation as opposed to the alternative hypothesis that schooling is a consumption activity for which no compensation is required.

The estimates presented in Table 10.1 are similar in magnitude to those obtained by Hansen (1963) using published data from the 1950 Census but somewhat lower than those obtained by Becker (1964) for 1940 and 1950. (Hanoch suggests that this difference may result from the crudity of the data used by Becker and from differences in estimated ages of entry into the labor force.)
Table 10.2
Time series returns to education in the United States.

<table>
<thead>
<tr>
<th>Year</th>
<th>Secondary</th>
<th>Higher</th>
<th>Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>1939</td>
<td>18.2</td>
<td>10.7</td>
<td>n.a.</td>
</tr>
<tr>
<td>1949</td>
<td>14.2</td>
<td>10.6</td>
<td>n.a.</td>
</tr>
<tr>
<td>1959</td>
<td>10.1</td>
<td>11.3</td>
<td>n.a.</td>
</tr>
<tr>
<td>1967</td>
<td>n.a.</td>
<td>n.a.</td>
<td>8.2</td>
</tr>
<tr>
<td>1968</td>
<td>n.a.</td>
<td>n.a.</td>
<td>8.7</td>
</tr>
<tr>
<td>1969</td>
<td>10.7</td>
<td>10.9</td>
<td>9.0</td>
</tr>
<tr>
<td>1970</td>
<td>11.3</td>
<td>8.8</td>
<td>9.0</td>
</tr>
<tr>
<td>1971</td>
<td>12.5</td>
<td>8.0</td>
<td>9.2</td>
</tr>
<tr>
<td>1972</td>
<td>11.3</td>
<td>7.8</td>
<td>8.5</td>
</tr>
<tr>
<td>1973</td>
<td>12.0</td>
<td>5.5</td>
<td>8.9</td>
</tr>
<tr>
<td>1974</td>
<td>14.1</td>
<td>4.8</td>
<td>8.5</td>
</tr>
<tr>
<td>1975</td>
<td>12.8</td>
<td>5.3</td>
<td>8.9</td>
</tr>
<tr>
<td>1976</td>
<td>11.0</td>
<td>5.3</td>
<td>8.3</td>
</tr>
<tr>
<td>1977</td>
<td>n.a.</td>
<td>n.a.</td>
<td>8.5</td>
</tr>
<tr>
<td>1978</td>
<td>n.a.</td>
<td>n.a.</td>
<td>8.5</td>
</tr>
<tr>
<td>1979</td>
<td>n.a.</td>
<td>n.a.</td>
<td>7.9</td>
</tr>
<tr>
<td>1980</td>
<td>n.a.</td>
<td>n.a.</td>
<td>8.3</td>
</tr>
<tr>
<td>1981</td>
<td>n.a.</td>
<td>n.a.</td>
<td>8.7</td>
</tr>
<tr>
<td>1982</td>
<td>n.a.</td>
<td>n.a.</td>
<td>10.2</td>
</tr>
</tbody>
</table>


In the preface to the second edition of *Human Capital*, Becker (1975) summarizes the evidence on the time-series pattern of the rate of return to investment in education in the United States based on his own work and on the research of others as suggesting that the rate of return tended to fall from 1900 to 1940 and then remained stable through 1970. It is widely believed that the rate of return to higher education fell sharply during the 1970s. This belief is supported by a summary of estimated rates of return to education from 1939–76 presented by Psacharopoulos (1981) which is reproduced in the first two columns of Table 10.2. His summary shows that the rate of return to a secondary education has fluctuated around an average value of about 10–12 percent over the entire period. In contrast, the return to college education was virtually constant from 1939 to 1969 at about 11 percent and then began a sharp fall to about 5 percent during the 1970s.

I have been unable to find more recent rate of return estimates for higher education in published sources. Consequently, I asked Finis Welch to estimate a set of cross-sectional statistical earnings functions using micro data from the March Current Population Surveys from 1968 through 1983 which could be used
to calculate rates of return to college education. The results of this exercise, which are presented in the third column of Table 10.2, provide a very different picture of recent trends in the rate of return to education than that given by the estimates summarized by Psacharopoulos. According to the CPS-based estimates, the rate of return to college education stayed within a narrow range of between about 8 and 9 percent during the entire period from 1967 to 1981 and rose to a little over 10 percent in 1982.

It should be emphasized that my rate of return calculations assume that the only cost of college is forgone earnings, that the typical college student spends exactly four years in obtaining his degree, and that the cross-sectional earnings profiles of a synthetic cohort are representative of the expected life cycle earnings path of typical members of the cohort of high school seniors in each year from 1967 to 1982. Each of these assumptions is patently counterfactual and relaxing them may make a difference. For example, Freeman (1977) finds that rising college tuition costs are partly responsible for the decrease in the rate of return to college education that he found during the early 1970s. He also makes an attempt to adjust for variations in expected earnings for true cohorts.

Recently, considerable attention has been given to possible changes in the structure of earnings caused by the dramatic increase in the number of young people entering the labor force as a consequence of the post-World War II baby boom and the rapid growth in the fraction of each cohort receiving college educations [see Freeman (1975, 1976, 1977, 1979), Welch (1979), Berger (1983a, 1983b), and Murphy, Plant and Welch (1983)]. The trends in the size and composition of the labor force are illustrated in Table 10.4 below by a set of average annual growth rates of the labor force by education level since 1920 with projections to 2000. The rapid acceleration of the growth of the total labor force

---

2 Professor Welch estimated regressions of log annual earnings on years of schooling and years of imputed experience in which experience is set equal to current age minus 16 for schooling less than 12 grades, age minus 17 for high school graduates, age minus 19 for 13 to 15 years of schooling and age minus 22 for 16 years and over. A variety of specifications of the functional form of this relationship were tried and a "preferred" form was chosen as the basis for the rate of return estimates presented in column C of Table 10.2. The preferred form involves a spline function which assumes that log earnings for any given schooling class grow linearly during the first 10 years of experience and follow a quadratic path thereafter. In addition, the linear spline is interacted with years of schooling to capture variations in early career earnings growth across schooling groups. Welch points out (personal correspondence, 20 December 1984) that the experience spline tracks the early career far better than the smooth quadratic popularized by Mincer. He writes: "Given this, I find it incredible that the profession sticks with the [smooth quadratic] model."

Using the preferred functional form, I calculated the estimated rate of return to a college education as follows. I first used the estimated regression coefficients to simulate life cycle paths of dollar earnings for a representative high school graduate from age 18 to 65 and a representative college graduate from age 22 to 65 and then calculated the rate of discount which brings the present value of the two simulated earnings streams into equality using the internal rate of return function in a spreadsheet program.

I am grateful to Professor Welch for supplying me with these regression estimates, but he should be held blameless for my interpretation of them.
Table 10.3
Average annual rates of growth in civilian labor force, ages 16–64.

<table>
<thead>
<tr>
<th>Years</th>
<th>Total</th>
<th>11 or less</th>
<th>12</th>
<th>13–15</th>
<th>16 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920–30</td>
<td>1.63</td>
<td>1.18</td>
<td>3.12</td>
<td>3.11</td>
<td>3.21</td>
</tr>
<tr>
<td>1930–40</td>
<td>1.44</td>
<td>0.48</td>
<td>4.09</td>
<td>3.87</td>
<td>3.16</td>
</tr>
<tr>
<td>1940–50</td>
<td>1.10</td>
<td>-0.24</td>
<td>3.35</td>
<td>3.34</td>
<td>3.71</td>
</tr>
<tr>
<td>1950–60</td>
<td>1.16</td>
<td>-1.03</td>
<td>3.77</td>
<td>3.12</td>
<td>4.37</td>
</tr>
<tr>
<td>1960–70</td>
<td>1.79</td>
<td>-1.73</td>
<td>3.94</td>
<td>5.07</td>
<td>4.14</td>
</tr>
<tr>
<td>1970–80</td>
<td>2.46</td>
<td>-2.72</td>
<td>2.94</td>
<td>5.75</td>
<td>6.44</td>
</tr>
<tr>
<td>1980–90</td>
<td>1.60</td>
<td>-3.76</td>
<td>0.80</td>
<td>3.21</td>
<td>5.15</td>
</tr>
<tr>
<td>1990–2000</td>
<td>0.96</td>
<td>-6.95</td>
<td>-0.81</td>
<td>2.15</td>
<td>4.05</td>
</tr>
</tbody>
</table>

Source: Dooley and Gottschalk (1984, Table 4).

during 1970–80 is a reflection of the baby boom, while the negative growth rates of those with fewer than 11 years of education and the extremely high growth rates of those with a college education indicate the effects of the dramatic increase in educational attainments in the population during this century.

To date, there appears to be agreement that the changing age-structure has had a significant effect on the structure of earnings, but there is less agreement about the likely persistence of the earnings disadvantage and low returns of those in the large cohorts. For example, Welch (1979) suggests that the major effects on relative earnings take place in the early phase of careers while Berger (1983) argues that this finding is a consequence of Welch’s econometric specification and finds evidence of greater persistence of cohort-size effects in his specification. In addition to changes in relative mean earnings, Dooley and Gottschalk (1984) show that there has been a significant increase in the variance of log earnings within schooling groups since 1970. They explain this increase, in part, as the consequence of increased post-school investment in human capital caused by expected increases in the rental rate on human capital resulting from the projected deceleration in labor force growth depicted in Table 10.3.

Rate of return studies have also been conducted in virtually every country in which at least fragmentary data on earnings by age and education exist. Results of many of these studies have been collated by Psacharopoulos (1973, 1981). His most recent summary table is presented in Table 10.4 which presents averages of marginal rate of return estimates for primary, secondary, and higher education from individual country studies. The countries are grouped by degree of economic development and by continent within the LDC category.

Table 10.4 also distinguishes “private” and “social” rates of return. The private rate of return assumes that the only cost of education is forgone earnings (because of public subsidy of direct schooling costs) and that earnings are net of
taxes. The social rate of return includes the direct cost of schooling and uses before tax earnings. Psacharopoulos notes that almost all the difference between the social and private rates of return is due to the direct costs of schooling. The reason is that (estimated) taxes tend to be approximately proportional to earnings so that an increase in the tax rate tends to reduce the opportunity cost of schooling and the benefits from schooling by the same proportion, leaving the rate of return unaffected.

It is readily apparent from Table 10.4 that estimated rates of return tend to be negatively related to the degree of economic development. In general, rates of return within the Advanced country category appear to be quite comparable to estimates for the United States. It may be noted that private and social returns diverge most markedly in the LDC and Intermediate categories. For example, in her study of educational wage differentials in Turkey, Krueger (1972) argues that there appears to be an excess of highly educated workers in Turkey relative to those with intermediate level skills. She argues that this is because private incentives to obtain higher education are very strong (the estimated private rate of return is 26 percent) even though the social return appears to be below rates available on financial or physical capital investments (the estimated social rate of return is only 8.5 percent). The apparent divergence between social and private rates of return to higher education in other parts of the developing world which is indicated in Table 10.4 suggests that her argument may be generalizable. 3

The brief survey of empirical estimates of rates of return to education in this section provides powerful support for the basic human capital hypothesis which regards education as an investment which must be compensated by higher lifetime earnings. Basically, there appears to be remarkable stability in educational wage differentials across time and space, although there are sufficient variations in both dimensions to provide fertile ground for explanation of the underlying determinants and consequences of changes in the supply and demand for human capital. I shall now turn to a rather detailed analysis and reinterpretation of the theoretical underpinnings of empirical human capital earnings functions.

3 It should be noted that most of the studies of rates of return in the developing countries rely on data on wage and salary workers. Such workers are surely unrepresentative of the labor force as a whole. Chiswick (1976) suggests that estimates based on wage earner data which omit the large and less educated self-employed sector will tend to overstate the rate of return. However, in a study of Iranian data, Henderson (1983) found that earnings functions for wage and salary workers were essentially indistinguishable from those of the self-employed when the latter group excludes the very unskilled in the "informal" sector (e.g. shoe-shine boys, etc.). Although I have not seen it discussed in my limited perusal of this literature, it seems to me that the measurement of earnings may be a serious problem in estimating rates of return in many LDCs because of the importance of household and non-market production in such societies. [However, see Kuznic and DaVanzo (1982) for an examination of the effects of alternative income measures on the distribution of family income in Malaysia.] It is possible that the rate of return estimates in Table 10.5 are overestimates for this reason since non-market income is likely to be of greater importance in rural areas and among the less educated.
Table 10.4
The returns to education by region and country type.

<table>
<thead>
<tr>
<th>Region or country type</th>
<th>Primary</th>
<th>Secondary</th>
<th>Higher</th>
<th>Social</th>
<th>Secondary</th>
<th>Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa (9)</td>
<td>29</td>
<td>22</td>
<td>32</td>
<td>29</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Asia (8)</td>
<td>32</td>
<td>17</td>
<td>19</td>
<td>16</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Latin America (5)</td>
<td>24</td>
<td>20</td>
<td>23</td>
<td>44</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>LDC average (22)</td>
<td>29</td>
<td>19</td>
<td>24</td>
<td>27</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Intermediate (8)</td>
<td>20</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Advanced (14)</td>
<td>a</td>
<td>14</td>
<td>12</td>
<td>a</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

*Not computable because of a lack of a control group of illiterates.

N = number of countries in each group.
Primary = primary educational level.
Secondary = secondary educational level.
Higher = higher educational level.

Source: Psacharopoulos (1981, Table II).

2.6. The human capital earnings function: Empirical results

In effect, the early rate of return studies allowed the functional form of the statistical earnings function to be dictated by the data. For example, many studies simply used tabulations of earnings by schooling by age or, when micro data was used [as in Hanoch (1967)], the regression specification was simply dictated by the best fit to the data.

A major development in the literature, initiated by Becker and Chiswick (1966) and carried to full fruition by Mincer (1974), sought to use the theory to restrict the functional form of the earnings function and thereby enhance the empirical content of the theory. This line of research attempts to integrate the theories of investment in education and on-the-job training pioneered by Becker (1964) and Mincer (1958, 1962) within an empirical framework which is compatible with more formal models of human capital accumulation such as the Ben-Porath (1967) model. This work was carried out with such ingenuity, sophistication and care by Mincer (1974), that the resulting function is often referred to as “the” human capital earnings function.

The standard human capital earnings function developed by Mincer (1974) is of the form:

\[ \ln y = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + u \]  

(12)

Using (7), it follows that the schooling coefficient, \( \beta_1 \), provides an estimate of the rate of return to education which is assumed to be constant in this specification. The concavity of the observed earnings profile is captured by the quadratic experience terms, \( x \) and \( x^2 \), whose coefficients, \( \beta_2 \) and \( \beta_3 \), are respectively
positive and negative. Since early data sources such as Census data did not record a worker's actual labor force experience, a transformation of the worker's age was used as a proxy for his experience. Mincer uses the transformation $x = a - s - 6$, which assumes that a worker begins full-time work immediately after completing his education and that the age of school completion is $s + 6$.

Mincer's justification for the earnings function in (12) represents a blend of theory and pragmatism. On the theoretical side, he assumes that the skills acquired by the worker through education and on-the-job training can be regarded as a stock of homogeneous human capital which influences the worker's productivity by the same amount in all lines of work for all employers. Following Becker's (1964) important distinction between firm-specific and general training, this implies that competition will force the worker (rather than the firm) to pay the costs of his training and will allow the worker (rather than the firm) to reap the returns from his accumulated investment.

The reason is that if the firm attempted to capture some of the returns from training investments, the worker could always move to another firm at a wage which reflects the full value of the human capital embodied in him. Thus, if the firm is to provide training, it will implicitly charge the worker by reducing his wages below his marginal product by the cost of training. Workers are willing to pay this implicit price because of the increase in their future earnings resulting from their increased productivity. It follows that the observed earnings of a worker at a given level of experience may be regarded as equal to the rental rate on his accumulated stock of physical capital minus the cost of his current investment.

Beginning with Ben-Porath (1967), a number of economists have attempted to characterize the life cycle earnings path that would follow from an optimal program of investment in education and on-the-job training. (See Chapter 11 by Weiss, in this Handbook, for a survey of this work.) These models commonly assume that the worker attempts to maximize the present discounted value of lifetime earnings net of the direct costs of investment. Maximization takes place subject to constraints imposed by a “human capital production function” which represents the worker’s ability to transform inputs of his own time and purchased goods (e.g. tuition, time of supervisor) into outputs of human capital and by his time budget which requires him to allocate his time between “learning” and “earning”.

These models have been quite successful in providing a rigorous foundation for the existence of life cycle earnings profiles which share some of the qualitative features of the Mincer earnings function in (12). Specifically, they suggest that a worker will tend to specialize in investment in the early portion of his life when his stock of human capital is low. This rationalizes specialization in education at the beginning of life. At some point, it pays the worker to combine earning with learning and he enters the labor force. Initially, the worker tends to invest at a fairly high rate so that the level of his observed earnings are low. However, as
time passes, his earnings will tend to grow rapidly both because of the rate of accumulation of the stock of human capital and because the optimal level of investment decreases. Eventually, the decrease in the rate of investment combined with depreciation on the existing stock of capital may result in a cessation of earnings growth. At this point, earnings reach a maximum and they tend to decrease until the age of retirement.

Unfortunately, the optimal human capital models are very difficult to implement rigorously in empirical work. First, they typically do not have a closed form solution so that the precise functional form for life cycle earnings implied by such a model is usually not known. The Mincer earnings function in (12) may be regarded as an approximation to this unknown functional form. Second, many of the concepts underlying the model, including the concept of human capital itself, are unobservable (or, at least, not usually measured in available data). In addition to human capital, the list of unobservables includes the rental rate on human capital, the rate of discount, the functional form of the human capital production function, the inputs of time and purchased goods used in investment, and the individual-specific parameters of the production function which may be interpreted as representing the interaction of individual's "learning ability" with the home, school, and work environments where learning takes place.

The earnings function in (12) represents a pragmatic method of incorporating some of the major implications of the optimal human capital models into a simple econometric framework which can be applied to the limited information available in Census-type data. Early in his book, Mincer states his key assumption. Specifically, he says that "For simplicity the rate of return is often treated as a parameter for the individual. This amounts to assuming that a change in an individual's investment does not change his marginal (hence, average, rate of return)" [Mincer (1974, p. 7)].

He then uses this assumption in combination with an assumption about the time path of investment over the individual's life cycle to derive the earnings function in (12). In particular, assume that an individual begins with a stock of human capital of $E(0)$ at the age of school entry, $t = 0$. Also assume that, at time $t$, he devotes a fraction, $k(t)$, of his earning capacity to investment in human capital and $1 - k(t)$ to earning, and that $\rho$ is the individual-specific rate of return. Given these assumptions, the instantaneous growth rate of his earnings capacity at time $t$ is

$$g(t) = \rho k(t). \quad (13)$$

Thus, at time $t$ his earning capacity is

$$E(t) = E(0)\exp\left\{ \int_0^t g(\tau) \, d\tau \right\} \quad (14)$$

and his actual earnings (i.e. earnings capacity minus current value of investment)
is
\[ y(t) = (1 - k(\tau))E(\tau). \] (15)

Schooling is regarded as an activity in which the individual devotes full time to investment (i.e. \( k(t) = 1 \) for ages 6 through 6 + s). From (15), it follows that earnings capacity upon school leaving is
\[ E(s) = E(0)e^{\rho s}. \] (16)

If no further investment took place after leaving school (i.e. \( k(t) = 0 \) for \( t > s \)), the individual's life cycle earnings profile would be horizontal at a value of \( y(s) = E(s) \). Taking the logarithm of both sides of (16), this implies that the schooling–earnings relationship is of the log-linear form:
\[ \ln y = \ln E(0) + \rho s. \] (17)

Theories of optimal human capital accumulation suggest that workers will continue to invest in on-the-job training after leaving school, but that amount of investment will tend to decline over time. The parabolic earnings function in (12) corresponds (approximately) to the assumption that the fraction of earnings capacity which is invested declines linearly during working life from an initial value of \( k(0) \) at the beginning of the work career to a value of zero at the end of the career.\(^4\) Thus, let \( k(x) = k(0) - (k(0)/n)x \), where \( n \) is the length of working life. In this case, earnings capacity is
\[ E(x) = E(s)\exp\left\{ \rho \int_0^x [k(0) - (k(0)/n)t] \, dt \right\} \]
\[ = E(s)\exp\left\{ \rho k(0)x - (\rho k(0)/2n)x^2 \right\}. \] (18)

Actual earnings net of investment cost are \( y(x) = (1 - k(x))E(x) \). Thus, (17) and (18) imply that
\[ \ln y = \ln E(0) + \rho s + \rho k(0)x - (\rho k(0)/2n)x^2 + \ln(1 - k(x)). \] (19)

Mincer treats the earnings function in (12) as an approximation to (19).

Estimates by Mincer (1974) of three alternative specifications of human capital earnings functions are presented in Table 10.5 based on data on white, non-farm men from the 1960 Census. Line 1 shows an estimate of the “schooling model” in

\(^4\)Mincer (1974) suggests several other possible assumptions about the time path of post-school investment which lead to somewhat different functional forms for the shape of the life cycle earnings profile. However, the quadratic function in (12) has proved by far to be the most popular in part because it is the simplest to estimate and in part because alternative functional forms do not appear to be superior on statistical grounds. For example, Heckman and Polachek (1974), using a Box–Cox test find that log earnings is the preferred dependent variable. Also, it may be noted that Heckman (1976) was unable to reject an earnings function of the form in (12) against the alternative hypothesis that earnings were generated by the Ben-Porath (1967) model.
Table 10.5
Estimates of human capital earnings functions.

<table>
<thead>
<tr>
<th>Equation forms</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( \ln y = 7.58 + 0.070s )</td>
<td>0.067</td>
</tr>
<tr>
<td>( (43.8) )</td>
<td></td>
</tr>
<tr>
<td>2. ( \ln y = 6.20 + 0.107s + 0.081x - 0.0012x^2 )</td>
<td>0.285</td>
</tr>
<tr>
<td>( (72.3) )</td>
<td>( (75.5) )</td>
</tr>
<tr>
<td>3. ( \ln y = 4.87 + 0.255s - 0.0029s^2 - 0.0043xs + 0.148x - 0.0018x^2 )</td>
<td>0.309</td>
</tr>
<tr>
<td>( (23.4) )</td>
<td>( (-7.1) )</td>
</tr>
</tbody>
</table>

\( y \) = annual earnings of white, non-farm males, 1959.

\( s \) = years of school completed.

\( x \) = years of experience measured by age-schooling-six.

\( t \)-ratios in parentheses.

Source: Mincer (1974, Table 5.1).

(17) which assumes no post-school investment. The estimated rate of return to schooling, given by the schooling coefficient, is 7 percent and the equation explains only 6.7 percent of variance in log earnings.

Omitting experience from the earnings function results in a downward bias in the schooling coefficient because schooling and experience tend to be negatively correlated due to the fact that at any given age those with more schooling of necessity have less experience. The extent of this bias is illustrated in line 2 which presents an estimate of the quadratic earnings function in (12). In this specification, the estimated rate of return rises to 10.7 percent and the coefficients of experience and experience squared imply that earnings growth is 8.1 percent at the beginning of working life and decreases continuously until it reaches zero after about 34 years of experience and becomes negative thereafter until retirement.\(^5\) The addition of the experience terms also markedly increases the explanatory power of the regression, raising \( R^2 \) to 28.5 percent.

The earnings function in (19) assumes that all workers have the same own rate of return to investment and that they all invest the same fraction of their earnings

\(^5\) It should be noted that the actual earnings growth of members of a given cohort of new entrants into the labor force will tend to be more rapid than the growth measured in the "synthetic" cohort of individuals from a given cross section if there is a positive trend of real wages in the economy. Sometimes, a constant growth rate of real wages is assumed in order to adjust for this bias [e.g. Ghez and Becker (1974)]. For example, if real wages are assumed to grow at 2 percent per year, the corresponding earnings function for a given cohort could be obtained by adding 2 percent to the coefficient of \( x \) in line 2 of Table 10.6. In this case, the earnings growth rate would be initially 10.2 percent and it would not fall to zero until after 42 years of experience. Note that the estimated rate of return to schooling is not affected by such an adjustment if it is assumed that the rate of wage growth is the same for all schooling groups.

The reduction in the rate of productivity change in the U.S. economy beginning in the 1970s together with evidence of changes in the wage structure by age and education discussed earlier suggest that cross-sectional data in more recent periods may be quite misleading indicators of the earnings functions faced by cohorts of current workers.
capacity at each level of experience (i.e. \( \rho \) and \( k(x) \) are both constant across workers). If workers differ in these characteristics, the estimated rate of return to schooling and the growth rate of earnings may vary across schooling classes. This possibility is explored in line 3 of Table 10.5 where schooling squared and the interaction term, schooling times experience, are added to the regression. The results indicate that the marginal rate of return to schooling is decreasing. Evaluating the derivative of log earnings with respect to education at 8 years of experience yields estimates of the marginal rates as 17.4 percent at 8 years of schooling, 15.1 percent at 12 years, and 12.8 percent at 16 years. Recall that this pattern of decreasing marginal returns is similar to that found in the rate of return studies discussed in the previous section. The negative interaction term indicates some tendency for percentage earnings differentials by schooling class to converge as experience increases. However, Mincer reports that both the non-linearity in schooling and the interaction term become insignificant when a variable controlling for weeks worked is added to the regression.\(^6\)

Earnings functions like those reported in Table 10.5, especially the form in (12) and line 2, have been estimated hundreds of times using both cross-sectional and longitudinal data sources from many countries. Almost all the earnings function estimates that I have seen indicate a concave log earnings–experience profile qualitatively similar to that implied by the earnings function in line 2 of Table 10.5. Psacharopoulous (1981) surveys estimates of the rate of return based on the schooling coefficient in earnings function regression with the following results. The average of the estimated (private) rates of return were 14.4 percent in the LDCs, 9.7 percent in the Intermediate countries, and 7.7 percent in the Advanced countries. It may be noted that these estimates are somewhat lower than the corresponding estimates obtained from direct calculation of internal rates of return which are given in Table 10.4. I am unable to provide an explanation for this.

\(^6\)More recently, a number of economists have tended to use weekly wages (i.e. annual earnings divided by weeks worked) in place of annual earnings as the dependent variable in earnings functions [e.g. Welch (1979)]. Given the failure of most human capital models to incorporate labor supply, unemployment, or retirement as endogenous variables, the choice between these variables is somewhat arbitrary.

The argument in favor of the weekly wage is presumably that it is a better measure of the effect of schooling or experience on earnings potential and, implicitly, that earnings potential is what people seek through their investment in human capital. Heckman (1976) provides a human capital model with endogenous labor supply in which such an approach is justified formally by the assumption that human capital has the same percentage effect on both market (i.e. earning) and non-market efficiency. On the other hand, several writers have argued that the payoff to investment in (market-oriented) human capital depends on the degree to which the capital will be utilized in market activities. This argument is often made in connection with explaining male–female differences in investment incentives and market earnings [see Mincer and Polachek (1974), Barzel and Yu (1984), Becker (1985), and Rosen (1983)] but could also apply to the extent that schooling and experience influence the risks of unemployment, etc. In this case, it may be argued that annual earnings provide a better measure of the return to investment in human capital.
The emergence of longitudinal data sets has allowed economists to investigate the evolution of the life cycle earnings of individuals. In one such study, Lillard and Willis (1978) attempted to determine the extent to which cross-sectional earnings differentials persist over time. They used data from the first seven years of the Panel Study of Income Dynamics (1967–73) to estimate a standard earnings function of the form in (12) in which the residual term is assumed to be composed of a person-specific “permanent” component and a serially correlated “transitory” component. If no explanatory variables other than year dummies are included they found that about 73 percent of the total variance in log earnings is due to the permanent component and that the transitory component displays a serial correlation of about 0.4. When schooling, experience, and experience squared are added to the regression, these variables explain 33 percent of the total variance in log earnings and 44 percent of the permanent component. Their estimates suggest that most of the cross-sectional variation in earnings across individuals is persistent and that a little over half of this variance is due to “unmeasured” factors which are not captured by observed schooling and experience differentials.

Longitudinal data has also permitted a closer examination of the trade-off between earnings growth and the initial level of earnings due to differential rates of investment in on-the-job training (OJT) which are predicted by Mincer’s overtaking concept which was discussed earlier. In terms of the model presented in this section, the argument is that increases in the fraction of earning capacity invested in OJT (i.e. variations in the parameter, k(0)) will lead to lower initial earnings and higher earnings growth. The level of earnings of individuals who differ in k(0) but are otherwise alike (i.e. have identical values of \(p\) and \(E(s)\)) will tend to be equal at the overtaking level of about 8–10 years of experience. At lower experience levels, there will be a negative correlation between earnings level and the growth rate of earnings and at higher levels the correlation will be positive.

In cross-section data, this prediction can be studied only by looking at the pattern of the residual variance of log earnings with respect to experience to see if it is U-shaped and reaches a minimum near the overtaking level. (It should also be recalled that a positive correlation between initial earnings capacity, \(E(s)\), and the rate of OJT investment implies that minimum variance will occur at a higher experience level and conversely for a negative correlation.) Mincer (1974) found a U-shaped pattern for individuals with 12 years of schooling but also found that the pattern was declining for those with 8 years and positive for those with 12 years of schooling. However, Dooley and Gottschalk (1984) find a U-shaped pattern for all schooling groups (with the exception of log weekly wages for college graduates) using within-cohort data from successive CPS cross-sections. Their estimates imply that minimum variance of log annual earnings tends to occur at about 23 years of experience and that the minimum for
log weekly wages occurs at about 13.5 years of experience. These estimates are consistent with a weak positive correlation between initial earnings capacity and the subsequent intensity of OJT investment. The difference between the results for annual and weekly earnings is probably due to the contribution to earnings variance of relatively high levels of job turnover and other transitory shocks during the early stages of the career.

In principle, the OJT hypothesis can be examined more precisely in longitudinal data because permanent components of variance can "control" for unmeasured differences in the levels of earnings across individuals, thereby removing the correlation between initial levels and subsequent growth that confounds cross-sectional estimates of overtaking. Hause (1980) exploits this aspect of longitudinal data with a small sample of Swedish white-collar workers during the early career stage and obtains results showing a substantial negative correlation between earnings levels and growth. His results indicate a (lower bound) minimum of the variance of log earnings occurs at a little over 5 years of experience and his results also suggest that variance of transitory shocks to earnings tend to be greatest in the initial years of working life and that the variance of these shocks diminish fairly rapidly. While Hause's estimates do not directly confirm the importance of OJT, they are certainly consistent with the hypothesis and no attractive alternative hypothesis has been proposed to explain such patterns of residual variance. In a somewhat similar model, Chamberlain (1978) also finds strong evidence for the OJT hypothesis using American data.

This brief survey of some of the empirical research based on human capital earnings functions has only skimmed the surface of a massive literature which utilizes this tool to study a wide variety of subjects which space constraints prevent me from describing in detail. Among these, to give a few examples, are studies of the black-white earnings differentials [e.g. Smith and Welch (1979)]; earnings differentials among other ethnic groups [e.g. Chiswick (1983a, 1983b)]; earnings of immigrants to the United States [e.g. Chiswick (1977, 1978, 1979)]; language as a form of human capital [e.g. McManus, Gould and Welch (1983)]; effects of school quality [e.g. Solmon (1975)]; evaluation of manpower training programs [e.g. Ashenfelter (1978)] and many other subjects. Another major area of research concerning the effect of ability differentials on earnings will be treated more extensively later in this chapter.

3. Homogeneous human capital models

3.1. Background

As a statistical model, the human capital earnings function developed by Mincer has provided the basis for a vast body of empirical research on the level and
distribution of life cycle earnings and the returns to education. This body of work reveals some striking empirical regularities concerning the structure of wage differentials which hold over periods of time and across societies which differ dramatically in technology, patterns of demand, and forms of social and economic organization. In my view, this body of work constitutes one of the major success stories of modern labor economics.

At the same time the pragmatic character of Mincer’s translation of human capital theory into an operational empirical tool has certain important drawbacks. From the point of view of individual behavior, the key assumption of human capital theory is the proposition that individuals choose to invest so as to maximize the present value of lifetime earnings. [Indeed, Rosen (1976a) argues that the entire economic content of human capital theory is contained in this hypothesis.] Mincer’s derivation of the human capital earnings function which was described in the preceding section is ostensibly based on this assumption. In fact, however, both the level of schooling and the time path of post-school investment are treated as exogenous.

Moreover, as noted above, Mincer treats the rate of return to human capital (i.e. \( \rho \)) along with initial earnings capacity [i.e. \( E(0) \)] and the fraction of capacity invested [i.e. \( k(0) \)] as unobservable individual-specific constants which may vary across individuals because of differences in ability, discrimination, etc. In effect, the resulting model provides an accounting scheme in which the distribution of observed earnings is related through an earnings function to the joint distribution of the observed variables, schooling and age, and the unobserved individual-specific parameters.

In this section I will describe some of the empirical and econometric difficulties that arise when individual optimization is taken into account. In order to keep the discussion simple, I will ignore post-school investment and concentrate on the schooling decision. Thus, throughout this section I assume that a given worker with \( s \) years of schooling has constant productivity and earns a constant labor income, \( y_i(s) \), from his entry into the labor force at age \( s + 6 \) to his retirement at age \( s + n + 6 \). Following Becker and Chiswick (1966) and Mincer (1974), this simplified model is called the “schooling model”.

The argument in this section may be summarized briefly as follows. Following Rosen (1976a), I show that the wealth maximization hypothesis is inconsistent with a simple log-linear schooling-earnings relationship of the form used in Mincer’s work. Then, using Rosen’s adaptation of Becker’s well-known 8

---

7 Of course, the economic content of human capital theory also includes the role of human capital in production and the consequences of market equilibrium of the supply and demand for human capital.

8 Actually, Mincer also allows the individual’s rate of return parameter on schooling investments to differ from the corresponding parameter on post-school investments and also introduces another unobserved individual-specific parameter to capture depreciation or obsolescence of the accumulated human capital stock.
Woytinsky Lecture model [Becker (1967, 1975)], I show how optimal investment in schooling varies with individual "ability" and "opportunity". The term "opportunity" refers to the terms on which an individual can finance investments in human capital and the term "ability" refers to his capacity to translate investments into higher productivity. This model is shown to lead to very difficult econometric problems due to self-selection. At the end of the section, I argue that many of these difficulties can be traced to the use of the "simplifying" assumption of "homogeneous" human capital.

3.2. Mincer's schooling model

At the beginning of his book, Mincer (1974) proposes what amounts to two very different approaches to the derivation of a human capital earnings function. One approach, which was described in the preceding section, treats the rate of return to schooling as an individual-specific parameter. In effect, this suggests that individuals have "human capital production functions" of the log-linear form:

\[ \ln y_i = \ln y_{0i} + \rho_i s_i, \quad (20) \]

where \( y_{0i} \) and \( \rho_i \) are viewed as individual-specific "ability parameters" of the \( i \)th individual. [See also (17) above.] The parameter \( y_{0i} \) may be regarded as the individual's basic earning capacity and the parameter \( \rho_i \) as his "learning ability" (i.e. his capacity to increase his labor productivity through additional schooling). Note that \( \rho_i \) also measures the \( i \)th individual's (constant) internal rate of return to investment in schooling.

If schooling is treated as exogenously determined and the model is estimated with the regression equation

\[ \ln y_i = \beta_0 + \beta_1 s + u_i, \quad (21) \]

the coefficients \( \beta_0 \) and \( \beta_1 \), respectively, provide estimates of the average level of initial earnings capacity and the average value of the rate of return parameter in the population. That is, \( \beta_0 = E(\ln y_i) \) and \( \beta_1 = E(\rho_i) \). The residual term, \( u_i \), is

\[ u_i = (\ln y_{0i} - \ln \bar{y}_0) + (\rho_i - \bar{\rho})s_i + \eta_i, \quad (22) \]

where \( E(\ln y_{0i}) = \ln \bar{y}_0 \) and \( E(\rho) = \bar{\rho} \) and \( \eta_i \) captures the effects of measurement error and transitory income components. Note that \( u_i \) is heteroskedastic because its variance is an increasing function of schooling.

The schooling model in (21) as derived from (20) appears to be consistent with the approach used by Mincer (1974) in most of his book. However, at the
beginning of the book (pp. 9–11), he proceeds to derive a log-linear earnings function in a completely different way. His derivation is as follows. Let $s$ and $s + d$ be two levels of schooling which differ by $d$ years and let $y(s)$ and $y(s + d)$ be two constant earning streams which are equal in present value when discounted at the market interest rate, $r$. From (2) it is easy to calculate that the respective present values of the two income streams are $V(s) = a e^{-rs}/r$ and $V(s + d) = a e^{-r(s+d)/r}$, where $a = (1 - e^{-r})$ is a correction for finite working life.

Equating these present values and rearranging, the result is

$$y = y_0 e^{rs},$$

where $y_0 = rV(s)$. Note that $y_0$ may be interpreted as the permanent labor income of a worker (adjusted for finite life) whose human wealth is $V(s)$. Taking the logarithm of both sides of (23), Mincer’s “schooling function” is

$$\ln y = \ln y_0 + rs.$$  

Clearly, the log-linear earnings functions in (20) and (24) are conceptually distinct. The function in (20) represents a hypothesis about the technology of human capital production, while (24) is simply a tautology which follows from the definition of present value. Moreover, it is also clear that the hypothesized technology in (20) is inconsistent with the hypothesis that individuals facing a given market interest rate, $r$, choose that level of schooling which maximizes the present value of their lifetime earnings. The problem, as Rosen (1976a) points out, is one of corner solutions. That is, given a constant internal rate of return, $\rho_i$, a wealth-maximizing individual will either choose zero schooling if $\rho_i < r$ or he will have an unlimited demand for schooling if $\rho_i > r$.

The corner solution problem is circumvented if it is assumed that each individual faces rising borrowing costs as he increases his investment in education. In this case, each individual invests to the point at which his marginal borrowing rate is equal to $\rho_i$. If all individuals faced the same schedule of borrowing rates, there would be a positive correlation between $\rho_i$ and the level of education chosen. This possibility appears to be contradicted by the data since, if anything, the estimated marginal rate of return to education (at least in advanced countries) appears to be a decreasing function of the level of schooling as can be seen from the rate of return estimates in Tables 10.1 and 10.5 above.

3.3. Rosen’s schooling model

In a section of his paper titled, “Education and Self-Selection”, Rosen (1976a) proposes a simple reinterpretation of the schooling model which meets these
theoretical objections. In effect, this model may be regarded as a simplified version of the Becker (1967, 1975) Woytinsky Lecture model or of a Ben-Porath (1967)-type model of optimal accumulation of homogeneous human capital under the conditions that post-school investment is ruled out, that the only cost of schooling is forgone earnings, and that each individual faces a constant interest rate.9

Let the human capital production function or “structural earnings function” for person \(i\) be

\[
\ln y_i = h(s; A_i),
\]

(25)

where \(A_i\) is a vector of exogenous variables which measure \(i\)'s economic ability. For now, assume that \(A_i\) is a scalar and that higher values of \(A\) indicate higher ability (i.e. \(h_A > 0\)). Using (7), note that

\[
\rho(s; A_i) = h_s(s; A_i)
\]

(26)

is the marginal internal rate of return to investment in schooling. In order to have an interior solution to the problem of optimal schooling choice, assume that the marginal rate of return to schooling is decreasing (i.e. assume \(h_{ss} < 0\) or, equivalently, that \(\rho_s < 0\)).

Following Becker (1967, 1975), assume that opportunities for financing their investments in education vary across individuals because, for example, of differential willingness or capacity of their families to support them or because of differential holdings of non-human wealth to serve as collateral for borrowing against future earning power. To stay within the wealth maximization framework, assume that person \(i\) faces a constant rate of interest,

\[
r_i = r(Z_i),
\]

(27)

at which he can borrow or lend where \(Z_i\) is a vector of exogenous variables such as family background and non-human wealth which influence his financing

9In his Woytinsky Lecture, Becker (1967, 1975) makes the intuitively appealing assumption that each person faces a rising supply curve of finance for educational investment. For example, some support an individual receives from his family may come in the form of a pure transfer with no opportunity cost to him, some may come in the form of loans at “low interest rates” and, if he wishes to pursue his education still further, he may have to borrow at high interest rates in the market.

It should be noted that he commits an analytical slip in his discussion by assuming wealth maximization as the individual's objective and he does not consider role of time preference. If the borrowing rate is not constant, the optimal investment problem becomes more complicated because the individual cares about the timing of the entire earnings stream, not just its present value. That is to say, it is no longer possible to appeal to the Fisher separation theorem to justify wealth-maximization as the first step in a two stage utility optimization problem in which the second stage is to choose the optimal consumption path subject to a wealth constraint. Among other things, when an individual faces a rising marginal cost of finance, subjective time preference plays a role in the investment decision. Thus, an individual with a high rate of time preference will choose less investment than another person of identical ability who faces the same financing conditions but has a lower rate of time preference. (See Chapter 11 by Weiss in this Handbook for more details.)
opportunities. For now, assume that $Z_i$ is a scalar and that increases in $Z$ are associated with improved (borrowing) opportunities so that $r' < 0$.

Person $i$'s optimal schooling choice is given by the problem

$$\max V(s) = ae^{-rs}y(s)/r$$
$$\text{s.t.}$$
$$y(s) = \exp\{h(s; A_i)\}. \quad (28)$$

The first-order condition for this problem implies that the individual should continue schooling until the marginal rate of return is equal to the interest rate. That is to say, using (25) and (26), the first-order condition may be written as

$$\rho(s; A_i) = r_i = r(Z_i). \quad (29)$$

The optimal schooling choice is obtained by inverting (29) to solve for $s$ so that

$$s = \rho^{-1}(A_i, r(Z_i)) = s(A_i, Z_i). \quad (30)$$

The individual's optimal earnings are then determined by substituting (30) back into the human capital production function in (25) to obtain:

$$\ln y = h(s(A_i, Z_i); A_i) = y(A_i, Z_i). \quad (31)$$

To aid in the discussion of this model, it is illustrated diagrammatically in Figure 10.2 for a low ability person (person 1) and a high ability person (person 2). The concave line labelled $h(s; A_1)$ is 1's human capital production function (i.e. his structural earnings function) and the concave line labelled $h(s; A_2)$ is 2's production function.

![Figure 10.2. Optimal schooling choice: equal opportunity and unequal ability.](image-url)
As Rosen (1977a) points out, the tautological version of Mincer's schooling function in (24) may be regarded as defining a set of iso-wealth curves in \((\ln y, s)\) space. Each curve has a slope of \(r_i\) and an intercept, \(\ln y_{0i} = \ln[r_i V(s)(1 - e^{-r_s})]\), given by an arbitrary level of the present value of lifetime earnings evaluated at the age of school entry, \(V(s)\). In Figure 10.2, it is assumed that there is equality of opportunity. That is, persons 1 and 2 are assumed to face the same interest rates. Therefore both share the same family of iso-wealth curves which are given by the positively sloped straight lines.

Maximum lifetime wealth is attained at the point of tangency of each person's production function and his iso-wealth curve at points \(a\) and \(b\) for person's 1 and 2, respectively. As drawn, the high ability person chooses a higher level of schooling and has higher earnings than the low ability person. He also has a higher level of human wealth. Note that equality of opportunity implies that the allocation of educational investment is efficient because it results in equalization of marginal rates of return across individuals.

Several points can be made about the empirical implications of this model and the econometric difficulties it presents. First, the situation depicted in Figure 10.2 illustrates the problem of ability bias caused by self-selection. Clearly, an estimated rate of return from data on schooling and earnings, given by the slope of a straight line connecting points \(a\) and \(b\), would be an overestimate of the marginal rate of return faced by either person. (Both have marginal rates equal to the common rate of interest.) In addition, the overall shape of a statistical earnings function estimated from such data would not resemble the structural earnings function of either person. Rather, the shape of the statistical earnings function, \(\ln y = \ln \varphi(s) + u\), would be heavily influenced by the shape of the distribution of ability in the population.

Another point is that under conditions of equality of opportunity and ability (i.e. all persons have equal values of \(A\) and \(Z\)) optimal schooling choices and the level of earnings would be the same for everyone. Thus, data from such a population would be incapable of identifying either the structural human capital earnings function or the rate of interest.

Moreover, if ability differences are to generate variation in schooling choice, they must influence the marginal rate of return to investment. That is, suppose that increased ability has "neutral" effect in that it has an equal percentage effect on earnings potential at each schooling level [i.e. suppose that the form of the structural earnings function is \(\ln y = A_i + h(s)\)]. In this case, optimal schooling choices would be identical for all ability groups because all individuals have identical marginal rate of return functions [i.e. \(\rho(s; A_i) = \rho(s)\)]. Conversely, the self-selection of high ability individuals to higher levels of schooling depicted in Figure 10.2 arises because it is assumed that an increase in \(A\) has a higher percentage effect on the productivity of an individual, the more schooling he acquires (\(\rho_{s,A} > 0\)).
From an econometric point of view, life would be easiest in the case of equality of ability and inequality of opportunity. In this case, all individuals would share identical structural earnings functions but would make different schooling choices and have different levels of earnings because they face different interest rates. For example, if all individuals have ability $A_1$, it is clear from Figure 10.2 that variation in the rate of interest will "trace out" the structural earnings function, $\ln y = h(s; A_1)$. In this special case, provided a suitable functional form for $h(\cdot)$ is chosen, a non-linear regression of $\ln y$ on $s$ would suffice to identify the structural earnings function. Life would be almost as easy if ability has a neutral effect as defined above. In this case, unobserved ability differences would simply generate random residuals about the estimated earnings function of a person of average ability.

It is worth noting that this example illustrates a point made earlier concerning the fact that optimizing economic behavior tends to censor the observations an econometrician needs to identify structural economic relations. The function, $h(s; A_1)$, is identified because it is assumed that there is an "imperfection" in the capital market which prevents equalization of marginal rates of return to investment in education across workers. Thus, it is only the misallocation of resources resulting from an assumed market imperfection which permits identification. If there were perfect capital markets, we would be back to the case of no variation in $s$ if abilities are identical or to the problem of ability bias due to self-selection if abilities are not equal.

It should be clear from this discussion that the econometric problems presented by this highly simplified model are severe. They would become still more difficult if the model is generalized to allow for post-school investment, non-constant interest rates, uncertainty, and other factors that may be important in explaining real world data. These problems are especially severe because, in practice, data on ability and on the financing opportunities available to individuals are not available. At best, some data sets contain proxies for ability in the form of IQ scores, scores on visual acuity tests, etc. and information on family income and other background variables which might proxy borrowing rates.

In the following section I argue that much of this difficulty can be traced to assumption of homogeneous human capital which is employed as a "simplification" by Becker, Mincer, and many (but not all) of the economists who have done theoretical and empirical work in human capital. In effect, the assumption of homogeneous human capital regards workers as bringing to the labor market a number of homogeneous "efficiency" units of labor which is proportional to their stock of accumulated human capital. Thus, all workers are perfect substitutes in production at ratios proportional to their endowment of efficiency units. Equivalently, the efficiency unit view assumes that a given investment in human capital increases an individual's physical productivity in all production activities by the same amount.
While this assumption is patently counterfactual, it is usually justified as a fairly innocuous simplification which enables the analyst to abstract from the details of occupational skills in order to focus on the major forces determining the distribution of earnings by schooling and age [see, for example, Becker (1975, p. 97)]. However, as I emphasized at the outset of this paper, human capital theory encompasses optimization on both sides of the market and assumes equilibration of the supply and demand for labor. If all types of labor are perfect substitutes, the demand for efficiency units of labor is perfectly elastic so that the relative wages of workers who differ in human capital stocks are fixed by technology. In order to generate variation in the amount of investment across workers, it is necessary to emphasize interpersonal differences in ability and opportunity which cause variation in the supply of human capital. This, in turn, leads to the self-selection issues discussed above.

An alternative which I explore in the following section is to drop the assumption of homogeneous human capital. The major features of Mincer's empirical analysis of human capital earnings functions emerge in an important special case of this model. Specifically, under conditions of “equality of opportunity” and “equality of comparative advantage” (a generalization of equality of ability), the simple log-linear earnings–schooling relation, the overtaking notion, and the U-shaped experience profile of residuals from the earnings function are generated by the model.

4. A model of heterogeneous human capital

4.1. The general model

In contrast to the homogeneous human capital model discussed above, assume that there are many types of human capital, each of which is specialized to a particular set of tasks. For convenience, I shall refer to each distinct set of tasks as an “occupation.”

Initially, assume that each occupation has a rigid educational qualification in terms of the duration and curricular content of the training required to practice the occupation. For example, suppose that some occupations such as janitors and ditch diggers require no formal education. Others such as plumbers and clericals require twelve years of school, but the training received by a plumber does not qualify him as a clerical and conversely. Similarly, accountants require college degrees but are unqualified for other occupations such as chemical engineering or computer salesman which also require college degrees. For simplicity, I continue to assume that there is no process of physical or mental maturation over the life cycle and no post-school training, depreciation, or obsolescence of skills so that
each person's productivity in his chosen occupation remains constant over his working life. Also, I continue to assume that there are no direct costs of schooling.

Formally, let there be \( m + 1 \) distinct schooling-occupation categories indexed from lowest to highest training requirements by \( j = 0, 1, \ldots, m \), where \( s_0 = 0 \) and \( s_0 \leq s_1 \leq \cdots \leq s_m \) are the minimum years of schooling needed to train for each occupation. For simplicity, assume that any schooling above the minimum requirement is unproductive. In general, workers vary in their occupational abilities (i.e. in their capacity to be trained for a given occupation). Let the vector

\[
l_i = (l_{i0}, \ldots, l_{im})
\]

be the ability endowment of the \( i \)th worker where \( l_{ij} \) is the number of efficiency units of labor (i.e. piece rate productivity) supplied by worker \( i \) in occupation \( j \), given that he has the requisite \( s_j \) years of schooling of the appropriate type.

The worker's opportunity set is given by a vector of potential earnings in each occupation,

\[
y_i = (y_{i0}, \ldots, y_{im})
\]

\[
= (w_0 l_{i0}, \ldots, w_m l_{im}),
\]

where the vector

\[
w = (w_0, \ldots, w_m)
\]

is a set of market-determined relative occupational "piece rates" or "skill prices". For example, suppose that \( l_{i0} \) measures the cubic feet of dirt worker \( i \) can dig per year if he received no education and that \( l_{mi} \) measures the number of heart transplants per year that he could perform if he became a heart surgeon, where the market piece rate is one dollar per cubic foot of dirt and \( w_m \) per heart transplant. Then he could earn \( y_{i0} = l_{i0} \) per year for \( n \) years beginning at age 6 or he could earn \( y_{mi} = w_m l_{mi} \) per year as a heart surgeon for \( n \) years at age \( s_m + 6.10 \)

According to the human capital hypothesis, the worker chooses that occupation and associated level and type of schooling which has the highest present value. It is important to point out that economists typically cannot observe the physical productivity of a worker (i.e. \( l_{ij} \)) or the market piece rate per unit of productivity (i.e. \( w_j \)), but can only observe earnings which is their product (i.e. \( y_{ij} = w_j l_{ij} \)) because most workers are paid by time rates (e.g. hourly wage rates or annual salaries) rather than by the piece [see Pencavel (1977) and Stiglitz (1975)]. However, the relationship between worker pay and productivity that is enforced by competitive labor markets implies that it is theoretically meaningful to distinguish between physical productivity and the price per unit of product even if this distinction cannot be verified by direct observation.
value. Let the present value to person $i$ of occupation $j$ be

$$V_{ij} = \int_{s_j}^{s_j+n} y_{ij} e^{-r_it} dt,$$

$$= \alpha_i e^{-r_j s_j y_{ij}/r_i}, \quad j = 0, \ldots, m,$$ (35)

where $i$'s earnings level, $y_{ij}$, is given by (34); $r_i$ is his (constant) rate of discount; and $\alpha_i = (1 + e^{-r_n})$. The worker's education decision rule is then

choose $s_i^* = s_k$ if $V_{ik} = \max(V_{i0}, \ldots, V_{im})$. (36)

Now consider the production side of the model. Within occupations, workers are assumed to be perfect substitutes in production at rates determined by their relative endowments, but they are imperfect substitutes across occupations either because they perform different tasks within firms in a given industry or because they enter into the production of different final products which are imperfectly substitutable in consumption.

Aggregate output of a composite good, $Q$, is given by the aggregate production function

$$Q = F(L_0, \ldots, L_m; K, t),$$ (37)

where

$$L = (L_0, \ldots, L_m)$$ (38)

is the vector of aggregate supplies of efficiency units of labor to each occupation $j$ ($j = 0, \ldots, m$), $K$ is the aggregate capital stock, and $t$ is a vector of variables summarizing the state of technology and pattern of consumer demand. For now, assume that both $K$ and $t$ are exogenous constants. Also assume that the dollar price per unit of $Q$ is unity.

Let $\{1, \ldots, N\}$ be the set of workers in the economy and let $a = (a_0, \ldots, a_m)$ be an assignment of workers to a given schooling–occupation class such that $i \in a_j$ ($j = 0, \ldots, m$) defines the set of workers in occupation $j$. Given the assignment, there are $N_j$ workers in occupation $j$, where $\sum_{j=1}^m N_j = N$ and the aggregate supply of efficiency units to the occupation is

$$L_j = \sum_{i \in a_j} l_{ij}.$$ (39)

Given the vector of aggregate supplies of efficiency units of labor, $L$, implied by assignment $a$, let

$$F = (F_0, \ldots, F_m)$$ (40)
denote the associated vector of marginal products, where \( F_j \) is the marginal product per efficiency unit of labor in occupation \( j \).

Labor market equilibrium is determined by the interaction of the aggregate supply and demand for workers in each occupation. Denote the market assignment in long-run competitive equilibrium by \( a(w^*) = (a_0(w^*), \ldots, a_m(w^*)) \). The equilibrium assignment occurs when the vector of market piece rates is \( w = w^* \) such that the aggregate number of efficiency units of labor in each occupation is supplied by workers who follow the decision rule in (36) is

\[
L(w^*, r) = (L_0(w^*, r), \ldots, L_m(w^*, r)),
\]

and market piece rates and marginal products per efficiency unit are equal, i.e.

\[
w_j^* \geq F_j, \quad \forall j = 0, \ldots, m,
\]

where the vector \( r = (r_1, \ldots, r_n) \) gives the discount rates faced by each individual in the population and the equality holds in (42) for all schooling–occupation categories for which there is a positive aggregate supply in equilibrium.

At the microeconomic level, this long-run equilibrium generates data on the length of schooling \( (s_i^*) \), occupation \( (i \in \alpha_j^*) \), and earnings \( (y_i^* = w_j^* l_{ij}) \) for each of the \( i = 1, \ldots, N \) individuals in the population. The schooling–earnings data generated by the market may then be described by the statistical earnings function

\[
y_i = \varphi(s_i) + u_i.
\]

In general, both \( \varphi(\cdot) \) and the distribution of the error term, \( u \), depend on production technology and the pattern of final demand which determine \( F(\cdot) \), and on the distribution of ability and opportunity in the population given, respectively, by the vectors \( l = (l_1, \ldots, l_n) \) and \( r = (r_1, \ldots, r_1) \).

### 4.2. Non-competing groups

The model outlined above in (32)–(43) is sufficiently flexible to be capable of generating a wide variety of relationships between schooling and earnings ranging from the Mincer-type schooling function, \( \ln y = \ln y_0 + rs \), in (24), to other possibilities that are in direct conflict with the spirit (if not the formalisms) of the human capital approach. Before considering the conditions under which Mincer’s results arise, it is instructive to illustrate this point by considering the following example of Cairnes-Mill “non-competing groups” which lies at the opposite extreme from the human capital model in terms of its empirical implications.
Assume that any given worker in the economy can be trained for one and only one occupation, but that there is diversity across workers in the occupation for which they are suited. Thus, let the first $N_0$ workers have ability endowments $(t_{i0}, 0, \ldots, 0)$ for $i = 1, \ldots, N_0$; the next $N_1$ workers have endowments $(0, t_{i1}, 0, \ldots, 0)$; and so on. Formally, each worker chooses that occupation for which his discounted lifetime earnings are highest, but the choice is trivial. Obviously, the aggregate supplies of labor to each occupation are perfectly inelastic and lifetime earnings net of training costs are pure economic rents. Given technology and the pattern of demand for final products, the equilibrium piece rate vector, $w^*$, in (42) depends solely on the distribution of ability in the population.

A priori, there is no reason in this example to believe that the statistical earnings function, $\varphi(s)$, in (43) is positively sloped or even monotonic. Moreover, as the capital stock, technology, and the pattern of final demands vary, the equilibrium piece rate vector, $w^*$, will tend to change in both level and pattern in ways that are difficult to predict. In turn, this will lead to a change in the level and shape of $\varphi(s)$ and a change in the distribution of $u$.

For example, the introduction of electronic computers vastly increases the speed with which accounting analyses can be prepared. If the elasticity of demand for accounting services is sufficiently inelastic, this would tend to reduce the demand for accountants. Since the supply of labor to accounting is perfectly inelastic, this shift in demand would reduce the equilibrium piece rate per efficiency unit of labor by accountants and, hence, reduce the earning of accountants relative to earnings in other professions. Conversely, the piece rate and earnings of accountants would increase if the demand for accounting services is sufficiently elastic.

More generally, in this extreme example of "perfectly" non-competing groups, one would expect that the relationship between schooling and earnings would be highly irregular in a given economy at a given time and that it would be extremely unstable over time and across countries because of variation in technology and demand patterns. Given the overwhelming evidence that schooling and earnings are positively and monotonically related in nearly all societies in all historical periods for which there is data on schooling and earnings, it is safe to infer that the capacities of the human agent are considerably more malleable than in the example just described.

By the same token, it is possible to imagine a society in which the assignment of workers to particular types of training and to occupations is arbitrarily determined by caste, hereditary guild membership, etc. Clearly, such an arbitrary allocation rule would generate an equally arbitrary and unstable schooling–earnings relationship because, once more, the supply of labor to each occupation is perfectly inelastic. Again, the evidence is against a hypothesis of arbitrary assignment in most societies.
4.3. Perfectly equalizing differentials

In view of this discussion, it is not surprising that the strongest version of the human capital hypothesis holds under conditions of equality of opportunity and a form of equality of relative ability that I call equality of comparative advantage. In this case, the long-run supply of labor (in efficiency units) to each occupation is perfectly elastic at a piece rate which is sufficient to equalize the present value to each individual of lifetime earnings in all occupations.

This pattern of equalizing differentials generates a Mincer-type statistical earnings function:

$$\ln y_i = \ln \varphi(s_i) + u_i$$

$$= \ln y_0 + rs_i + A_i,$$

(44)

where the error term, $u_i = A_i$, which is equal to person $i$'s "absolute advantage", is homoskedastic and statistically independent of $s_i$. Thus, (44) can be estimated consistently by ordinary least squares even when ability differentials (i.e. differences in absolute advantage) are not observed.

In addition, this earnings function is remarkably stable in the long run under conditions of varying technology, capital stock, and demand patterns. Specifically, if the interest rate, $r$, remains constant, (44) remains perfectly stable for different patterns of final demand (holding resources and technology constant) and only its constant term, $\ln y_0$, shifts as technology and resources vary. Thus, a theory of heterogeneous human capital based on the hypotheses of equality of comparative advantage and equality of opportunity constitutes an extraordinarily simple and powerful theory of educational wage differentials. Moreover, when post-school investment is introduced the resulting earnings functions possess all of the properties of Mincer's human capital earnings functions.

In contrast to the relation $\ln y = \ln y_0 + rs$ in (36), which simply follows from the definition of equal present value, it is important to point out that the earnings function in (44) holds only under certain very strong conditions. Thus, a human capital theory based on the theory of perfectly equalizing differentials is eminently falsifiable with data.

The theory of equalizing differences is one of the oldest theories of wage differentials in economics, going back to Adam Smith (see Chapter 12 by Rosen in this Handbook). It is also the basic framework employed by Friedman and Kuznets (1945) in their classic study of income differences among independent professionals which was, in turn, an important precursor to the development of modern human capital theory by Becker and Mincer. Indeed, both Smith's and Friedman and Kuznets' work figure prominently in Mincer's first paper on human capital [Mincer (1958)] and clearly have deeply influenced the subsequent
development of his work. Theories of the role of comparative advantage in labor markets also have a long history. Pioneering modern statements by Roy (1951) and Tinbergen (1951) have been followed by the work of Rosen (1978), Sattinger (1975, 1980), and others. Finally, as we have seen, the importance of financing opportunities has been emphasized by Becker (1967, 1975).

Despite these historical precedents, I have been unable to find a systematic exposition of the conditions under which the conventional human capital earnings function arises as the outcome of general equilibrium in the labor market, although Rosen (1977a) provides a brief description of the approach I elaborate here. Since the theory to be presented essentially duplicates the main results of Mincer's theory, I view it as a reinterpretation of his theory. The reinterpreted theory has several major advantages. First, it provides a clear and rigorous statement of the conditions under which the standard results occur. It also have certain stability properties which have not been emphasized in the past. Second, the model is a special case of a more general theory within which the empirical implications of departures from these conditions can be analyzed.

Viewed as an econometric model, (44) rests on two fundamental empirical hypotheses, one economic and the other non-economic. The economic hypothesis corresponds to the condition of equality of opportunity which is defined, as before, as the situation in which all individuals face a common interest rate, \( r_i = r \) for all \( i = 1, \ldots, N \). This condition will hold if the economic system provides sufficiently good access to finance and sufficiently free entry into schools and occupations to permit the marginal rate of return to educational investment to be equalized across individuals.

The non-economic hypothesis is that humans are sufficiently alike in their basic capacities that the distribution of educational and occupational choices is not influenced by ability differences. Of course, this will be true if all individuals have identical ability endowments, \( l_i = \bar{l} = (\bar{l}_0, \ldots, \bar{l}_m) \) for all \( i = 1, \ldots, N \) individuals in the population. A somewhat more general condition, called equality of comparative advantage is that individual ability endowments are identical up to a factor of proportionality.

Specifically, there is equality of comparative advantage if

\[
l_i = e^{A_i} \bar{l} = e^{A_i} (\bar{l}_0, \ldots, \bar{l}_m), \quad \forall i = 1, \ldots, N,
\]

where \( A_i \) is a person-specific scalar constant which provides a one-factor (i.e. one-dimensional) measure of ability or “absolute advantage”. Assume that \( A_i \) is scaled such that the mean ability level in the population is \( E(A_i) = 0 \) so that \( \bar{l} \) is the ability vector of the average person.

The derivation of the earnings function in (44) is simple. If the potential earnings of individual \( i \) follow the pattern

\[
y_{ij} / Y_{ij} = e^{\gamma t_i},
\]

where
then, from (35), it follows that the present value of lifetime earnings is equated across all \( j = 0, \ldots, m \) schooling-occupation choices faced by the individual; i.e.

\[
V_{ij} = \alpha y_{ij} e^{r_j} / r_i = \alpha_i y_{i0} / r_i
\]

so that \( V_{i0} = \cdots = V_{im} \), where \( \alpha_i \) is a finite life correction defined in (35).

Given equality of opportunity and equality of comparative advantage, there exists a unique vector of relative occupational piece rates such that the earnings pattern in (46) holds for each individual in the population. From (45), equality of comparative advantage implies that \( y_{ij} = w_j l_{ij} = w_j l_j e^{\delta_j} \) and equality of opportunity means that \( r_i = r \). Hence, the structure of potential earnings will follow the pattern in (46) for each individual if the market piece vector is \( w^* = (w_0^*, \ldots, w_m^*) \), where

\[
w_j^* = w_0^* \left( \frac{l_0}{l_j} \right) \exp(r_s) \quad \text{for} \ j = 0, \ldots, m. \tag{48}
\]

If the vector of market piece rates is \( w^* \), each individual will be indifferent among all potential schooling-occupation choices because each provides the same present value. When an individual is indifferent among alternative opportunities, I assume that his actual choice is random. This assumption implies that schooling choice and ability are statistically independent if \( w^* \) is the market piece rate vector and there is equality of opportunity and comparative advantage.

The piece rate \( w_j^* \) may be interpreted as the "supply price" per efficiency unit of labor in occupation \( k \) given that \( w_j = w_j^* \) for all other occupations \( j \neq k \). That is, if \( w_k < w_k^* \) then no one will choose occupation \( k \), and if \( w_k > w_k^* \), then all individuals will choose occupation \( k \). Hence, the long-run supply of labor to occupation \( k \) is perfectly elastic at a price of \( w_k^* \) per efficiency unit.

Since the likelihood that an individual will choose occupation \( k \) is independent of his ability, the expected earnings of a worker who chooses that occupation is \( \bar{y}_j^* = \alpha l_j w_k^* \), where \( \bar{a} = E(e^\delta) \) is the arithmetic mean ability level of the population and \( \bar{y}_j^* = E(y_j^*) \) is the arithmetic mean earnings of those who choose occupation \( k \). Similarly, the aggregate supply of labor to occupation \( k \) is \( L_k = \bar{a} l_k N_k \), where \( N_k \) is the number of workers in the occupation.

To complete the description of long-run competitive equilibrium, we need to consider the vector of demands for labor in each occupation derived from the aggregate production function in (37). As a special case, first consider the case of homogeneous human capital. Specifically, assume that labor of each occupational type is a perfect substitute for labor of any other occupational type at a technologically determined rate. Thus, let

\[
Q = F \left( \sum_{j=0}^m \theta_j L_k; K; t \right) = F(L; K, t), \tag{49}
\]
where the $\theta_j$'s are constant parameters determined by technology and $L = \sum \theta_j L_j$ is a scalar measure of the aggregate supply of labor in efficiency units.

Given (49), the marginal product per efficiency unit of labor in occupation $j$ is $F_j = \theta_j F'(L)$. Hence, the vector of relative marginal products in (40) is a vector of constants

$$F = (F_0, \ldots, F_m) = (\theta_0, \ldots, \theta_m), \quad (50)$$

and the demand for efficiency units of labor in any occupation $j$ is perfectly elastic at a piece rate $w_j = \theta_j$.

Since the $\theta$'s are technologically determined parameters, there is no reason to assume that they will follow any particular pattern. Almost surely, the piece rate associated with the "demand price" per efficiency unit for some particular schooling--occupation choice will produce higher lifetime earnings than any other choice. If so, all individuals would choose that occupation and there will be no observed variation in schooling.

Diversity in occupational choice and in the duration and curricular content of schooling depends on imperfect substitution among efficiency units of labor of different types. Given imperfect substitutability, the demand curve for efficiency units of each type of labor is downward-sloping with respect to its own piece rate. In general, the aggregate supplies of labor to each occupation will adjust so as to satisfy the equilibrium conditions in (41) and (42).

For example, suppose that all $m + 1$ occupations are "necessary" in the sense that the marginal product of each type approaches infinity as its quantity approaches zero [e.g. $F(\cdot)$ is Cobb-Douglas]. In this case, the equilibrium quantities of each type of labor will be positive (i.e. $L_j^* > 0$ for all $j = 0, \ldots, m$) and the equilibrium piece rates will be given by $w_j = w_j^*$ as defined in (48), where $w_0^* = F_0$ is the marginal product per efficiency unit of unschooled labor evaluated at the equilibrium vector of aggregate labor supplies $(L_0^*, \ldots, L_m^*)$.

More generally, it is possible that the equilibrium supply of some schooling--occupation categories will be zero because, as its quantity approaches zero, the marginal product per efficiency unit of labor in such an occupation is less than its supply price. Typically, this will be the case if a given type of training produces a type of skill which is a close substitute in production for the skills produced by alternative types of training or if the product of that occupation is a close substitute in consumption for the products produced by individuals in other occupations. For example, accountants trained to keep ledgers by hand are close substitutes in production for accountants trained to use computers and wheelwrights produce components for a product which is a close substitute in

$^{11}$ That is, suppose that occupations $j$ and $k$ are both chosen by a positive number of workers. In order for this to happen, it must be the case that $w_j^* = \theta_j$ and $w_k^* = \theta_k$ where $w_j^*$ and $w_k^*$ are defined by (46) and $\theta_j$ and $\theta_k$ are technologically determined constants. Both of these equalities would hold only under a highly improbable coincidence.
consumption for the product produced by auto manufacturers. Given contemporary technology, the value marginal productivity of those with "obsolete" training is lower than the supply price to the occupation and the long-run equilibrium supply of labor to such an occupation is zero. However, all occupations with positive equilibrium supplies will have long-run equilibrium piece rates given by the \( w^* \) vector defined in (46).

The individual-level data generated by this model of "perfectly equalizing differentials" is determined by the equilibrium assignment vector \( \alpha^*(w^*) = (\alpha^*_0(w^*), \ldots, \alpha^*_m(w^*)) \), where \( w^* \) is the equilibrium piece rate vector defined in (48). Any given individual \( i \) is randomly assigned to schooling-occupation class \( j \) with probability \( p_j = N_j^*/N \), where \( N_j^* \) is the equilibrium number of workers in occupation \( j \) implied by the relation \( L_j^* = \tilde{a}N_j^* \) and \( N \) is the total number of workers for \( j = 0, \ldots, m \). The earning of workers \( i \in \alpha_j^*(w^*) \) are given by

\[
y_{ij} = w_j^*l_{ij}
\]

\[
= \left[ \alpha_0^* \left( \tilde{I}_0/\tilde{I}_j \right) e^{\tau_j} \right] l_j e^{A_i}, \quad \forall i \in \alpha_j^*(w^*), \quad j = 0, \ldots, m,
\]

\[
= w_0^*l_j e^{\tilde{\tau}_j + A_i},
\]

(51)

where \( y_{ij}^* \) is distributed randomly with mean \( \tilde{y}_j^* = \tilde{a}l_j w_0^* \) and variance \( \text{var}(e^{A_i}) \).

Taking the log of both sides of (51) yields the log-linear earnings function in (44),

\[
\ln y_i = \ln y_0 + r s_i + A_i,
\]

where \( y_0 = w_0^*l_0 \). The present value of lifetime earnings of all individuals of the ability \( V_i^* = \alpha y_0 e^{A_i} \) is the same regardless of their actual choice of schooling and occupation. The assumption of random choice among indifferent alternatives implies that \( A_i \) is homoskedastic and that there is zero covariance between \( A_i \) and \( s_i \). Hence, a regression of schooling on log earnings will provide consistent estimates of \( \ln y_0 \) and \( r \) even if ability (i.e. \( A_i \)) is unobserved.

As mentioned above, this long-run earnings function is remarkably stable as technology, the pattern of final demand, or the stock of capital change, although it may vary in the short run. For example, consider a once and for all change in technology caused by the introduction of computers such that the piece rate productivity endowment of the typical individual (for whom \( A_i = 0 \)) changes from the vector \( \tilde{l} \) to the vector \( \tilde{l}' \), where \( \tilde{l}' > \tilde{l} \) in certain occupations.

In the short run, assume that the number of workers in each occupation is fixed. The effect of improved technology is to increase the number of efficiency units of accounting labor which reduces the marginal product per efficiency unit or market piece rate (i.e. \( w_j = F_j \) decreases because \( F_{jj} < 0 \)). Since each accountant enjoys a ten-fold increase in the number of efficiency units of labor he possesses, the earnings of accountants, given by the product of the marginal product per
efficiency unit and the number of efficiency units, may either rise or fall depending on the degree of substitutability between accounting services and other inputs. Moreover, because of cross-effects, market piece rates in other occupations also change. Types of labor which are complementary to accounting enjoy an increase in earnings and those which are substitutes suffer a decrease.

In the long run, the supply of labor to each schooling-occupation class will change in the same direction that short-run occupational earnings changed. Given equality of opportunity and comparative advantage, the vector of aggregate labor supplies \((L_0, ..., L_m)\) and piece rate vector \((w_0, ..., w_m)\) will adjust in the long run until the present value of each individual's potential lifetime earnings in each schooling-occupation choice are equated. Thus, from (46), this implies that the equilibrium piece rate in accounting must fall to one-tenth its previous value. Assuming that the interest rate, \(r\), remains unchanged, the structure of occupational wage differentials continues to be described by (47) and the statistical earnings function will continue to be of the log-linear form in (44). The only change in the earnings function caused by the technological improvement will be an upward shift in the intercept which reflects the higher level of aggregate productivity.

In general, changes in technology, fluctuations in income, baby booms and busts and so on can be expected to cause short-run changes in occupational wage rates and thereby cause cross-sectional and longitudinal educational wage differentials to depart from the long-run earnings function. The precise nature of the departure depends in part on the degree of substitutability or complementarity in production of the skills supplied by different occupation-education categories. Such demand-side factors have received the most attention in recent analyses of the effects of the baby boom which were discussed earlier. (See also Chapter 8 by Hammermesh in this Handbook.)

However, the heterogeneous human capital model suggests that several supply-side factors may also be of importance. As I have developed it, the model assumes that human capital has a putty–clay structure such that the only way of altering the skill composition of the labor force is for newly trained workers with skills in high demand fields eventually to supplant older workers whose training is in areas experiencing a decline in demand. A useful extension of the model would be to incorporate the possibility of changes in occupation by workers after they have received their education.

For example, one might distinguish between “inflexible” and “flexible” educational investment. A perfectly inflexible education is one which qualifies a worker for only one occupation. That is, ex ante, the typical worker could choose to obtain any element in the vector \((l_0, ..., l_m)\) but, ex post, his productivity endowment is \(l_j\) given a choice of occupation \(j\) and zero for all other occupations. In this case, changes in the skill composition of the labor force are achieved only through generational turnover and relative occupational wages will tend to fluctuate considerably as adjustment takes place. More flexible education (e.g.
Ch. 10: Wage Determinants

liberal arts vs. vocational education?) would augment a worker's productivity in more than one occupation, thus increasing the cross-elasticity of occupational labor supplies and reducing the degree of short-run wage fluctuations.

Another supply-side factor which has received very little attention is the role of interest rate variations. The theory suggests that the supply of educated labor should be sensitive to variations in the real interest rate. It is interesting to note that the ex post real rate of interest was very low and perhaps even negative during the 1970s when the rate of return to higher education in the United States began to fall. It is too early to tell whether the current extremely high real interest rates will result in a reversal of this trend.

4.4. Life cycle earnings growth and perfectly equalizing differentials

What are the implications of a model of perfectly equalizing differences when post-school investment in on-the-job training is allowed? The answer comes in two parts. The first part is that the equalizing difference approach itself can say nothing about the shape of the life cycle profile. For this, one needs a model of optimal human capital accumulation such as those surveyed by Weiss in this volume. However, since the model assumes that there are many different occupations which may vary in technological possibilities for learning, it would be appropriate to assume that each occupation offers one or more optimal accumulation paths. Thus, certain occupations may be learned quickly once one has the appropriate schooling while others may require a lifetime to master.

The second part of the answer, which I shall develop below, is that most of the stability properties of the schooling function derived above hold when there is equality of comparative advantage and opportunity. In addition, the resulting set of earnings functions have all of the properties which Mincer (1974) exploited to such great advantage. In particular, under conditions similar to those assumed by Mincer, (a) the schooling coefficient in a regression of log earnings on schooling and experience will estimate the rate of return to schooling; (b) the overtaking experience level is approximately $1/r$; and (c) the variance of log earnings will tend to be U-shaped with a minimum at the "overtaking" level of experience.

The second part of the answer presumes that we have solved out for a set of optimal paths of human capital accumulation which may be chosen by those who have completed a given number of years of schooling and now are about to embark on their work careers. Without loss of generality, assume that there are $m$ different possible paths for the $j = 1, \ldots, m$ different occupation-schooling categories such that

$$l_{ij}(x) = l_{ij}\exp\left(\int_0^x g_{ij}(t) \, dt\right), \quad (52)$$

where $l_{ij}(x)$ are the number of efficiency units of labor that individual $i$ has
accumulated in occupation \( j \) after \( x \) years of work experience and \( g_{ij}(t) \) is the instantaneous growth rate of his efficiency units at time \( t \). It is assumed that the time path of accumulation of occupational skill implied by \( g_{ij}(t) \) is determined according to an optimal program that depends upon the skills embodied in individual \( i \) at the time he leaves school, on the particular characteristics of occupation \( j \), and on his rate of discount.

Given his schooling–occupation choice and given that he follows the optimal post-school investment program for that occupation, the individual's life cycle earnings path will be given by

\[
y_{ij}(x) = w_j l_{ij}(x),
\]

where \( w_j \) is the market piece rate for that occupation. The individual then chooses an occupation–schooling category which provides the (optimized) life cycle earning path which maximizes the present value of lifetime earnings where

\[
V_{ij} = \int_0^n y_{ij}(x) e^{-r(s_j + x)} \, dx
\]

is the formula for the present value of earnings in occupation \( j \) with schooling level \( s_j \) for \( j = 0, \ldots, m \).

As before, the case of perfectly equalizing differentials occurs when there is equality of opportunity and equality of comparative advantage. In the context of post-school investment, equality of comparative advantage means that

\[
(\lambda_0(X), \ldots, \lambda_m(X)) = \mathbb{E}^{A_i}(\bar{\lambda}_0(X), \ldots, \bar{\lambda}_m(X)).
\]

That is, (55) implies that individuals at the same level of experience differ in occupational productivity only by a scalar factor of proportionality. In combination with (52), this also implies that each individual, regardless of his ability (i.e. regardless of \( A_i \)), will have the same path of instantaneous growth rates of earnings, given his choice of schooling and occupation.

In order to attract individuals into all occupations, the piece rate vector, \((w_0, \ldots, w_m)\), will adjust until the present value of each occupation is the same for each individual (i.e. \( V_{i0} = \cdots = V_{im}, \forall i = 1, \ldots, N \)). Given the equilibrium piece rate vector, \((w_0^*, \ldots, w_m^*)\) each individual will be indifferent among his schooling–occupation choices. Thus, the long-run supply of labor to each occupation will be perfectly elastic at an average wage rate of \( y_{ij}^*(x) = w_j^* l_j(x) \) for individuals with experience \( x \). Hence, the experience–earnings profiles in each occupation–schooling class will tend to be stable in the long run under conditions of equality of opportunity and comparative advantage for the same reasons given earlier for the stability of the earnings–schooling relationship. However, the shape of average experience–earnings profiles within and between schooling
classes may change if growth profiles differ across occupations and the occupational mix changes as a result of shifts in the pattern of labor demand.

The main features of Mincer's human capital earnings function are replicated in this model. For example, recall from (11) that the overtaking level of experience is exactly $1/r$ years if individuals make constant dollar post-school investments in on-the-job training each year. The same result holds for the model of perfectly equalizing differentials if each individual experiences linear growth in occupation-specific productivity.

Thus, consider two occupations, 1 and 2, and assume that $\hat{I}_j(x) = \hat{I}(s_j) + k_jx$ ($j = 1, 2$) is the number of efficiency units of labor in occupation $j$ that an educationally qualified worker of average ability (i.e. $A_i = 0$) utilizes at experience level $x$ and let $y(s_j, x) = w_j\hat{I}_j(x)$ be his earnings at that level of experience. If $k_1 > k_2$, workers in occupation 1 have more rapid wage growth than workers in occupation 2. If the relative piece rate, $w_1/w_2$, adjusts so as to equate the present value of lifetime earnings in the two occupations, it is straightforward to show that $y_1(x) = y_2(x)$ at $x = x^* = 1/r$ (assuming an infinite working life).

In this model, the schooling coefficient in a regression of log earnings on schooling and experience provides an estimate of the rate of return to schooling (i.e. $\rho = r$) if the distribution of post-school growth opportunities is independent of duration of schooling. In this case, the data will show no interaction between schooling and experience and the distribution of the residuals from the regression will be U-shaped with a minimum at $x^*$. Otherwise, it is necessary to estimate the rate of return directly (i.e. by calculating it from age-earnings profiles predicted from the regression) or by using Mincer's short-cut method by comparing log earnings at the overtaking point and the distribution of residuals may follow a different pattern.

4.5. **Generalizations**

The Mincer earnings function can occur under conditions which are somewhat weaker than those specified in the preceding section. For example, it is not necessary to assume that each occupation has rigid schooling qualifications nor is it necessary to assume that all workers have equal comparative advantage. For simplicity, I will discuss these generalizations in the context of the schooling model with no post-school investment.

---

12 Note that the worker may not utilize all of his labor because he devotes part of his time to training as in the Ben-Porath (1967) model. Alternatively, however, he may experience growth in productivity solely because of so-called "costless" learning-by-doing in which case he utilizes all of his labor potential at each point in time. Of course, since competition forces present values of all occupations to be the same, an occupation which offers better learning opportunities will force an entrant to pay for the opportunity by accepting a lower initial wage, a point made by Becker (1975).
To relax the assumption of rigid schooling qualifications, assume that each worker $i$ of ability $A_i$ is endowed with a separate human capital production function for each occupation $j$ of the form $l_{ij} = h_j(s, A_i)$. Given his occupational choice, he chooses an optimal schooling level so as to equate the internal rate of return and the interest rate in the manner described by Rosen's schooling model in (26)–(31). The analogue to the assumption of equal comparative advantage in this case is that the production functions for each occupation are of the form $l_{ij} = A_i h_j(s)$ for all $i = 1, \ldots, N$ workers and all $j = 0, \ldots, m$ occupations because the internal rate of return to schooling for each occupation is independent of ability. Thus, regardless of ability, each worker in occupation $j$ will choose the same level of schooling. Because of differences in the productivity of investment in education among occupations, the optimal schooling level will vary depending on occupational choice. The assignment of workers to occupations and the equilibrium distribution of earnings by occupation–schooling class is then determined in exactly the way described in the preceding sections.

It should also be recognized that wage differentials need not be perfectly equalizing for all workers in order to achieve the Mincer earnings function. All that is necessary is sufficient long-run mobility of some workers to maintain the equilibrium structure of educational wage differentials described in the preceding section. For example, suppose that half of the labor force has ability endowments satisfying the equal comparative advantage condition while the other half consists of individuals whose talents are completely specialized to one or another of the $m$ occupations. Assume that the supply of efficiency units of labor to each occupation by the latter set of individuals is given by the vector $(L_0, \ldots, L_m)$. In this case, the supply of efficiency units of labor to occupation $j$ is perfectly inelastic at an aggregate supply of $L_j$ provided by those individuals whose talents are specialized to $j$ and is perfectly elastic at a piece rate $w^*_j$ given by (48) for any aggregate supply $L_j > L_j$. As long as the pattern of labor demand is such that demand curves intersect the elastic portions of the occupational supply curves, the equilibrium pattern of earnings will generate the Mincer earnings function in (44).

5. Inequality of opportunity and ability

Even with the weaker conditions just described, it is not obvious on a priori grounds that one should expect the Mincer earnings function to hold exactly. In this section I will outline the implications of departures from the conditions of equality of opportunity and of comparative advantage for the equilibrium pattern of educational and occupational wage differentials. For expositional simplicity, I ignore post-school investment and concentrate on the "schooling model". The section is concluded with a brief survey of the empirical literature on "ability bias" in the returns to education.
5.1. Inequality of opportunity

First, consider the effect of relaxing the assumption of equality of opportunity while maintaining the assumption of equality of comparative advantage. Specifically, suppose that there is a distribution of discount rates. In this case the equilibrium labor market assignment will sort those individuals with the highest discount rates into occupations with the lowest schooling requirements, those with the next highest discount rates into occupations with the next lowest schooling requirement, and so on until those with the lowest discount rates are left to be assigned to those occupations with the highest schooling requirement.

The resulting equilibrium earnings function will be of a non-linear form:

$$\ln y_i = \varphi(s_i) + A_i,$$

such that $\varphi'' < 0$. Thus, the empirical earnings function will display a pattern of decreasing marginal rates of return to schooling similar to those in Hanoch’s (1967) study which were reported earlier in Table 10.1. In addition, the shape as well as the level of the earnings function will tend to vary as the pattern of labor demands for each occupation shift.

These propositions are illustrated in Figures 10.3 and 10.4 for a simple example. Assume that there are three occupations $j = 1, 2, 3$ with schooling requirements $s_1 < s_2 < s_3$. Also assume that all individuals have identical productivity endowments, but that they differ in their discount rates. The distribution of discount rates is given in Figure 10.3.

Assume that labor demand conditions are such that one third of the labor force is in each occupation. Let the points marked $\bar{r}$ and $r$ partition the distribution in Figure 10.3 into thirds corresponding to the areas marked “A, B, and C.” I first show that the equilibrium labor market assignment is such that workers for whom $r < \bar{r}$ choose $s_1$; those for whom $\bar{r} < r < \bar{r}$ choose $s_2$; and those for whom $r < \bar{r}$ choose $s_3$. I then show that equilibrium log earnings corresponding to this assignment is given by points $a$, $b$, and $c$ in Figure 10.4, where the slope of the line segment $ab$ is $\bar{r}$ and the slope of the segment $bc$ is $r$. Thus, the marginal rate of return to schooling is decreasing (i.e. the rate of return from $s_1$ to $s_2$ is $\bar{r}$ and from $s_2$ to $s_3$ is $r$ where $\bar{r} > r$).

Figure 10.3. Interpersonal distribution of interest rates.
The argument is simple. Let earnings in occupation 1 be $\ln y_1 = \ln y_a$ corresponding to point $a$ in Figure 10.4. If earnings in occupation 2 are at point $b$ where $\ln y_2 = \ln y_b = \ln y_a + \bar{r}(s_2 - s_1)$, the marginal individual with discount rate $\bar{r}$ will receive equal present value in either occupation 1 or 2 and, hence, will be indifferent between them. Given the wage structure corresponding to points $a$ and $b$, all individuals for whom $r > \bar{r}$ will prefer occupation 1 and those for whom $r < \bar{r}$ will prefer occupation 2. Thus, to maintain one third of the labor force in occupation 1, the equilibrium wage in occupation 2 must be at point $b$. By the same argument, the marginal individual with discount rate $r$ will be indifferent between points $b$ and $c$ where $\ln y_c = \ln y_b + r(s_3 - s_2)$ and a wage of $\ln y_3 = \ln y_c$ will lead individuals for whom $\bar{r} \leq r \leq r$ to choose occupation 2 and those for whom $r < \bar{r}$ to choose occupation 3. Thus, the wage structure represented by points $a$, $b$, and $c$ is the only one which will elicit a supply of one-third of the labor force to each occupation.

Note that this earnings function is sensitive to the distribution of labor demand. For example, suppose that the demand for the high skill occupation 3 increases while the demand for occupation 1 and its wage, $y_a$, remain constant. The wage in occupation 2 must also remain constant at point $b$ to maintain the incentives of those previously in occupation 1 to remain while the wage in occupation 3 must rise above $y_c$ to induce additional entrants. In the new equilibrium, the marginal rate of return to additional schooling from $s_2$ to $s_3$ will have risen.

5.2. Inequality of ability

In the discussions of perfectly non-competing groups and perfectly equalizing differentials, I have already described two forms of inequality of ability which
have radically different implications for the shape and stability of the earnings function. There are many other possible forms of inequality of ability and a corresponding plethora of possible empirical relationships between schooling and earnings.

Beginning with Becker's (1964) early work, a major concern has been the possibility that estimates rates of return to education overstate the "true" rate of return because of a positive correlation between schooling and ability. A large literature has developed since then which attempts to test for the presence of "ability bias" and to provide estimates of the true rate of return. Implicitly, this concept of ability bias assumes that ability is essentially a one dimensional characteristic. Clearly, alternative assumptions are possible and perhaps more plausible. Strength, intelligence, agility, dexterity, visual acuity, creativity, and so on are words describing various distinct "abilities" which are thought to be of differential importance in different tasks or occupations and which are possessed, presumably, in varying levels and proportions by different workers.

In general, it is extremely difficult (some would argue impossible) to obtain direct measures of "true" abilities. At best, we have proxy measures of certain dimensions of ability such as scores in IQ tests, tests of visual acuity, etc. Despite misgivings about the meaning of test scores, early attempts to determine the extent of ability bias simply included test scores in earnings regressions. [See, for example, Taubman and Wales (1974), and Griliches and Mason (1972).] More recently, statistical methods have been developed, especially by Chamberlain and Griliches, which permit true abilities to be regarded as unobservable latent variables. This approach often requires considerable information not found in typical census-type data such as, for example, data on siblings including fraternal and identical twins. Even with unusually rich data, the capacity of economists to ascertain the extent of ability bias turns on a set of difficult issues concerning identification, treatment of errors in the measured variables, and so on. Griliches (1977, 1979) provides excellent surveys of the literature in this area together with a detailed interpretation of the statistical assumptions and empirical results in this literature. A brief synopsis of these issues and the empirical results will be given later.

In addition to the question of unobservables, there is a question of the extent to which abilities are "innate" or "acquired". In essence, this is the fundamental issue raised by the human capital concept. There is no question that there is a high degree of heterogeneity in the skills which different individuals actually possess and utilize in the performance of different occupations. However, as we have seen, if individuals are innately alike in comparative advantage, the supply of skills adjusts endogenously so as to equalize net advantages across occupations and an unbiased estimate of the rate of return to education can be obtained without controlling for ability differences in absolute advantage. This suggests that the issue of ability bias turns on the degree to which there is variation in comparative advantage which is correlated with endogenous schooling decisions.
5.3. The Roy model

A. D. Roy (1951) provided an early and highly innovative verbal presentation of a model of the economic implications of exogenous ability variation for the occupational choice, the structure of wages, and the distribution of earnings. In recent years the implicit mathematical structure of the Roy model has been expressed explicitly by several authors. The most complete mathematical statement of the model is probably contained in Heckman and Sedlacek (1981) who also develop its econometric implications to study the effects of minimum wage legislation on employment and wages.

The Roy model has also been extended to allow for endogenous skill acquisition through education by Willis and Rosen (1979) who used it to study issues concerning ability bias, self-selection, and the wealth maximization hypothesis in educational choice. They found that education was selective on ability, but that a one-factor representation of ability was inadequate in the data they examined.

In this section I provide a fairly detailed sketch of the Willis–Rosen version of the Roy model as a background both for discussion of their empirical findings and the findings from other investigations of the ability bias issue. For purposes of comparison with other studies, I augment the Willis–Rosen–Roy model by introducing a set of exogenous innate abilities which underlie occupation-specific abilities. However, for simplicity, I omit consideration of life cycle earnings growth which Willis–Rosen do incorporate into their model. The model of perfectly equalizing differences which leads to the Mincer earnings function (and no ability bias) is a special case of this model.

Let each individual be endowed with two exogenous innate abilities called "strength" and "intelligence" which are denoted, respectively, by $\xi_1$ and $\xi_2$. Assume that these abilities are jointly normally distributed in the population with zero means (i.e. $\mu_1 = \mu_2 = 0$), unitary variance (i.e. $\sigma_1^2 = \sigma_2^2 = 1$), and correlation $\rho_{12}$.

Assume that there are only two occupations, A and B, where A requires a college education and B requires a high school education. An individual's abilities are assumed to influence his occupation-specific productivity multiplicatively so that the logarithm of the occupation-specific productivities of the individual are

$$a_i = \alpha_0 + \alpha_1 \xi_{1i} + \alpha_2 \xi_{2i},$$

and

$$b_i = \beta_0 + \beta_1 \xi_{1i} + \beta_2 \xi_{2i},$$

where $a_i = \ln l_{ai}$, $b_i = \ln l_{bi}$, and the $\alpha$'s and $\beta$'s are fixed coefficients (i.e. factor
loadings) which indicate the importance of each ability to occupation-specific productivity.

Given the assumption that $\xi_1$ and $\xi_2$ are jointly normal, it follows from (57) that $a$ and $b$ are also jointly normal with the following parameters:

\[
\begin{align*}
\text{means:} & \quad \mu_a = \alpha_0, \quad \mu_b = \beta_0; \\
\text{variances:} & \quad \sigma_a^2 = \alpha_1^2 + 2\alpha_1\alpha_2\rho_{12} + \alpha_2^2, \\
& \quad \sigma_b^2 = \beta_1^2 + 2\beta_1\beta_2\rho_{12} + \beta_2^2, \\
\text{covariance:} & \quad \sigma_{ab} = \alpha_1\beta_1 + (\alpha_1\beta_2 + \alpha_2\beta_1)\rho_{12} + \alpha_2\beta_2.
\end{align*}
\]

Note that the correlation between occupation-specific productivities (i.e. $\rho_{ab} = \sigma_{ab}/\sigma_a\sigma_b$) will tend to be positive if strength and intelligence are useful in both occupations even if strength and intelligence themselves are uncorrelated (i.e. $\rho_{12} = 0$).

For convenience, define

\[
\begin{align*}
a_i &= \mu_a + e_{ai}, \\
b_i &= \mu_b + e_{bi}, \\
-\varepsilon_i &= e_{ai} - e_{bi},
\end{align*}
\]

where $\mu_a = \alpha_0$ and $\mu_b = \beta_0$ are the population means of $a$ and $b$, $e_{ai} = \alpha_1\xi_{1i} + \alpha_2\xi_{2i}$, and $e_{bi} = \beta_1\xi_{1i} + \beta_2\xi_{2i}$. Thus, $E(e_a) = E(e_b) = E(\varepsilon) = 0$; $E(e_a^2) = \sigma_a^2$; $E(e_b^2) = \sigma_b^2$, and $E(e_a e_b) = \sigma_{ab}$ as defined in (58). Also, $E(\varepsilon) = 0$, $E(\varepsilon^2) = \sigma_\varepsilon^2 = \sigma_a^2 + 2\rho_{ab}\sigma_a\sigma_b + \sigma_b^2$, $E(\varepsilon e_a) = \sigma_{ea} = \sigma_{ab} - \sigma_a^2$, and $E(\varepsilon e_b) = \sigma_{eb} = \sigma_b^2 - \sigma_{ab}$.

The annual earnings of individual $i$ are $Y_i = W_a a_i$ if he chooses $A$ and $Y_{bi} = W_b b_i$ if he chooses $B$, where $W_a$ and $W_b$ are the market piece rates in $A$ and $B$. Using (6), the rate of return to a college education for a given individual is

\[
\rho_i = \rho_i(s_a, s_b) = \frac{\ln y_{ai} - \ln y_{bi}}{(s_a - s_b)}
= \frac{\ln(W_a/W_b) + (\mu_a - \mu_b) + (e_{ai} - e_{bi})}{(s_a - s_b)}.
\]

A wealth-maximizing individual will choose to enroll in college (i.e. choose $A$) if $\rho_i > r_i$ and will stop at the end of high school (i.e. choose $B$) if $\rho_i \leq r_i$.

Define the index function

\[
I_i = (\tilde{\rho} - r_i)(s_a - s_b)
= (\mu_a - \mu_b) + \ln(W_a/W_b) - r_i(s_a - s_b),
\]

where $\tilde{\rho}$ is the mean rate of return to college education in the population. Thus, a
wealth-maximizing individual will follow the decision rule:

choose A if $I_i > \varepsilon_i$; otherwise choose B, \hspace{1cm} (62)

where, from (59), \(-\varepsilon_i = \varepsilon_{ai} - \varepsilon_{bi}\).

Clearly, the decision to choose college (i.e. A) is more likely the lower is the
discount rate, $r_i$, and the higher is the individual’s productivity in occupation A
relative to occupation B (i.e. the lower is $\varepsilon_i = \varepsilon_{bi} - \varepsilon_{ai}$). In turn, the selectivity
of college choice on innate ability (i.e. $\xi_1$ and $\xi_2$) depends on the relative usefulness
of strength and intelligence in the two occupations, on the correlation between
the two abilities in the population and on the correlation between abilities and
the discount rate.

To focus on the determinants of selection on ability, assume that there is
equality of opportunity (i.e. $r = r_i$ for all $i = 1, \ldots, N$ individuals). The probability
that an individual chosen at random from the population will choose A is then

$$Pr(\text{choose A}) = Pr\left(\frac{I_i}{\sigma_i} > \frac{\varepsilon_i}{\sigma_i}\right) = F\left(\frac{I_i}{\sigma_i}\right),$$

(63)

and the probability of choosing B is $1 - F(I_i/\sigma_i)$, where $F(I_i/\sigma_i)$ is the c.d.f. of a
standard normal distribution evaluated at $I_i/\sigma_i$. Note that (63) also gives the
fraction of the population who go to college and supply labor to occupation A.

If individuals were randomly assigned to schooling–occupation classes, average
earnings in A and B would be $\ln y_a = \ln w_a + \mu_a$ and $\ln y_b = \ln w_b + \mu_b$, respectively, and the observed rate of return to college would be $\bar{\rho}$ as defined in (61).

Given the decision rule in (62), the expected earnings of individuals who actually
choose A and B will typically diverge from $\ln y_a$ and $\ln y_b$. This divergence may
be called “selectivity bias” due to self-selection, although the term “bias” is fully
appropriate only if our goal is to estimate the earnings potential of a randomly
chosen person. In addition, the rates of return received by those who choose
college will tend to be higher than $\bar{\rho}$ and the potential returns of those who did
not choose college will be lower than $\bar{\rho}$.

Let $\bar{\rho}_a$ be the average rate of return to college received by those who choose A
and $\bar{\rho}_b$ be the average potential rate of return that those who choose B could have
received if they had gone to college. (60) implies that the distribution of rates of
return is normally distributed with mean $\bar{\rho}$ and variance $\sigma_r^2$ and (61) implies that
the marginal individual who chooses college will receive a rate of return equal to
the interest rate, $r$.

The distribution of rates of return is depicted in Figure 10.5 by a normal
distribution with mean $\bar{\rho}$ and standard deviation $\sigma_r$. The distribution is parti-
tioned into two parts at the point where $\rho = r$. As drawn, most of the population
chooses B (i.e. the area to the left of $r$) and the remaining fraction chooses A (i.e.
the area to the right of $r$). The mean rate of return received by those who choose
college is indicated by $\bar{\rho}_a$ which is the mean of the right-hand portion of the normal distribution truncated at $r$. Similarly, the point marked $\bar{\rho}_b$ is the mean of the left-hand portion of the truncated distribution.

Mathematically, the expressions for $\bar{\rho}_a$ and $\bar{\rho}_b$ are

\[ \bar{\rho}_a = \bar{\rho} - \lambda_a \geq r \]

and

\[ \bar{\rho}_b = \bar{\rho} - \lambda_b \leq r, \quad (64) \]

where the "inverse Mills ratios",

\[ \lambda_a \equiv E(\varepsilon/\sigma_e | I/\sigma_e > \varepsilon/\sigma_e) = -f(I/\sigma_e)/F(I/\sigma_e) < 0 \]

and

\[ \lambda_b \equiv E(\varepsilon/\sigma_e | I/\sigma_e < \varepsilon/\sigma_e) = f(I/\sigma_e)/(1 - F(I/\sigma_e)) < 0, \quad (65) \]

are, respectively, the formulas for the means of the upper and lower tails of a truncated normal distribution [see Heckman (1979)]. Note that $\lambda_a$ is always negative and $\lambda_b$ is always positive. An important implication of this analysis is that $\bar{\rho}_a$, the average of the "true" rates of return received by those who go to college, is always greater than the interest rate if $\sigma^2$ is non-zero.

The terms $\lambda_a$ and $\lambda_b$ are also important in determining the mean earnings of those who actually choose A and those who actually choose B. Expected log earnings of those who choose college is

\[ \ln \tilde{y}_a = E(\ln y_a | I > 0) = \ln \tilde{y}_a + E(\varepsilon_a | I/\sigma_e > \varepsilon/\sigma_e) = \ln \tilde{y}_a + (\sigma_{\varepsilon_a}/\sigma_e)\lambda_a = \ln \tilde{y}_a - \alpha_\lambda \lambda_a, \quad (66) \]
and, by a parallel derivation,

$$\ln \tilde{y}_b = E(\ln yb|I < 0) = \ln \bar{y}_b + \beta_\lambda \lambda_b,$$

where the selectivity bias coefficients for A and B are, respectively,

$$\alpha_\lambda = (\sigma_a - \rho_{ab} \sigma_b)(\sigma_a/\sigma_e)$$

and

$$\beta_\lambda = (\sigma_b - \rho_{ab} \sigma_a)(\sigma_b/\sigma_e).$$

From an economic point of view, this self-selection process produces an efficient allocation of resources in the sense that the aggregate supply of efficiency units in A is maximized for any given aggregate supply of efficiency units in B because the equilibrium assignment selects individuals according to their comparative advantage. As the demand for A rises, \(w_a/w_b\) tends to rise causing the mean of the distribution of \(\rho\) in Figure 10.5 to increase. This induces workers who were previously in B to shift to A. The sign of the “selectivity bias” on \(\ln \tilde{y}_a\) (i.e. \(\alpha_\lambda\)) indicates whether the workers drawn from B into A tend to have lower or higher earnings potential in A than those workers previously in A. Similarly, the sign of the “selectivity bias” on \(\ln \tilde{y}_b\) (i.e. \(\beta_\lambda\)) indicates whether the workers drawn to A tend to have higher or lower earnings potential in B than the workers who remain in B.

If the rate of return to college is calculated by comparing the actual earnings of college and high school workers, the estimates rate of return is given by

$$\hat{\rho}_a = \left[\ln \tilde{y}_a - \ln \tilde{y}_b\right]/(s_a - s_b).$$

Using (64)–(68), the estimated rate of return may be expressed as

$$\hat{\rho}_a = \bar{\rho} - \lambda_a - \beta_\lambda (\lambda_b - \lambda_a)$$

$$= \bar{\rho}_a - \beta_\lambda (\lambda_b - \lambda_a),$$

where \(\lambda_b - \lambda_a > 0\). Thus, the estimated rate of return to college will overstate the actual average return received by those who attend college (i.e. \(\bar{\rho}_a\)) if \(\beta_\lambda < 0\) and will understate it if \(\beta_\lambda > 0\). The intuition is that the average forgone earnings of those going to college is understated by the earnings of high school graduates in the former case and overstated in the latter case.

The direction of selectivity bias and its influence on the estimated rate of return to a college education depends on the underlying parameters, \(\sigma_a\), \(\sigma_b\), and \(\rho_{ab}\), which describe the population distribution of productivity in the two occupations. These parameters, in turn, depend on the underlying distribution of innate ability and the role of innate ability in determining occupational productivity. Four possible patterns are described below as Cases 1 through 4.
Case 1: Equality of comparative advantage. Equality of comparative advantage holds if \( \sigma_a = \sigma_b \) and \( \rho_{ab} = 1 \). In this case there will be no selectivity bias and the estimated rate of return will be equal to the interest rate. Specifically, recall that equality of comparative advantage holds if \( (\lambda_a, \lambda_b) = e^{A_x}(\hat{I_a}, \hat{I_b}) \) where \( A_i \) is a scalar measure of ability with variance \( \sigma_i^2 \). In this case, \( a_i = \ln I_a + A_i \) and \( b_i = \ln I_b + A_i \). Thus, \( \sigma_a = \sigma_b = \sigma_A \) and \( \rho_{ab} = 1 \) so that the bias terms, \( \alpha_x \) and \( \beta_x \), in (68) are zero and the estimated rate of return is equal to \( r \). Geometrically, this corresponds to a situation in which the rate of return has zero variance across people so that the distribution of \( \rho \) in Figure 10.5 becomes degenerate at \( r \).

In terms of innate ability, the conditions under which equality of comparative advantage holds can arise in two different ways as can be seen by examining (58). Specifically, these conditions hold if either (a) strength and ability are perfectly correlated (i.e. \( \rho_{12} = 1 \)) and \( \alpha_1 + \sigma_2^2 = \beta_1 + \sigma_2^2 \) or (b) strength and ability have identical percentage effects on productivity in each occupation (i.e. \( \alpha_1 = \beta_1 \) and \( \alpha_2 = \beta_2 \)). Put differently, in either of these cases, economic ability will appear to be a one-dimensional factor even though, in the latter case, ability tests in a non-economic context may distinguish distinct components of ability such as strength and intelligence.

Case 2: Positive hierarchical sorting. This case arises when \( \sigma_a / \sigma_b > \rho_{ab} > \sigma_b / \sigma_a \). It is called “positive hierarchical sorting” because those who go to college are drawn from the upper portion of the distribution of potential earnings in A while those who stop at high school are drawn from those in the lower portion of the distribution of potential earnings in B. Note that the parameter values for which this occurs imply that \( \rho_{ab} \) is sufficiently positive and that \( \sigma_a^2 > \sigma_b^2 \).

For example, suppose that \( \rho_{ab} = 1 \) so that there is a perfect correlation between talent in A and in B. Then the least talented person in A will be more talented than the most talented person in B. This extreme case of hierarchical sorting is illustrated in Figure 10.6 where the marginal distributions of \( a_i \) and \( b_i \), respectively, are drawn on the horizontal and vertical axes and their joint distribution is the degenerate bivariate normal whose density lies along line \( dd \) which passes through \( (\mu_a, \mu_b) \) at point \( m \) and has slope \( \sigma_a / \sigma_b \). The index function in (61), rewritten as \( a = r(s_a - s_b) - \ln(w_a / w_b) + b \), is given by line II which intersects \( dd \) at point \( e \). Anyone whose endowment point, \( (a_i, b_i) \), lies above II will achieve higher present value by choosing B and anyone whose endowment is below II will do best to choose A. Since everyone’s endowments lie on \( dd \), it is clear that all individuals whose endowments lie on the segment of \( dd \) below point \( e \) will choose B while all those whose endowments lie on the segment of \( dd \) above point \( e \) will choose A. The shaded areas of the two marginal distributions indicate the hierarchical sorting in labor market equilibrium.

The special case depicted in Figure 10.6 was used by Roy (1951) as a possible explanation for the tendency of the distribution of labor incomes to be more
skewed than a log-normal distribution. That is, note that the overall distribution of log earnings in the diagram consists of a lower part given by the shaded area of the B distribution with variance \( \sigma_b^2 \) and an upper tail given by the A distribution which has a larger variance, \( \sigma_a^2 \), so that the composite distribution is skewed to the right.

More recently, Rosen has used a similar argument to explain the extremely high earnings of “superstars” [Rosen (1981)] and the skewed distribution of managerial salaries at successively higher levels of hierarchically structured firms [Rosen (1982)]. In both cases the argument turns on the assertion that (a) there is a one-factor distribution of ability (i.e. \( \rho_{ab} = 1 \)) and (b) that the “scope” for talent is greater for, say, a major league baseball player than for a minor leaguer or for a corporation president compared to a middle manager (i.e. \( \sigma_a > \sigma_b \)). In terms of the relationship between innate abilities and occupational productivity in (57), greater “scope” for talent in A implies that \( \alpha_1 + \alpha_2 > \beta_1 + \beta_2 \).

**Case 3: Negative hierarchical sorting.** This case corresponds to the condition \( \sigma_a / \sigma_b < \rho_{ab} < \sigma_b / \sigma_a \). The negative hierarchical sorting implied by this condition would hold if there is a high positive correlation between productivities in A and B but with a greater scope for talent in the occupation which requires only a high school degree. Geometrically, the extreme form of this case would be illustrated by relabeling the axes in Figure 10.6. Needless to say, this case does not appear to be empirically important.

**Case 4: Non-hierarchical sorting.** The final possibility is the case of “non-hierarchical sorting” occurs when \( \sigma_a / \sigma_b > \rho_{ab} \) and \( \sigma_b / \sigma_a > \rho_{ab} \). This case will
occur if \( \rho_{ab} \) is sufficiently small or if the scope for talent in A and B is about the same (e.g. if \( \sigma_a = \sigma_b \), then \( \rho_{ab} < 1 \) is sufficient for non-hierarchical sorting). In this case, those who are best at B will tend to go to high school and those who are best at A will tend to go to college.

As an extreme example, suppose that innate abilities are uncorrelated (i.e. \( \rho_{12} = 0 \)) and that only strength is useful in B and only intelligence is useful in A (i.e. \( \alpha_1 = \beta_2 = 0 \)) so that \( \rho_{ab} = 0 \). Geometrically, this case would correspond to a situation in which the degenerate bivariate distribution of \( a \) and \( b \) represented by line \( dd \) in Figure 10.6 is replaced by an elliptical set of iso-probability contours whose major axis is horizontal. The index line, II, partitions this bivariate distribution on a slant so that the probability that individual \( i \)'s comparative advantage is in A is an increasing function of \( a_i \) and the probability that his comparative advantage is in B is an increasing function of \( b_i \). On average, the strongest workers will choose B and the most intelligent workers will choose A. Thus, average productivity of those in B will exceed \( \mu_b \) and the average productivity of those in A will exceed \( \mu_a \).

5.4. Empirical studies of ability bias

A large and complex literature on the question of ability bias has arisen in the wake of the claim that a comparison of earnings of individuals who differ in education can be used to estimate the rate of return to investment in human capital. Since Griliches (1977, 1979) has ably reviewed all but the most recent literature in this field, I will summarize its methodology and findings very briefly in this section.

A major problem in dealing with questions concerning the role of ability and opportunity factors in determining earnings is the difficulty of finding data sets that contain information on ability and family background together with good information on individual education and earnings. To a considerable extent, economists have begun with data collected for other purposes and, in several important cases, they have resurveyed individuals who appear in existing data sets in order to add economic information. Since "opportunism" has been a dominant force in generating the data bases used to study these issues, there are often serious questions concerning the representativeness of a given sample and the comparability of variables across data sets.

Two major types of data have been used. One type provides information on psychometric mental and physical ability tests such as IQ, AFQT, tests of visual acuity and so on. These tests represent direct attempts to measure ability. Initially, such measures tended to be taken at face value and were entered directly into earnings regressions as "controls". More recently, test scores have often been regarded as "indicators" of underlying unobservable "true abilities"
in latent variable models. The other major type of data uses data on siblings (i.e. brothers, dizygotic and monozygotic twins) in order to control for unobservable family effects including genetic and environmental influence which influence ability and/or opportunity. Under certain assumptions, one sibling can in effect be used as a "control" for the unmeasured family effects of the other sibling.

One example of the first type of data is the NBER-Th sample which was based on data on a sample of men who had volunteered for pilot, bombardier, and navigator programs of the Army Air Force during World War II which was originally gathered by the psychologists Robert L. Thorndike and Elizabeth Hagen. These men had taken a battery of tests of mental and physical abilities in the Air Force and were resurveyed in 1955 by Thorndike and Hagen after the war to determine their educations, occupations, and income. Later Paul Taubman and F. Thomas Juster at the National Bureau of Economic Research discovered these data and organized a resurvey of a subset of these men in 1969. The resulting data set provides information on education, income at up to five different points in the life cycle, test scores, and fairly detailed information on various measures of family background such as parental education, father's occupation, mother's work activity, and so on. As is true of most "opportunity" data sets, the individuals in the NBER-Th sample are not representative of the population. For example, they all have at least a high school education and all scored in the upper half of the AFQT ability test. These data have been used in a number of studies of earnings including Taubman and Wales (1974), Hause (1975), Lillard (1977), and Willis and Rosen (1979).

An example of the second type of data is the NRC twins sample which contains information on about 1000 monozygotic (MZ) and 900 dizygotic (DZ) twin pairs based on a National Research Council sample of white male army veterans. This data set also represents another case in which economists have resurveyed a sample which was collected for another purpose—in this case primarily for bio-medical research. Again, Paul Taubman was the economist who initiated the resurvey. More recently, Taubman and his associates have surveyed the children of these twins to study intergenerational issues.

The effect of measured ability on earnings in the NBER-Th data can be seen in Figure 10.7 which is reproduced from a descriptive study by Lillard (1977). Lillard combined individual test scores into one measure using factor analysis. This ability measure was then entered in a fully interactive manner with age and education in a third degree polynomial together with linear family background effects in a least squares earnings regression. The figure shows predicted age-earnings profiles for individuals with 12, 16, and 20 years of schooling who have sample average ability and sample average values of other variables. It also shows the earnings function for those whose measured ability is one standard deviation above or below the sample average.
An important point to note from this diagram is that ability interacts positively with age (or experience). Thus, at early ages the more able earn slightly less than the less able, but by the time peak earnings are reached around age fifty the more able earn significantly more than those with less ability. Put differently, it is clear from the figure that higher levels of both ability and education are associated with higher dollar growth of income. An obvious hypothesis to explain these patterns is that the more able tend to invest more in on-the-job training or that they choose jobs with greater growth potential.

The figure also suggests that ability effects may be substantially understated in studies which rely on data for men under 35. Unfortunately, such an age limitation is characteristic of a number of data sets which have been used to study ability effects. Lillard suggests this as a possible explanation for the small magnitude of the ability effects found by Griliches and Mason (1972) and in the literature surveyed by Jenks et al. (1972). This is also a problem in the more recent series of papers by Chamberlain and Griliches [e.g. Chamberlain and Griliches (1977)] which use data from the National Longitudinal Survey of Young Men.

In the Lillard study and in a number of earlier studies, measured abilities were essentially taken at face value in the sense that they were simply entered as regressors to estimate the effect of schooling net of measured ability. There is, however, an obvious question whether tests really measure "true ability" and, if so, how well. To the extent that the measured ability measures are imperfect or incomplete representations of "true ability" there remains the possibility that the effects of schooling and experience will still be subject to ability bias even when measured ability is controlled. In addition, there are questions concerning the
treatment of schooling as an endogenous choice variable which are not addressed or inadequately addressed in much of this literature.

Following Griliches (1977, 1979), the discussion of studies of ability bias can begin with a simple earnings function:

$$\ln y = \alpha + \beta s + \gamma A + u,$$

(70)

where $\beta$ is regarded as the "true" measure of the rate of return to education and $A$ represents a set of unmeasured variables including ability, family background, or other variables apart from schooling which are thought to influence earnings. For expositional simplicity, experience effects are assumed away. Least squares estimates which omit $A$ will result in a biased estimate of $\beta$. Using the standard formula for omitted variable bias, the expected value of the schooling coefficient is $E\beta_{s} = \beta + \gamma b_{as}$, where $b_{as} = \text{cov}(A_{i}/\text{var}(s))$ measures the association between schooling and the left out variable(s).

The literature surveyed by Griliches (1977, 1979) gives primary emphasis to the treatment of ability as an unobservable. Of the studies surveyed, by far the greatest degree of ability bias was found by Behrman et al. (1977) using data from the NRC twin sample which was described above. Griliches' analysis of their results provides an excellent outline of many of the theoretical and statistical issues that arise in attempts to deal with the question of ability bias. I will attempt to convey the flavor of his analysis in the following brief summary.

Behrman et al. (1977) argue that ability (and other unobservable variables determining economic success such as drive, ambition, etc.) effects may be regarded as the consequence of the genetic and environmental contribution of the family. If it is assumed that the unobserved component, $A_{i}$, of person $i$ is a pure "family effect" which captures these genetic and environmental effects, then data on siblings, especially twin data, may be used to control for these unobservable effects and permit an unbiased estimate of $\beta$.

To illustrate this, assume that the earnings model is given by (70), that schooling is treated as exogenous and that the family effect represented by $A_{i}$ can be decomposed into additive genetic and environmental components as follows:

$$A_{i} = G_{i} + E_{i},$$

(71)

where $G_{i}$ and $E_{i}$ are, respectively, the genetic and environmental components which have variances $\sigma_{G}^{2}$ and $\sigma_{E}^{2}$ and covariance $\sigma_{GE}$.

The basic idea behind using sibling data is that the "within-family" return to schooling can eliminate at least part of the covariance between $A$ and $s$ which exists between random pairs of individuals. Thus, consider the following within-
family model obtained by taking first differences of the earning function in (70):

$$\ln y_i - \ln y_{i'} = \beta (s_i - s_{i'}) + \gamma (A_i - A_{i'}) + u_i - u_{i'}$$

$$= \beta (s_i - s_{i'}) + \gamma (G_i - G_{i'}) + \gamma (E_i - E_{i'}) + u_i - u_{i'},$$ \hspace{1cm} (72)

where $i$ and $i'$ denote a given twin pair for $i$, $i'=1, \ldots, n$ twin pairs and $A_i$ is assumed to have the error component structure given in (71).

When it is applied to MZ twins, the within-family model can completely eliminate ability bias if it is assumed that individuals who grow up in the same family have identical environmental components and that individuals who have identical genes have identical genetic components. Given these assumptions, $E_i - E_{i'}$ and $G_i - G_{i'}$ are both zero for MZ twins. Since cov$(u_i)$ is assumed to be zero, it follows that an estimate of $\beta$ using data on MZ twins will be unbiased. In the case of DZ twins, the environmental component is eliminated by first differencing and the variance of the genetic component, $G_i - G_{i'}$, is considerably smaller than it would be for randomly chosen pairs.

Behrman et al. (1977) found evidence of substantial ability bias when they applied the model in (72) to data from the NRC twins sample. They first estimated $\beta$ using data from random pairs and obtained an estimate of about 8 percent for $\beta$ which is similar to estimates found in representative samples of the U.S. population. If individuals are chosen at random from the population, a least squares estimate of $\beta$ is subject to ability bias due to covariance between $A$ and $s$ as explained above. When data on DZ twin pairs were used, the estimate of $\beta$ fell to about 6 percent and when data on MZ twins were used it fell to only 2.7 percent. They interpreted these results as showing that a major portion of the apparent return to education is due to correlation between schooling and unmeasured family (especially genetic) components.

Griliches (1979) argues that this interpretation is suspect for two major reasons. First, it is not clear a priori that all omitted variables are purely family effects. Once individual-specific components are allowed, he shows that the within-family estimates are not necessarily less biased than estimates for randomly chosen individuals. Second, the effects of statistical problems other than unobserved ability components may be accentuated in the within-family regressions. A major example of this is the possibility that schooling is measured with error. This problem is likely to be minor when the variance of schooling is relatively large as it is in the general population. However, the noise-to-signal ratio and hence the degree of bias due to errors in variables tends to become large in the within-family regressions. Griliches argues that plausible assumptions concerning errors in the schooling measure can explain most of the differences in the estimates of the returns to schooling among random pairs, DZ twins and MZ twins found by Behrman et al.
It is important to point out that the earnings model in (70) treats schooling as exogenous and treats the coefficients $\alpha$, $\beta$, and $\gamma$ as constant across individuals. In contrast, as has been argued at length above, when schooling decisions are endogenous and the condition of equality of comparative advantage holds there is no bias in the least squares estimate of $\beta$ because $\text{cov}(A \Delta s) = 0$. However, if there is interpersonal variation in comparative advantage, there will tend to be correlation between $A$ and $s$ because of self-selection. Moreover, because variation in comparative advantage implies interpersonal variation in the rate of return to education, $\beta$ (and the other parameters) will not be constant across individuals. Thus, $\beta$ may be regarded as an estimate of the sample average of individual-specific marginal rates of return. Since individual rates of return influence schooling decisions, it follows that $u$ will tend to be correlated with $s$ because it contains individual deviations from $\beta$. Hence, even if “true ability” could be observed perfectly, least squares estimation of $\beta$ may be subject to simultaneous equations bias.

Willis and Rosen (1979) use data from the NBER-Th sample in an attempt to deal with some of the problems presented by self-selection and unobserved ability and opportunity components. They formulate an econometric model based on the Roy model which utilizes the distinction between ability and opportunity factors emphasized in Becker’s Woytinsky Lecture. Information on observed ability and opportunity variables is used to correct for selectivity on the unobservables. The model permits them to determine whether ability selection is hierarchical or non-hierarchical. It also provides evidence on the hypothesis that schooling choices are based on maximization of the present value of earnings and provides an estimate of the elasticity of college enrollment with respect to the rate of return to college education.

In the Willis–Rosen model individuals are assumed to choose that level of schooling which maximizes the present value of lifetime earnings. For reasons of computational feasibility, the choice is restricted to two schooling categories. They are labeled A for more than high school and B for high school graduate. (Recall that all members of the NBER-Th sample are at least high school graduates.)

The life cycle earnings profile of each individual, conditional on his schooling choice, is described by two parameters corresponding to an initial level and a constant growth rate of earnings. These “structural earnings functions” are

\[
\ln y_{ai} = X_i \beta_a + u_{1i}, \\
g_{ai} = X_i \gamma_a + u_{2i}, \\
\ln y_{bi} = X_i \beta_b + u_{3i}, \\
g_{bi} = X_i \gamma_b + u_{4i},
\]

(73) (74)
where $y_{ij}(x) = \tilde{y}_{ij}\exp(g_{ij}x)$ is the earnings of person $i$ at experience level $x$ given that he has schooling level $j$ (= A, B). $X_i$ is a vector of observable "ability" variables which affect the individual's initial earnings (i.e. $\tilde{y}_{ij}$) and growth rate of earnings (i.e. $g_{ij}$) given his schooling choice $j$. Earnings growth is calculated as the average real growth rate of earnings between the respondent's first job (in 1946, on average) and his most recent earnings in 1969. The influence of unobservable ability variables on earning potential is captured by $u_{1i}$ through $u_{4i}$.

It is important to note that neither observed nor unobserved ability components necessarily influence earnings potential in the same way in both A and B. Specifically, no restriction is placed on the variances or covariances of the unobservable components, $u_1$ through $u_4$. Similarly, the coefficients of the observable ability variables, $(\beta_a, \gamma_a, \beta_b, \gamma_b)$, may differ between A and B. It follows that the rate of return to college education may vary across individuals because of differences in both observed and unobserved components.

Opportunities to finance educational investment are also assumed to vary across individuals. Each individual is assumed to face a constant interest rate,

$$r_i = Z_i \delta + u_{5i},$$

(75)

where $Z_i$ is a vector of observable "opportunity" variables which influence the individual's rate of interest and $u_{5i}$ reflects unobservable opportunity variables. Since there is no direct data on individual-specific discount rates, $r_i$ is also treated as an unobservable.

The decision rule for college enrollment is obtained by defining the index function,

$$I_i = \ln(V_{ai}/V_{bi}),$$

where $V_{ai} = \frac{\tilde{y}_{ai}/(r_i - g_{ai})}\exp(-r_\tau)$ and $V_{bi} = \frac{\tilde{y}_{bi}/(r_i - g_{bi})}$, respectively, denote person $i$'s present value of lifetime earnings if he chooses to go $s$ years beyond high school or stop at high school graduation. (For simplicity, working life is assumed to be infinite.) Individual $i$ will choose to enroll in college if $I_i > 0$ and otherwise will stop at high school. Using a Taylor Series approximation around the population mean values ($\tilde{\beta}_a, \tilde{\beta}_b, \tilde{\tau}$) yields:

$$I_i = \alpha_0 + \alpha_1(\ln \tilde{y}_{ai} - \ln \tilde{y}_{bi}) + \alpha_2 g_{ai} + \alpha_3 g_{bi} + \alpha_4 r_i,$$

(76)

with $\alpha_1 = 1$, $\alpha_2 = 1/(\tilde{\tau} - \tilde{g}_a) > 0$, and $\alpha_3 = -1/(\tilde{\tau} - \tilde{g}_b) < 0$.

Two key assumptions are required to make this model operational. One concerns the functional form of the joint distribution of the unobservable ability and opportunity components, $u_{1i}$ through $u_{5i}$. These are assumed to be normal with mean zero. No restriction is placed on the variances or covariances of these components.

The second key assumption concerns identification. Ideally, we would like to observe the effect on earnings of a change in education of a person with given
ability. Since ability is not completely observable and since any given person can only choose either A or B, this ideal is unattainable. The best that we can hope for is to observe the average effect of an increase in education in a group of people in which the difference in education is uncorrelated with ability. As I pointed out earlier, this goal can be attained if opportunities (i.e. \( r_i \)) can be varied independently of abilities. Thus, to identify the effect of selection on ability it is necessary for at least one of the opportunity variables, \( Z_i \), to differ from the ability variables, \( X_i \). In addition, if we wish to identify the effect of variation the rate of return to college on college enrollment, it is necessary for at least one \( X \) variable to differ from the \( Z \) variables.

In their paper, Willis and Rosen assume that the \( X \) and \( Z \) variables do not overlap. They associate the \( X \) variables with the battery of test scores taken by the respondents in the military. These include tests of reading, mathematics, mechanical aptitude, and dexterity. The \( Z \) variables are assumed to be a set of family background variables including father’s education and occupation, mother’s work experience, religion, and number of siblings. This identifying restriction emphasizes the importance of the family as a source of finance for higher education and assumes that any direct family effects on ability (e.g. genetic effects or training received within the family) are adequately captured by the test scores. While this restriction is not testable, some indirect empirical support for it is suggested by the preponderance of evidence from other studies that family background variables appear to have little direct effect on earnings, but operate primarily through their influence on schooling attainment [see Griliches (1979)]. On the other hand, the fact that the respondents were eligible for G.I. Bill educational subsidies may undercut its plausibility for the NBER-Th sample.

Willis and Rosen’s estimation strategy, based on an econometric model by Lee (1977), involves three steps. First, they estimate a “reduced form probit” equation which describes the probability that an individual with observed characteristics given by \((X_i, Z_i)\) will choose to go beyond high school. This equation is used to form an estimate of the inverse Mills ratios, \( \lambda_{ai} \) and \( \lambda_{bi} \) [which were defined in (65)], for each individual in the sample. Second, the estimated values of \( \lambda_{bi} \) and \( \lambda_{ai} \), respectively, are entered as regressors along with \( X_i \) into the equations for the initial level and growth rate of earnings for in (73) and (74) to obtain an estimate of \( \beta_{da} \) and \( \beta_{db} \) which are corrected for selectivity bias. [See Heckman (1976) for the justification for this procedure.] As explained earlier, the estimated coefficients of \( \lambda_{ai} \) and \( \lambda_{bi} \) provide evidence on the nature of the selectivity of

---

13Actually, it is theoretically possible to correct for selectivity bias even if the \( X \) and \( Z \) variables are identical by using the fact that inverse Mills ratios are non-linear functions of these variables while the earnings parameters in (73) and (74) are assumed to be linear functions of \( X \). However, the use of non-linearities for identification is perilous because it places very heavy reliance both on correct specification of the functional form of the unobservables and on correct specification of the structural regression.
Ch. 10: Wage Determinants

schooling on unobserved ability components. Third, estimates of $\ln \tilde{y}_{ai}$, $\ln \tilde{y}_{bi}$, $\tilde{g}_a$, and $\tilde{g}_b$ are entered into a "structural probit" equation along with $Z_i$ to obtain consistent estimates of the coefficients $\alpha_1$, $\alpha_2$, and $\alpha_3$ in (76). Using these estimates together with estimates of $\tilde{g}_a$ and $\tilde{g}_b$ from sample mean growth rates, it is possible to obtain an estimate of $\tilde{r}$, the mean rate of interest in the population.

Their empirical findings indicate significant non-hierarchical selection on ability equivalent to Case 4 described in Section 5.3. That is to say, individuals who enroll in college have higher lifetime earnings in A than those who did not enroll, while those who did not enroll have higher lifetime earnings in B than the enrollees would have had if they had chosen B. The most important measured ability factor appears to be the mathematical test score which significantly increases lifetime earnings of the college educated but has little effect on the high school graduates.

From the earlier discussion of the Roy model, recall that non-hierarchical sorting implies that there is more than one distinct ability factor and that the direction of ability bias is uncertain. For the average sample member, the estimated uncorrected rate of return to college of 9 percent is lower than the corrected rate of return of 9.8 percent. On average, those who attended college had a rate of return of 9.9 percent while those who did not attend college had a return of 9.3 percent. Estimates of the average rate of discount (i.e. $\tilde{r}$) range between 9.8 and 12.4 percent. Finally, at the sample mean, the estimates imply that a one percent increase in the permanent earnings of college-educated workers relative to high school graduates would increase the probability of college enrollment by about 2 percent. Alternatively, assuming four years of college, this calculation implies that an increase in the rate of return to college by one percentage point would increase college enrollment by 8 percent. This is a sizeable elasticity but it is, of course, smaller than the perfectly elastic response that would be expected under conditions of equality of opportunity and comparative advantage.

In a similar model applied to data from Project Talent, Kenny et al. (1979) find evidence of selectivity on the earnings of high school graduates but no evidence of selectivity for the college educated. However, their data do not permit estimation of corrected rates of return to college or estimation of a structural college enrollment function.

Unfortunately, it is not certain how well either the qualitative or quantitative findings reported by Willis and Rosen generalize because, apart from Kenny et al., this type of model has not been estimated with other data sets. The main reason for this is probably the scarcity of data sets that contain sufficient ability and opportunity measures and also have earnings data covering a large portion of the life cycle.

Given the complexity of the issues and the non-representative character of the data sets that have been employed in the literature on ability bias, it is difficult to reach any firm conclusions about the magnitude or even the direction of the bias...
in U.S. data and there seem to be few, if any, studies using non-U.S. data. My impression is that the simple Mincer-type earnings function does a surprisingly good job of estimating the returns to education even though more general econometric models suggest that the conditions of equality of opportunity and equality of comparative advantage upon which it is based are not strictly true.

6. Other topics

Throughout the chapter to this point, it has been assumed that investment in human capital raises worker productivity and that the market wage received by a worker at any point in the life cycle is equal to the value of his current marginal productivity. Both of these assumptions have been questioned in recent literature and a survey of the determinants of earnings would be seriously incomplete without mentioning them. Given the length of this survey, however, I will not be able to do them full justice with a detailed treatment.

6.1. Education and economic growth

The central question which provided the original impetus for the modern development of human capital theory was not the issue of earnings differentials by age and education which has dominated both this survey and the human capital literature since 1960. Rather, as is clear from Schultz’s (1961) Presidential Address to the American Economic Association, the central question concerned the extent to which growth in the average quality of labor over time resulting from investment in human capital could help to account for the “residual” in U.S. productivity growth which growth in conventionally measured inputs of capital and labor in early studies by Solow (1957) and others left unexplained.

The answer to this question given in the “growth accounting” literature [see, for example, Dennison (1962) and Griliches (1970)] was that a considerable fraction of the residual could be explained by investment in human capital. A simple example of the basic methodology of these studies goes as follows. Assume that the relative productivity of workers in a base period who differ in education is given by their relative wages. The growth of “quality corrected” aggregate labor input in another period can then be calculated as the weighted sum of the man-hours of labor contributed by that group where the weight is given by relative base period wage of each education group. Since there has been substantial growth in the average educational attainment of the U.S. labor force over time, the growth of “quality-adjusted” aggregate labor input is more rapid than the growth of the unadjusted aggregate and the overall unexplained residual in productivity growth is reduced.
A basic assumption underlying this methodology is that an increase in educational attainment causes an increase in labor productivity. From the inception of human capital theory, a variety of critics have expressed skepticism about this causal assumption. Most frequently the skeptics, especially non-economists, argued that the higher pay received by the more educated reflects the operation of “credentialism” rather than higher productivity. This line of criticism is unpersuasive to economists who ask why profit-seeking firms would choose to sacrifice profits by paying wage premia merely for “sheepskins”.

In his well-known “signalling” model, Spence (1974) showed that profit-seeking firms may indeed pay wage premia to more educated individuals even if education has no effect on productivity [see also Arrow (1973)]. In addition, he shows that wealth-maximizing individuals will be willing to make educational investments because of these wage gains. Thus, he argues, it is possible for there to be a market equilibrium in which more educated workers receive higher pay even if education has no effect on worker productivity. The basis for this surprising result is the assumption that information about worker productivity is distributed asymmetrically (i.e. workers know their own productivity but firms cannot tell which workers are most productive) and that more able workers can invest in a signal more cheaply than the less able.

A simple numerical example conveys the nature of this argument [Spence (1973)]. Imagine that there are two types of workers who differ in productivity and that there are equal numbers of each type in the population. High ability workers have a marginal product of 2 and low ability workers have a marginal product of 1. Workers may choose either zero or s years of schooling at a positive cost, but schooling does not augment their productivity. With perfect information about worker productivity, employers would pay high and low ability workers wages of 2 and 1, respectively, and no worker would invest in schooling.

Now suppose that firms have no knowledge of worker productivity but that workers know their own type. If firms offer a wage of 2 to high ability workers, low ability workers have an obvious incentive to misrepresent themselves. One possible equilibrium is that firms pay the expected marginal product of a randomly chosen worker (i.e. one-half). An alternative is that high ability workers might choose to use education as a “signal” to employers of their innately higher productivity. Schooling will be a credible signal if, in fact, high ability individuals choose to invest in schooling while those with low ability do not.

This type of self-selection can occur if the cost of schooling is negatively correlated with ability. To justify this assumption, it might be assumed, for example, that high ability people can get through school with less effort than low ability people. Thus, consider the following signalling equilibrium. Assume that the direct cost of investment in schooling to high ability people is 0.5 and to low
ability people is 1.5 and that firms offer a wage of 2 to educated workers and a wage of 1 to uneducated workers. The net benefit from education is equal to the wage to educated workers minus the opportunity wage to an uneducated worker minus the direct cost of education. Hence, the net benefit to education for a high ability worker is 0.5 and is \( -0.5 \) for a low ability worker. In this situation, workers self-select themselves into educational categories according to ability and firms find that the high pay given to educated workers is justified by their higher productivity. Once established, the equilibrium tends to be self-fulfilling.

The theoretical literature on signalling and screening models and aspects of the effect of asymmetric information in the labor market (and other markets) has grown rapidly since the early work of Spence and Arrow. Some of this work considers questions about the sensitivity of the results to alternative equilibrium concepts [see, for example, Riley (1975, 1979)]. The signalling model also raises questions about the efficiency of market determination of educational investment. In the simple example given above, there is clearly overinvestment in education from a social point of view since aggregate labor productivity is unaffected by education but net output is reduced by the cost of education as compared with a situation in which no education takes place. Efficiency issues become more complicated if the allocation of high and low ability workers matters. For example, suppose that low ability workers are completely unproductive if they are assigned to a "high ability" job. In this case, it is socially worthwhile to spend some resources to assign workers to the job in which they are most productive. To the extent that education plays this role, it has social productivity. [See, for example, Stiglitz (1975a).]

Signalling theories have attracted skeptics, too. For example, Becker (1975) points out that a college education is a very expensive test instrument and that it is likely that firms can find cheaper ways to determine worker ability. For example, a number of economists have explored the possibility that labor contracts can be structured in such a way that workers will self-select themselves according to their ability. If such schemes are feasible, the need for investment in a signal is eliminated. In the simple example given above, a piece rate system would suffice if worker productivity is known to the firm ex post. In more complicated situations, no first best labor contract may be feasible.

Ideally, the direction of causation between investment in education and worker productivity would be determined by empirical test. However, the signalling–human capital debate provides an extreme (and clear-cut) example of the tendency for efficient behavior to censor non-experimental economic data in such a way that information crucial to the test is removed.

The basic point is easily illustrated by considering a simple "human capital" example which parallels Spence's signalling example described above [see Spence (1981)]. In the human capital example, education causes an increase in the productivity of all workers by the same amount from a marginal product of 1
with no schooling to 2 with \( s \) years of schooling. The direct cost of schooling is 0.5 to high ability workers and 1.5 to low ability workers. As before, in the signalling example, high and low ability workers, respectively, have marginal products of 2 and 1 regardless of their schooling level and their direct costs of schooling are respectively 0.5 and 1.5. Individual ability is assumed to be unobservable (ex ante) to either firms or to the econometrician. Ex post, the average productivity of a group of workers can be observed.

Given experimental data, determination of which model is correct is perfectly straightforward. The experimenter simply randomly assigns individuals to different schooling levels and observes their average productivity ex post. If the signalling theory is correct, he would expect to find that workers in both schooling groups have an average productivity of 1.5 and, if schooling is productive, he would expect to find the average product of the uneducated and educated groups of workers to be 1 and 2, respectively. Now suppose that the econometrician must rely on market data on schooling and earnings. Note that the equilibrium distribution of workers and level of wages by educational level in the human capital and signalling examples are identical. That is, in both models low ability individuals choose zero schooling and receive a wage of 1 and high ability individuals choose \( s \) years of schooling and receive a wage of 2.

The identification issue illustrated by this example appears to be generic to tests of signalling versus human capital interpretations of educational investment. Note that any information on individual ability which is available to an econometrician is also likely to be available to firms and, therefore, would not be related to the "unobservable" ability components for which education is a signal. The issue is further complicated by the fact that the observed effect of schooling on earnings may consist of both productivity and signalling components. Thus, the empirical problem is to determine the relative importance of these components. As a consequence of these difficulties, the empirical literature on this issue is neither large nor very persuasive. [See Riley (1979) for a review and critique of earlier empirical studies.]

The main empirical tactic of any promise rests on an attempt to classify occupations in terms of an a priori view about the degree to which individual-level productivity in those occupations is observable. For example, Wolpin (1977) argues that screening by education is less important for the self-employed than for employees and then examines differences between the levels of educational attainment and the effect of schooling on earnings for the two groups. He argues that his results do not support the signalling hypothesis, but Riley (1979) argues that Wolpin's results do provide mild support for signalling if they are properly interpreted. In his own empirical work, Riley (1979) chooses a strategy which lets the data "speak for themselves" by classifying occupations into relatively screened and unscreened categories on the basis of occupation-specific earnings functions. Then some additional differences between the two groups are used as tests of the
importance of screening. Riley reports some support for the screening view, but emphasizes the tentativeness of his conclusions.

6.3. Specific human capital

Although the important distinction between "general" and "firm-specific" investment in human capital was introduced very early in the development of human capital theory by Becker (1962, 1964) and Oi (1962), most subsequent theoretical and empirical work in the field tended to ignore the issues raised by specific capital until quite recently. The distinction between the two types of investment is simple. Purely general training received by a worker within a given firm is defined as investment which raises the potential productivity of the worker in other firms by as much as it is raised within the firm providing the training. Purely specific training raises the worker's productivity within the firm providing the training, but leaves his productivity unaffected in other firms.

As explained earlier, competition implies that workers rather than firms will tend to pay the costs and receive the returns from any general training they receive. In effect, general capital is completely embodied in the worker. Consequently, it is efficient for the worker to "own" his general capital and be free to employ it wherever it receives the highest reward. In contrast, the productivity of specific capital is jointly dependent on the productive characteristics embodied in the worker and the characteristics of other firm-specific inputs.

In this case, Becker (1962, 1964) points out that it may be inefficient for either the worker or the firm to have exclusive ownership of specific human capital. For example, if the worker pays the full cost of training and attempts to reap the full returns, the firm may inflict a capital loss on the worker by dismissing him without suffering any loss itself. Symmetrically, if the firm pays the cost of specific training and attempts to reap the returns by paying the worker his opportunity wage, the worker can without cost inflict a capital loss on the firm by quitting to work elsewhere. Becker suggested that the solution to the problem of jointness is for the worker and the firm to share both the costs and the returns so that each agent would suffer a loss if the worker–firm relationship is terminated. However, Becker was unable to provide a theory of the factors that determine the worker's and the firm's shares. Without such a theory, it is not possible to derive implications for the life cycle pattern of worker earnings. Since the theory of general training does produce such implications, it tended to provide the theoretical underpinnings for empirical studies of earnings discussed earlier in this chapter.

One of the major implications of the specific capital concept is to emphasize the importance of the duration of a worker–firm match in determining the total pay-off to the investment. Recently, this has led to renewed interest in developing
theories and methods to measure the duration of jobs and to assess the determinants and implications of life cycle labor force mobility. For example, Hall (1982) has shown that “lifetime jobs” are of more importance in the United States than had commonly been believed. In a still more recent study, Randolph (1983) estimates that a typical U.S. worker has about a 50 percent chance of having a job that lasts more than half of the length of his total career in the labor force. However, he finds that the expected duration of a given job is only about 3 years because of a high exit probability in early phases of a job.

An obvious first question to be raised about specific training is how important it is empirically. Unfortunately, specific capital is no more directly observable than is general human capital. Thus, answers to this question tend either to involve classification of various types of training expense according to a priori notions about their degree of specificity or to attempt to extend the theory in order to obtain indirect evidence by testing its implications for observable behavior.

Most of the hypothesized examples of specific capital that I have run across appear to involve issues of imperfect information rather than the task-specific skills which are often conjured up when describing what human capital “really is” because the technological “know-how” involved in task-specific skills is unlikely to be unique to a given firm. For instance, possible examples of firm-specific capital include a salesman’s knowledge about the characteristics and needs of the firm’s clients, a middle-manager’s knowledge of the firm’s operating procedures, the identity of other employees who know how to fill out a given form and so on.

As another example, the costs of hiring a worker are often treated as a specific investment by the firm. Many of these costs arise because it is costly to inform potential workers of the availability of a job and also costly to screen applicants for their suitability. Similarly, many aspects of search costs incurred by workers seeking jobs can be viewed as firm specific. The concept of information about the quality of a job match as a form of specific human capital has been exploited in a theoretical model of job matching by Jovanovic (1979) which focuses on the implications of the matching process for job turnover.

Jovanovic’s model assumes that the joint productivity of a given match between a firm and a worker is not known at the time of hiring by either the worker or the firm. Rather, the quality of the match is gradually revealed by the worker’s productivity record on the job. As information begins to accumulate in the early phases of the job, poorly matched workers learn this fact and tend to quit in order to search for a better match. Initially, the probability (or hazard) of quitting tends to rise with tenure on the job. As time passes, however, the remaining workers tend to be those for whom the quality of the match is high relative to the expected value of alternatives and the probability of quitting tends to decrease with increased tenure. This pattern of hazard rates has been con-
firmed empirically by Randolph (1983) who finds that the hazard rate increases for about the first 12 months of job tenure and decreases thereafter.

In addition to exploring the empirical implications of Jovanovic's model for job turnover, Mincer and Jovanovic (1981) also examine its implications for earnings. If specific capital is important, the theory suggests that increases in job tenure, holding labor force experience constant, should have a positive effect on earnings. They find significant tenure effects which indicate that about one-third of wage growth in the early portion of the career and 20–25 percent in mid-career can be attributed to specific investment with the remainder due to general investment. [See also Bartel and Borjas (1981).] Recently, Hashimoto and Raisian (1984) have used a similar approach in an attempt to determine the relative importance of specific training in the United States and Japan. They provide evidence that expected job tenure is longer than in the United States and that the tenure-related component of wage growth in Japan tends to be relatively larger.

6.4. Agency theories of life cycle wages

Conventionally, it is assumed that a worker's productivity is the cause of the economic reward he receives. In an important paper, Alchian and Demsetz (1972) argue that the reverse line of causation may be equally important because of the effect of the system of compensation on worker incentives. The basic incentive problem arises because the self-interest of workers is not coincident with the interests of the firm. For example, the owners of the firm value the worker's output but do not have any direct preferences concerning the disutility of effort the worker experiences in producing that output. Conversely, the worker has no direct preferences for the output of the firm. Rather, he is concerned only with his own income and effort.

In the language of agency theory [Ross (1973)], the worker is an "agent" of the firm which is the "principal".\(^\text{14}\) The principal's problem is to design the organization of production and system of rewards (and penalties) in such a way as to make the worker's behavior coincide with the principal's objectives subject to the constraint that the worker receives a level of utility at least as great as he could receive in his next best alternative. A large literature on "agency" theories of wage determination has arisen in the past decade. I shall only briefly describe some of the elements of this literature with special emphasis given to its implications for life cycle wage patterns. (See Chapter 14 by Parsons in this Handbook for more details on this class of problems.)

\(^{14}\) However, see Carmichael (1983) for an interesting model in which the principal himself becomes an agent after the reward system has been agreed upon.
If a given worker's productivity is independent of other inputs (including the effort of other workers) and his output is easily monitored by the firm, Alchian and Demsetz point out that a piece rate system induces an efficient level of effort by the worker. However, if it is difficult to observe output or if there is "team production" (i.e., interactions in production) the worker's product cannot be used as the basis of rewards. The alternative of simply paying an hourly wage provides no incentive for the worker to expend effort. To avoid shirking, the firm may expend resources in attempt to monitor the worker's effort.

As a way of reducing monitoring costs, the firm may attempt to design a compensation scheme which reduces the worker's incentive to shirk (or to engage in other misfeasances or malfeasances such as stealing). This approach was taken by Becker and Stigler (1974) and elaborated by Lazear (1979) who used it in an attempt to explain the phenomenon of mandatory retirement. (See also Chapter 5 by Lazear in this Handbook.)

Very briefly, the argument is that workers can be induced to behave "honestly" (e.g., in accord with an ex ante implicit or explicit contract specifying the level of effort) if some portion of payment to the worker is deferred and the employer follows the practice of dismissing the worker if he is discovered to violate the terms of the contract. In effect, the deferred payment acts as a performance bond because the worker loses the value of the deferred payment if he is dismissed. If the value of this loss is sufficiently high at each point in time during the period of the contract, the worker will be deterred from shirking.

Competition will ensure that the present value of a worker's productivity and the payments made to him over the period of his employment are equal. Hence, a deferred payment scheme implies that the worker will be paid less than his marginal product during the initial phases of the job and more than his marginal product later on. Given that he is being paid more than he is worth, the senior worker would prefer to continue working beyond the ex ante optimal duration of the job but the firm would lose profits if he were to do so. Thus, Lazear argues that mandatory retirement can be regarded as a contractual mechanism by which the firm enforces the optimal duration of the employment relationship.

From the viewpoint of the theory of life cycle earnings, an important implication of deferred compensation schemes is that they break the close link between the evolution of productivity and earnings which is a feature of investment in (general) human capital. The question of the relationship between productivity and wage growth is addressed by Medoff and Abraham (1980, 1981). They examine evidence based on the relationship between a worker's wages and evaluations of his performance by supervisors. They interpret this evidence as suggesting that productivity and pay are not as closely linked as is suggested by

---

15 See Stiglitz (1975b) for an analysis of the trade-off between piece rates and time rates when workers are risk averse.
conventional human capital theory. In contrast, Brown (1983) finds evidence in favor of the human capital theory. He uses data from the Panel Study of Income Dynamics to examine the relationship between the importance of on-the-job training on the current job and wage growth. His results suggest that wage growth tends to be attenuated when training becomes unimportant. A clear-cut resolution of this question awaits future research.

7. Conclusion

A combination of advances in economic theory, collection of new data, and creation of new statistical and econometric techniques has been the hallmark of the development of modern labor economics. Nowhere, in my view, has this combination been more fruitful than in the analysis of the determinants of earnings. In the main, the initial insights of Becker and Mincer who first developed human capital theory have been repeatedly confirmed with data from around the world. Indeed, the reinterpretation of their theory offered in this chapter tends to strengthen this assessment. Moreover, the empirical findings have stood up remarkably well to the possibility that the return to investment in human capital is the result of innate ability differentials rather than compensation for the cost of adjustment.

The recent stress on the role of specific as opposed to general human capital and the development of agency theories of the employee–employer relationship may result in the modification of some of the received doctrine, but these theories also serve to enrich the scope of the theory by pointing toward interesting and potentially important connections between wages, job mobility, and institutional practices. Future progress in this area will hinge crucially on the development of data which links information on the individual characteristics of workers and their households with data on the firms who employ them. I see no comparable promise that the signalling hypothesis will receive a convincing test against the conventional human capital theory because of the inherent identification problem described earlier.

I believe that an important and promising area of future research lies in the further exploration of the general equilibrium interaction of the supply and demand for human capital which has begun with the recent studies of cohort size effects discussed earlier. Theoretical considerations suggest that there may be important interrelationships between changes in the age distribution due to variations in population growth, changes in the age structure of life cycle productivity due to human capital investment and the equilibrium interest rate. In addition, the underlying influence of the family and the government on the supply and demand for human capital need to be considered. Some initial explorations of these interrelationships are presented in Willis (forthcoming) in a
steady-state overlapping generations model, but theoretical work on non-steady-state problems and most of the relevant empirical work awaits future research. Also, the potential for increased knowledge from international comparative studies is great. The discovery and development of new data, especially micro data, will be the key to progress in this area.

References


Ch. 10: Wage Determinants

Krueger, Ann O. (1972) "Rates of return to Turkish higher education", *Journal of Human Resources*.
Murphy, Kevin, Mark Plant and Finis Welch (1983) "Cohort size and earnings", paper presented to the IUSSP/IASSA seminar on *Economic consequences of population composition in developed countries*, Vienna.
State University of New York at Stony Brook.


