Human Capital and the Life Cycle
By Mark R. Killingsworth*

In what follows I develop a life-cycle model of labor supply and human capital accumulation. The model assumes that the individual maximizes lifetime utility by allocating time to work, leisure and human capital formation; and the model allows for two different kinds of human capital -- "training" and "experience" (i.e., "learning by doing"). The model thus effects a synthesis of diverging viewpoints in the analysis of labor supply in two important respects. First, it offers a unified treatment of the relationship between wage rates, hours of work and investment in human capital; and second, it provides a comprehensive treatment of the nature and significance of "human capital" itself.

*Assistant Professor of Economics, Fisk University, and Visiting Research Associate, Industrial Relations Section, Princeton University. I thank Alan S. Blinder for numerous helpful discussions of many of the ideas here; my debt to the paper by him and Yoram Weiss is evident throughout what follows. My debt to Phyllis Ventigli, who typed this manuscript, is likewise substantial.
Most previous work has not treated the relationship between wage rates, hours of work and investment in human capital in a unified way. In one view, persons take the wage as given and maximize utility by allocating time between leisure and work; in another, persons take leisure time as given and maximize income by allocating time between work and investment in human capital. In Paula Stephan's words (p. 35):

[One traditional] view sees the wage rate as exogenous and hours of work as endogenous...The human capital view sees the combined number of hours of work and training as exogenous and the wage rate as endogenous...Both are, of course, partial views of the world. Yet these partial views are seldom joined. As a profession we teach our students that the hours-of-leisure variable is solved in a utility-maximization framework while the wage rate variable is solved by the technique of lifetime income maximization.

In contrast, the present paper treats the wage rate, hours of work and hours of leisure as endogenous; all three variables are determined jointly in a process of utility maximization.

Of course, in this respect the present paper is not unique; several other models also take the same approach, e.g., those of Gilbert Chez and Gary S. Becker and of Alan Blinder and Yoram Weiss. But while the following pages are not unique in their emphasis on wage rates, hours of work and hours of leisure as variables which are endogenously determined in a process of utility maximization, they are, however, useful in carrying work in this area further than has previously been done, since they also develop propositions based on it which are suitable for detailed
empirical investigation. 1/

A second respect in which the present paper synthesizes diverging approaches to the analysis of the supply of labor is its emphasis on a comprehensive treatment of the nature and significance of human capital itself. All previous models of "human capital" have in effect assumed that human capital is either all "training" or all "experience" (i.e., "learning by doing"). 2/

1/ Chez and Becker are largely content with stating first-order conditions for a maximum, and devote rather little effort to rigorous derivation of what these conditions imply about life-cycle behavior (though there is a great deal of intuitive discussion of life-cycle phenomena); and while Blinder and Weiss examine the equilibrium-dynamics properties of their model (i.e., the time paths of the endogenous variables for given values of the exogenous variables) in a much more careful and thorough fashion than do Chez and Becker, they do not attempt to study the comparative-dynamics properties of their model (i.e., the ways in which the time paths of the endogenous variables may change in response to changes in the values of the exogenous variables). Moreover, empirical work on such models is truly in its infancy; the Blinder-Weiss paper is entirely theoretical in orientation, and the empirical work of Chez and Becker is in general somewhat rudimentary and in many respects rather unsatisfactory. (For example, their econometric work is usually based on moving averages for "synthetic cohorts," i.e., moving averages by age group as obtained from US Census samples. Thus, data for a "30-year old" in their study are actually averages over persons aged 29-31; data for a "31-year old" are in fact averages for persons 30-32; etc. It is well-known that OLS estimates of regression parameters will be inefficient when based on data of this kind.) James Heckman's recent empirical work (1975)(forthcoming) is econometrically more sophisticated, but is restricted to the case where a "fully interior" solution (i.e., a situation in which leisure, work and investment time are all positive) is assumed to obtain.

2/ "Training" models include those of Chez and Becker and Blinder and Weiss; "experience" models include those of Malcolm R. Fisher and Weiss. Investment in human capital of the "training" type occurs when time is devoted specifically to this purpose (e.g., to schooling or to formal or informal on-the-job training (OJT)); on the other hand, investment in human capital of the "experience" or "learning by doing" type occurs as a mere by-product of time spent actually working, and not as a result of time devoted to other purposes. Thus training can occur during time spent at work (but not "working"), while experience is acquired during time spent actually working. (For a more precise definition of these two terms, see below.)
Now, Becker (and, following him, Jacob Mincer (1974)) appear at times to take the view that the distinction between experience and training, at least as usually stated, is false. As usually stated, the distinction is made more or less as follows: the acquisition of training occurs as a result of the diversion of time from actual work to training (which may often occur on the job), and thus must involve a sacrifice of current income; while the acquisition of experience occurs as a mere by-product of time spent at work, and thus cannot involve any sacrifice of current income.  

But this cannot be a meaningful distinction, say Becker and Mincer; for example, as Mincer says, "even if apparently costless differential 'learning by doing' opportunities exist among jobs, competition tends to equalize the net returns, thereby imposing opportunity costs on such learning" (1974, p. 132; see also 1974, p. 65, and Becker, pp. 53-55, for similar arguments). However, while the importance of this point is clear, it is also clear that once individuals do use net returns (however defined) to choose jobs, they begin accumulating experience and hence human capital as they work, and thus, if experience does affect productivity, that their wages are in part a function of their accumulated work time. In other words, while it is certainly true that -- as Becker and Mincer stress -- there are indeed ex ante opportunity costs associated with the acquisition of experience

\[\text{If "training" is "general" (i.e., raises productivity equally in all firms in which it is used), hours spent training will result in lower current income (but, presumably, higher future income) for the individual (see Becker, pp. 8-29); while time spent gaining "experience" (i.e., time spent actually working) will result in higher current income (and, presumably, higher future income also). For example, a machinist who devotes time to an OJT program is receiving "training," i.e., reduces current income in order to earn still more income later; while a machinist who devotes time to work may be gaining "experience," i.e., receives some income now and will (by virtue of his experience) receive still more income later.}\]
(i.e., costs associated with choosing one job with one kind of opportunity for "learning by doing" rather than another), there are no ex post costs (i.e., costs associated with acquiring experience within a given job).

The fundamental distinction remains: once a job has been chosen, experience accumulates as a mere by-product of time spent at work.

To be sure, training and experience are not wholly different: one can raise one's wage by accumulating more of either kind of human capital; and, as just explained, the acquisition of either does involve a sacrifice. But while one can accumulate more training without changing leisure time (one can simply spend less time actually working and more time getting training, while leaving unchanged the total amount of time spent at work), one cannot accumulate more experience without changing leisure time (i.e., one must spend more time at work). Thus, there does seem to be a real difference between the two kinds of human capital; and the present paper both recognizes this difference and attempts to allow for it in analyzing life-cycle behavior. To repeat: previous work has assumed that all human capital is either training or experience; the present paper recognizes that both training and experience constitute forms of human capital, and treats the effects of both kinds.

The rest of the paper proceeds as follows. I first outline the model and the concepts embedded in it, and summarize its major equilibrium-dynamics implications; in Part II, I describe the model in more detail and present rigorous demonstrations of the assertions which appear in Part I. I then discuss econometric issues in Part III.
I. Outline of the Model and its Equilibrium-Dynamics

All models abstract from reality; that is what makes them models. The closer one attempts to get to "reality," the more complicated one's model gets; and, conversely, the simpler one's model is, the less "realistic" it will be. Thus, in constructing a model, one must steer a difficult course between false simplicity and unmanageable complexity. In the present case, my objective is to explain much with little -- to see how a relatively simple model embodying only a few easily-grasped hypotheses can be used to account for, and predict, the life-cycle behavior of a number of variables, such as labor supply, wages and earnings. In many cases the assumptions are quite specific (e.g., that the "output" of training is a linear function of the amount of time devoted to training). I therefore stress at the outset that in many cases more general assumptions lead to conclusions which -- while harder to arrive at -- are essentially the same as those which emerge from the much more restrictive assumptions adopted here. That is, one frequently finds that the difference between models of the kind considered here and more general (and hence more complicated) models is simply that the latter are much more difficult to manipulate. Of course, in other cases one finds that models which are more general than the kind developed here admit of a greater variety of patterns of life-cycle behavior, i.e., do indeed have different (or at least richer) implications. But if the purpose of models is to explain observed events, a simple model which implies nothing but what has been observed (e.g., an inverted-U-shaped age-earnings profile) may be preferable to a much more complicated model which admits of other possible behavior patterns which may never have been observed empirically.
(e.g., a sinusoidal age-earnings profile). Indeed, as will be evident in what follows, I have chosen to construct the former kind of model rather than the latter.

With this in mind, I now turn to the model itself.

In what follows, I shall assume that the individual's objective is to maximize $F$, the value of his stream of utility from the present ($t = 0$) to death ($t = T$), where

$$ F = \int_0^T [U(C(t)) + V(L(t))]e^{-\beta t} dt + B(A(T)) $$

(1)

$B(A(T))$ is the utility of his estate (i.e., of the value of $A(T)$, his assets at time $T$), $U(C(t))$ is the utility at time $t$ of the consumption of goods $C$, and $V(L(t))$ is the utility at time $t$ of $L$, the fraction of total time available which is devoted to leisure.

I shall assume (a) that marginal utility is positive and diminishing in $C$, $L$, and $A_T$, -- i.e., that, for all $t$,

$$ U'(C(t)) > 0, \quad U''(C(t)) < 0 $$
$$ V'(L(t)) > 0, \quad V''(L(t)) < 0 $$
$$ B'(A(T)) > 0, \quad B''(A(T)) < 0 $$

(2-1)

and (b) that the marginal utility of $A(T)$, $C$ or $L$ becomes infinite as the amount of $A(T)$, $C$ or $L$ approaches zero (i.e., that positive consumption and positive leisure become infinitely precious as their amount gets very small) and that a positive estate is, similarly, regarded as infinitely
desirable by the individual). Hence

\[ \lim_{A(T) \to 0} B'(A(T)) = \infty; \lim_{C(t) \to 0} U'(C(t)) = \infty; \lim_{L(t) \to 0} V'(L(t)) = \infty \quad (2-2) \]

Note that utility is separable both in time (so that, apart from the fact that it is enjoyed one period later, a given bundle of C and L received at \( t+1 \) would provide exactly the same utility as the same bundle received at t) and in its arguments (so that the marginal utility of C(t) depends only on the amount of C(t), and similarly for the marginal utility of L(t) and of A(T)). Note also that assumptions (2-2) ensure that the individual will always have positive A(T), C(t) and L(t) \( (0 \leq t \leq T) \). Finally, the individual's subjective rate of discount, which converts future utilities into present utilities, is \( \rho \).

Next, I make a number of assumptions and definitional statements about the allocation of time. By definition, any time not devoted to leisure is "market time," i.e., time which is used with a view to receiving income either now or in the future. Thus,

\[ L(t) + H(t) = 1, \text{ with } 0 \leq H(t) \leq 1 \text{ and } 0 \leq L(t) \leq 1 \text{ for all } t \quad (3) \]

where \( H(t) \) is the fraction of total time available which is "market time."

In turn, I assume that \( H(t) \) may in turn be allocated either to training or to work; and thus -- where \( F(t) \) is the fraction of \( H(t) \) which is devoted to training at time \( t \), with \( 0 \leq F(t) \leq 1 \) for all \( t \) -- that the fraction of total time devoted to training is \( F(t)H(t) \), while the fraction of total market time spent actually working is \( (1-F(t))H(t) \). When \( F(t) = 0 \), all market time is devoted to working; the individual "specializes" in working and
receives no training. On the other hand, when \(0 < F(t) < 1\), the individual devotes some of his market time to work and some to training; in other words -- following Sherwin Rosen's notion of firms as entities which offer opportunities both for working and for acquiring training (see Rosen, 1972 and 1975) -- I shall say that when \(0 < F(t) < 1\) the individual is (usually) receiving on-the-job training, and is learning as well as earning. (However, note that \(0 < F(t) < 1\) need not necessarily signify that the individual is receiving on-the-job training; it might simply mean that the individual is engaged in part-time schooling and part-time (pure) work. But this is a complication which I shall not pursue here.) Finally, when \(F(t) = 1\), all market time is devoted to training; the individual "specializes" in training and receives no current income. In this case, I shall say that the individual is in school -- in keeping with Rosen's notion of schools as "firms" which specialize in providing training and which provide no work opportunities.

Hence (3) may be rewritten as follows:

\[
(1-F(t))H(t) + F(t)H(t) + L(t) = 1,
\]

with \(0 \leq F(t) \leq 1\), \(0 \leq H(t) \leq 1\), \(0 \leq L(t) \leq 1\) for all \(t\).

Note that, expect for periods during which \(F(t) = 1\) (which, as just noted, may be called periods of (full-time) schooling), \(H(t)\) is essentially equivalent to what is called "hours worked" or "labor supply" in most data sets. In fact, however, as just noted, what data sets call "hours worked" is in fact "hours spent at work," for when \(0 < F(t) < 1\), \(H(t)\) hours are spent at work but only \((1-F(t))H(t)\) hours are spent actually working; the rest,
are spent receiving training. (Of course, when $F(t) = 0$ and no training occurs, all hours spent at work are spent actually working; but it is only in this case that empirical measures of "hours worked" will correspond to hours spent actually working.) In general, while most data sets contain measures of $H(t)$, measures of $F(t)$ (and thus of $F(t)H(t)$) are almost never available. Thus, much of this paper will focus on the implications of the model for the behavior of $H(t)$, as $H(t)$ is usually the only allocation-of-time variable of which empirical measures are directly available and about which testable propositions can be formulated.

I shall also assume a perfect capital market in which the individual may freely lend and borrow at a market rate of interest $r$; that the individual is endowed with an initial level of non-human assets, $A(0)$ (which could be either positive or negative, denoting positive or negative financial net worth at $t = 0$) and has an initial endowment of human wealth or human capital (consisting of "raw ability," prior experience, previous schooling and training, etc.) of $K(0)$; that the price of consumer goods is $p$; and that the "full-time" or "potential" wage (the wage paid per period of work when the entire period is spent actually working), $W_p(t)$, is equal to $yK(t)$, where $y$ is the service yield or rental per unit of human capital and $K(t)$ is the amount of human capital at time $t$ (and which, for the moment, I shall not define precisely). Hence $Y(t)$, income per period at $t$, depends on $W_p(t)$, the full-time or potential wage, and on the fraction of the period actually devoted to working. Specifically, I shall follow Rosen and Blinder and Weiss in assuming that
\[ Y(t) = W_p(t)g(F(t))H(t) = yK(t)g(F(t))H(t) \]  \hspace{1cm} (4)

where \( g(0) = 1, g(1) = 0, g'(F(t)) < 0, g''(F(t)) < 0 \) for all \( t \)

From (4) one immediately gets (5), which simply says that the time rate of change of non-human assets is equal to income from assets, \( rA(t) \), plus earnings, \( Y(t) \), less consumption expenditures:

\[ A(t) = rA(t) + yK(t)g(F(t))H(t) - pC(t) \]  \hspace{1cm} (5)

(4) generally, and the function \( g(.) \) in particular, deserves a word or two of explanation. (4) embodies essentially three assumptions: one concerns the relation between earnings per hour and hourly wages; another concerns the effects of human capital on wages; and the last concerns the effects of training on wages.

The first assumption is that wages per hour and earnings per hour are identical, i.e., I ignore the possibility that they may diverge (due, say, to the effects of piecework schemes or overtime premia). Hence I assume that \( W_0(t) \), the observed wage (i.e., pay per hour of time spent at work -- the wage measure available in actual data sets), is given by

\[ W_0(t) = Y(t)/H(t) = yK(t)g(F(t)) \]

Note that, except for periods during which \( F(t) = 0 \) and no training occurs, \( W_p(t) \), the "potential" or "full-time" wage, will not be observed, for in general not much is known about the \( g(.) \) function and empirical measures of \( F(t) \) are usually unavailable. Thus, much of this paper will focus on the implications of the model for the behavior of \( W_0(t) \), as this is the only
wage variable of which empirical measures are directly available and about which testable propositions can be formulated.

The second assumption implicit in (4) is that the wage is homogeneous of degree one in \( K \), the stock of human capital: if one has twice as much \( K \) as someone else, one is twice as productive in the market, and thus, \textit{ceteris paribus}, one's wage will be twice as large. In particular, if \( F(t) = 0 \) and \( H(t) = 1 \) -- i.e., if all available time per period is devoted to work -- the \( Y(t) = yK(t) \), i.e., \( yK(t) \) is the "full-time" or "potential" wage.

The third assumption underlying (4) is that the observed wage (and also, at any given level of hours of market time) is a decreasing and concave function of \( F(t) \), the fraction of \( H(t) \) which is devoted to training. That \( W_0(t) \) should be a decreasing function of \( F(t) \) is fairly clear; after all, an increase in \( F(t) \) implies a decrease in the fraction of time spent at work which is actually devoted to work, so \( W_0(t) \) -- pay per hour spent at work -- should fall, for each hour spent at work results in less output for one's employer. Thus \( g'(F(t)) < 0 \), \( g(1) = 0 \) (full-time training is completely unproductive for an employer), \( g(0) = 1 \) (full-time work is fully productive), and \( 0 < g(F(t)) < 1 \) for \( 0 < F(t) < 1 \) (part-time work is partly productive). That \( W_0(t) \) should be a concave function of \( F(t) \) -- i.e., that \( g''(F(t)) < 0 \) -- is less obvious. Indeed, Ghez and Becker in effect assume that \( g(F(t)) = 1 - F(t) \) and hence that \( g''(F(t)) = 0 \), i.e., that earnings, \( Y(t) \), are simply the product of the full-time wage, \( W_p(t) = yK(t) \), and the fraction of full-time which is actually spent working, \((1-F(t))H(t)\). (See, for example, p. 21.) Hence in their model the wage per hour of time spent at work, \( W_0(t) \), is always equal to the full-time wage, \( W_p(t) \), times the fraction of each hour.
of time spent at work which is actually devoted to working. This certainly seems a sensible and straight forward formulation of the relationship between \( W_0(t) \) and \( F(t) \); what, then, underlies the somewhat mysterious (and certainly more complicated) \( g(.) \) function? The answer, as Rosen and, following him, Blinder and Weiss have stressed, is an assumption that the provision of training by a firm requires increasing marginal sacrifice of its inputs (materials, physical capital, the time of other workers, etc.) -- that successive increases in the fraction of market time which a worker devotes to training become increasingly more expensive to employers and thus lead to successively larger reductions in the wage per hour spent at work. (For further discussion on this point, see Blinder and Weiss.) Thus, the \( g(.) \) function in (4) embodies what may be called the hypothesis of "increasing marginal sacrifice" in training, as adopted by Rosen and Blinder and Weiss, rather than the notion of "constant marginal sacrifice" implicit in the Ghez-Becker model.

So much for the relationships between \( Y(t), W_p(t), W_0(t), F(t), H(t) \) and \( K(t) \); but what of \( K(t) \) itself? Following my earlier discussion, I shall assume that the stock of "human capital", \( K(t) \), consists of two components: the stock of training, \( N(t) \), and the stock of experience, \( X(t) \). In general, then, one could simply say that \( K(t) = k(N(t), X(t)) \) and refer to \( k(.) \) as the "human capital function." To say this alone would be to say very little; but more can in fact be said. For example: (1) Presumably training and experience are imperfect substitutes for each other, so one may also say that \( \frac{\partial K}{\partial N} \), the elasticity of substitution between \( N \) and \( X \), is less than infinity. (2) Presumably an increase in training or experience leads to an increase in \( K \)
(i.e., in $W_p$), so $k^N_N > 0$ and $k^X_X > 0$; and presumably these increases in $N$ or $X$ are not subject to increasing marginal returns, i.e., $k^N_{NN} \leq 0$ and $k^X_{XX} \leq 0$.

While it is possible to derive a few implications about life-cycle behavior even if the $k(.)$ function is not specified further, I choose -- following the considerations discussed at the beginning of this section -- to adopt a very specific form of $k(.)$, i.e., the Cobb-Douglas function

$$K(t) = N(t)\alpha X(t)\beta \quad \text{with} \quad 0 < \alpha < 1, \quad 0 < \beta < 1$$

As is well-known, this function implies (a) that there are diminishing marginal returns to both $N$ and $X$ in the "production" of $K$ (and hence of $W_p$), in the sense that successive increases in either $N$ or $X$ lead to smaller and smaller increases in $K$ and $W_p$; (b) that the elasticities of $K$ (and thus of $W_p$) with respect to $N$ and to $X$ are constant and equal to $\alpha$ and $\beta$, respectively; (c) that the elasticity of substitution between $N$ and $X$ in "producing" $K$ (i.e., in "producing" the full-time or potential wage) is unity; and (d) that there are increasing, constant or decreasing "returns to scale" (e.g., that a doubling of both $N$ and $X$ will mean that $K$ will increase by a factor or more than two, exactly two, or less than two) as $\alpha + \beta$ exceeds, equals or is less than unity. Obviously, this is a fairly restrictive specification of the relationship between $N$ and $X$, on the one hand, and $K$, on the other. I therefore stress that in many cases the propositions which hold for this specification, (6), also hold under the more general specification, $K(t) = k(N(t), X(t))$. The main reason why I adopt (6) rather than the more general specification is that, as will be evident shortly, adoption of (6) greatly facilitates the task of describing the individual's solution to the problem of maximizing lifetime
utility.

Now that the relationship between $K$, $N$ and $X$ has been specified, one final task remains: to specify the way in which stocks of $N$ and $X$ accumulate. The stock-flow relationships for $N$ and $X$, I shall assume, are as follows:

\[
N(t) = N(t) \left[ nF(t)H(t) - \delta_N \right];
\]

(7-1)

\[
n > 0, \ n > \delta_N \text{ if } \delta_N > 0, \ N(0) > 0
\]

\[
X(t) = X(t) \left[ xH(t) - \delta_X \right];
\]

(7-2)

\[
x > 0, \ x > \delta_X \text{ if } \delta_X > 0, \ X(0) > 0
\]

or, equivalently,

\[
\frac{\dot{N}(t)}{N(t)} = nF(t)H(t) - \delta_N \quad (7-1')
\]

\[
\frac{\dot{X}(t)}{X(t)} = xH(t) - \delta_X \quad (7-2')
\]

(7-1), the stock-flow relationship for $N$, implies that increments to the stock of training are "produced" using (and are a linear function of) two "inputs": the amount of time devoted to training, $F(t)H(t)$; and the stock of training itself, $N(t)$. Hence (7-1) implies that persons with more training can more easily acquire still more training -- that "training" is productive not only as a means of generating income, but also as a means of generating still more training. On the other hand, the stock of training may decay or become obsolete; decrements to the stock of training due to such obsolescence are assumed to occur at a constant percentage rate $\delta_N$. (Note,
however, that I make no assumptions on the sign of $\delta_N$; in other words, $\delta_N$ could well be positive if, for example, economy-wide advances in productivity -- "training-augmenting technical progress" -- augment units of training by enough to offset any effects of obsolescence or decay.) So $N(t)$, the net absolute rate of change in the stock of training, is found by subtracting decrements, $\delta_N H(t)$, from increments, $N(t)P(t)H(t)$. (7-1') puts the same point a little differently: as (7-1') implies, the net \textbf{percentage} rate of change of $N(t)$, $N(t)/N(t)$, is a linear function of time devoted to training, $F(t)H(t)$, and of the rate of obsolescence of training, $\delta_N$ (which could, however, be either positive or negative, as just noted). (The assumption that $N > \delta_N$ if $\delta_N > 0$ means that it is possible to make net, as well as gross, investments in training for some $F(t)H(t) > 0$.)

(7-2), the stock-flow relationship for $X$, is quite similar in nature to (7-1). (7-2) implies that increments to the stock of experience are "produced" using (and are a linear function of) two "inputs": the amount of market time as a whole, $H(t)$; and the stock of experience itself, $X(t)$. Hence (7-2) implies that persons with more experience can more easily acquire still more experience -- that exposure to market processes and activities "does more" for persons who have already had some exposure, and are therefore (it is assumed) better able to benefit from further exposure. (In effect, then, since (7-1) and (7-2) assume that training and experience are inputs in the production of further training and experience, respectively, they imply that there are "learning by doing" effects not only in the production of market goods but also in the production of training and experience.)
(Further, (7-2) implies that decrements to the stock of experience occur at a constant percentage rate $\delta_X$, where the fact that no assumptions are made on the sign of $\delta_X$ implies that "experience-augmenting technical progress" in the economy as a whole could outweigh any processes of obsolescence or decay in the stock of experience.) In other words, $X(t)$, the net absolute change in the stock of training, is found by subtracting decrements, $\delta_X X(t)$, from increments, $X(t)H(t)$. (7-2'), like (7-1'), puts the point a little differently: as (7-2') implies, the net percentage rate of change of $X(t)$, $X(t)/X(t)$, is a linear function of time devoted to the market, $H(t)$, and of the rate of obsolescence of experience, $\delta_X$ (which could, however, be either positive or negative, as just noted). (The assumption that $x > \delta_X$ if $\delta_X > 0$ means that it is possible to make net, as well as gross, investments in experience for some $H(t) > 0$.)

While in most respects (7-2) and its equivalent, (7-2'), are quite straightforward, the assumption that $X(t)$ is a function of $H(t)$ -- total market time -- deserves a bit of explanation. While the literature does not appear to contain a clear statement on this point, the usual notion of experience seems to be that "experience" is acquired during time spent actually working, and this would seem to dictate that one specify $X(t)$ as a function of $(1-F(t))H(t)$, the portion of total market time which is actually devoted to work, rather than $H(t)$. I nevertheless specify $X(t)$ as a function of $H(t)$ for two reasons, one practical and the other theoretical.

The practical reason for specifying $X(t)$ as a function of $H(t)$ is as follows. If $X(t)$ were specified as a function of $(1-F(t))H(t)$, one
would in effect assume that persons who have just entered the labor market after schooling (and who have therefore had \( F(t) = 1 \) up to now) do so without any experience at all. Together with (6), this would imply that a school-leaver's potential wage, \( W^* \), is initially zero (at least for the first instant after schooling ends and work begins). To be sure, it turns out, as I show below, that many of the results which hold when one assumes (6) also hold under more general assumptions about the function linking \( K(t) \), \( N(t) \) and \( X(t) \) which need not imply zero potential wages at the outset of the working life. (Indeed, the results for a model which uses (6) and which assumes that \( X(t) \) depends on \( (1-F(t))H(t) \) rather than on \( H(t) \) are almost identical to the present model, which uses (6) and assumes that \( X(t) \) depends on \( H(t) \).) But since adoption of (6) turns out to simplify the analysis considerably, it seems desirable to introduce other assumptions in such a way as to ensure that they, together with (6), do not have implausible implications -- rather than to follow the reverse procedure, of first introducing other assumptions and then specifying the \( K(t) \) function, \( k(.) \), in such a way as to ensure that it and the other assumptions of the model do not have implausible implications.

The second reason for specifying \( X(t) \) as a function of \( H(t) \) is theoretical. Specification of \( X(t) \) as a function of the time spent actually working, \( (1-F(t))H(t) \), entails the assumption that one never acquires any "experience" during schooling (when \( F(t) = 1 \)) or during that portion of one's work time which is devoted to training -- even though, as many (notably graduate students) will attest, schooling and training can in a very real sense be "work." In other words, schooling and training impart not only cognitive and motor skills and what may loosely be termed
"knowledge" (e.g., languages, mathematics and so on), but also certain kinds of qualities which may be called "experience," e.g., the habit of punctuality, the habit of following rules and instructions, the habit of planning and making decisions, etc. Thus, in keeping with my previous discussion, I distinguish between "training" and "experience" by defining the former as something which can be increased without any increase in $H(k)$ or reduction in $L(t)$, merely by an increase in $Y(t)$; and the latter as something which can only be increased by an increase in $H(t)$ (i.e., a reduction in $L(t)$).

Equations (1) - (7) are thus the foundations of the model. To summarize: the individual is assumed to maximize lifetime utility,

$$ F = \int_0^T [U(C(t)) + V(L(t))] e^{-\beta t} \, dt + B(A(T)) \tag{1} $$

where

- $U'(C(t)) > 0$, $U''(C(t)) < 0$
- $V'(L(t)) > 0$, $V''(L(t)) < 0$
- $B'(A(t)) > 0$, $B''(A(t)) < 0$

$$ \lim_{A(T) \to 0} B'(A(T)) = \infty; \quad \lim_{C(t) \to 0} U'(C(t)) = \infty; \quad \lim_{L(t) \to 0} V'(L(t)) = \infty \tag{2-1} $$

Indeed, one might argue that one acquires both "experience" and "training" merely by living and keeping one's eyes and ears open. But such "learning by living" and "training by living" effects are, I shall assume, adequately reflected in the "decay" parameters $\delta_N$ and $\delta_Y$, for, as noted earlier, these need not be of any particular sign. The effect of the distinction made here between learning by doing and training, then, is that one acquires more experience and training -- over and above what is acquired merely by living -- by devoting more time to the market and to training, respectively.
and to allocate time to leisure and to the market (market time in turn being allocated between training and working), i.e.,

\[(1-F(t))H(t) + F(t)H(t) + L(t) = 1, \quad (3')\]

with \(0 \leq F(t) \leq 1, 0 \leq H(t) \leq 1, 0 \leq L(t) \leq 1\) for all \(t\)

Experience and training accumulate according to the stock-flow relations

\[\dot{N}(t) = N(t) [\delta F(t)H(t) - \delta_N]; \; n > 0, \; n > \delta_n \text{ if } \delta_N > 0, N(0) > 0 \quad (7-1)\]

\[\dot{X}(t) = X(t) [\delta H(t) - \delta_X]; \; x > 0, \; x > \delta_x \text{ if } \delta_X > 0, X(0) > 0 \quad (7-2)\]

Experience and training in turn determine human capital, which is proportional to the potential wage:

\[K(t) = N(t)^\alpha X(t)^\beta; \quad 0 < \alpha < 1, \; 0 < \beta < 1 \quad (6)\]

\[W_p(t) = yK(t) = yN(t)^\alpha X(t)^\beta \quad (6')\]

Given the notion of "increasing marginal sacrifice" in the allocation to time to training, one has

\[Y(t) = W_p(t)g(F(t))H(t) = yK(t)g(F(t))H(t) = yN(t)^\alpha X(t)^\beta g(F(t))H(t) \quad (4)\]

where \(g(0) = 1, g(1) = 0, g'(F(t)) < 0, g''(F(t)) < 0\)

and so the rate of change of assets is

\[\dot{A}(t) = rA(t) + Y(t) - \rho C(t) = rA(t) + yN(t)^\alpha X(t)^\beta g(F(t))H(t) - \rho C(t) \quad (5)\]

In Part II, I introduce a few additional assumptions which are mostly technical in nature and then derive, in a fairly rigorous way, the implications

\[1/\text{By now it should also be obvious that I assume a completely static world of perfect foresight, i.e., one in which parameters such as } y, n, \text{ etc. are known constants.}\]
of the model for life-cycle behavior. Before turning to Part II, the reader may find it helpful to refer to Figs. 7 - 9, which depict graphically, and thus may serve as an advance summary of, the main results derived in Part II.

II. Implications of the Model for Life-Cycle Behavior

While one may analyze life-cycle behavior under assumptions (1) - (7) in a variety of ways, I find it convenient to state and solve the individual's maximization problem in terms of optimal control theory. After formulating the model as a problem of optimal control in Part IIA, I then present in Part IIB a stage-by-stage discussion of a "normal" life cycle. In Part IIC, I use the results obtained in Part B to describe various possible lifetime profiles of $H(t)$, $W_o(t)$ and $Y(t)$, i.e., observed hours of work, observed wage rates and earnings.

IIA. The Life Cycle in Terms of Optimal Control

Formally, I have assumed that the individual's objective is

$$\text{Maximize} \quad \int_0^T e^{-pt} [u(C(t)) + v(t)] dt + B(A(T))$$

subject to

$$0 \leq P(t) \leq 1$$
$$0 \leq H(t) \leq 1$$
$$0 \leq L(t) \leq 1$$
$$0 \leq C(t)$$
subject to

\( (1-F(t))H(t) + F(t)H(t) + L(t) = 1 \)
\( K(t) = N(t)^a X(t)^b \)
\( A(t) = rA(t) + yK(t)g(F(t)H(t) - FC(t)) \)
\( N(t) = N(t)[nF(t)H(t) - \delta_N] \)
\( \dot{X}(t) = X(t)[xH(t) - \delta_X] \)
\( N(0) > 0 \)
\( X(0) > 0 \)

By introducing three costate variables \( S_1(t) = e^{-pt} \lambda_1(t), S_2(t) = e^{-pt} \lambda_2(t) \) and \( S_3(t) = e^{-pt} \lambda_3(t) \)—one may write the Hamiltonian of this problem as

\[ \mathcal{H} = e^{-pt} \left[ [U(C(t)) + V(L(t))] + \lambda_1(t) N(t)[nF(t)H(t) - \delta_n] \right. \]
\[ \left. + \lambda_2(t) X(t)[xH(t) - \delta_X] + \lambda_3(t)[rA(t) + yN(t)^a X(t)^b g(F(t)H(t) - FC(t))] \right] \]

where the \( \lambda_i(t) \) \((i = 1, 2, 3)\) are dynamic analogues of the Lagrange multipliers of static optimization problems, and may be interpreted as the shadow price or value of training \((i = 1)\), of experience \((i = 2)\) and of assets or non-human wealth \((i = 3)\). By (2), (3) and the maximum principle of L.S. Pontryagin et al., the following are necessary conditions for an optimal solution to the individual's maximization problem:

\[ U(C(t)) = \lambda_3(t)P \]  (8-1)
\[ V'(1-H(t)) = \lambda_1(t)H(t) + \lambda_2(t)X(t) + \lambda_3(t)Y(t) + r(t), \]  
\[ \text{with } \lambda_1(t)H(t) + \lambda_3(t)X(t) = 0. \]

\[ \lambda_1(t)N(t) = \lambda_3(t)X(t)[-g'(0)] \rightarrow F(t) = 1, \]
\[ \lambda_1(t)N(t) = \lambda_3(t)X(t)[-g'(0)] \rightarrow F(t) = 0, \]  
\[ \lambda_1(t)N(t) = \lambda_3(t)X(t)[-g'(F(t))] \text{ otherwise} \]  

\[ \dot{\lambda}_1(t) = \rho \lambda_1(t) - \lambda_1(t)[H(t)/N(t)] \]  
\[ \dot{\lambda}_2(t) = \rho \lambda_2(t) - \lambda_2(t)[X(t)/H(t)] \]  
\[ \dot{\lambda}_3(t) = (\rho - r) \lambda_3(t) \]  
\[ \dot{\lambda}_4(t) = \beta'(A(t)) \]  
\[ \lambda_1(T) = \lambda_2(T) = 0, \]  
\[ \lambda_3(T) = \beta'(A(T)) \]

and also equations (7) and (5), the stock-flow relationships for \( N(t), X(t) \) and \( A(t) \), respectively. (To simplify notation, from now on I shall for the most part drop time operators for any \( \tau, 0 < \tau < T \). Thus, from now on, "\( \lambda \)" should be understood to mean \( X(t) \), etc.

Equations (8-1) - (8-3) embody the usual marginal benefit-marginal cost comparisons of economic analysis. (8-1) says that the marginal utility of consumption is equated to the marginal utility of the income which must be used to purchase it. (8-2) says that the marginal utility of leisure

\[ \dot{Y} \text{ By (8-1), } \lambda_3(t) > 0 \text{ always, so } \dot{\lambda}_3(t) \lesssim 0 \text{ as } \rho - r \lesssim 0. \]
(i.e., of reducing market time) is never less than the sum of the marginal utilities of the training, experience and income which are foregone when one reduces market time in order to enjoy more leisure. (Indeed, the marginal utility of leisure could be strictly greater than this sum when all time is devoted to leisure.) (Note that assumptions (2) exclude the possibility that either C or L could ever be zero.) (8-3) says that the fraction of market time devoted to training will be unity (in between zero and unity) whenever the marginal utility of full-time (part-time) training exceeds (equals) the marginal utility of the income which one foregoes in order to pursue such training. Equations (8-4) - (8-6) give the rates of change of the shadow prices of training, experience and assets, respectively. Equations (8-7) and (8-8) are the so-called transversality conditions. Initial endowments of training and experience are strictly positive, so that, by (7-1) and (7-2), the stocks of training and experience are always positive. Thus, (8-7) states, at the end of "life," i.e., at T, neither training nor experience has any (further) value. (8-8) states that the shadow price or value of assets at time T is equal to the marginal utility of assets at time T (i.e., of bequests).

I note parenthetically that (8-3) discloses an interesting implication of the stock-flow assumption for N, (7-1). The left-hand side of the first-order condition given by (8-3), \( \lambda_1(t)N(t)n \), may be interpreted as the (instantaneous) marginal benefit of investment in training, while the right hand side, \( \lambda_3(t)Y(t)[-g'(F(t))] \), may be interpreted as the instantaneous marginal cost of such investment. The stock-flow specification given by (7-1) thus turns out to imply somewhat more than just that investment in
training affects one's productivity in producing further training as well as one's productivity doing actual work (as is already evident from inspection of (7) and (4)). As (8-3) indicates, (7-1) also turns out to imply that investment in training is "neutral" in its effects on productivity in these two spheres, at least in the sense that the (instantaneous) marginal cost-marginal benefit comparison which potential investors make is (in this model) independent of \( N \) per se and depends only on the ratio of \( K \) to \( N \). That is, a simultaneous doubling of both \( K \) and \( N \) will leave unchanged the marginal costs and marginal benefits of further training. (Indeed, if \( \alpha + \beta = 1 \), a simultaneous doubling of \( N \) and \( X \) will likewise leave the marginal benefit-marginal cost comparison unchanged.) This kind of "neutrality" is of a much more specific sort than the kind discussed by Yoram Ben-Porath; "neutrality" in Ben-Porath's terminology seems to mean simply that \( N \) (or for that matter \( X \)) affects productivity both in the production of goods and in the production of more \( N \). (See p. 135.) The present model possesses "neutrality" in this sense, but it also possesses "neutrality" in another, namely, that investment in \( N \) in this model is unaffected by the level of \( N \) and depends only on the ratio of \( K \) to \( N \).

(Note also that if \( \alpha = 1 \) and \( N \) were the only kind of human capital, i.e., if \( K = N \), then of course \( N \) would be completely neutral both in the sense of Ben-Porath and in my sense, since \( K/N = 1 \) always. This is precisely the case considered by BlINDER and WEISS.1/)

1/Mincer (1970, p. 148; emphasis supplied) says that in the Ben-Porath "pure training" model human capital is "neutral" because "...it increases productivities in the market at the same rate as it does in the production of additions to the stock of human capital." However, while in the Ben-Porath model human capital does indeed increase productivity in both kinds of production, it is not at all clear that it does so "at the same rate" in both or what "at the same rate" might mean. Indeed, while Ben-Porath notes that \( K \) is neutral in his model in the first sense given here, he says nothing about whether it is neutral in the second. (See p. 135.)
So much for necessary conditions. A sufficient condition for a maximum is that $\mathcal{K}^*$, the maximized Hamiltonian, be concave in the state variables $A$, $N$ and $X$ for given $\lambda_1$, $\lambda_2$, $\lambda_3$ and $t$. (See Weiss, pp. 1312-1313.) A necessary and sufficient condition for the concavity of $\mathcal{K}^*$ is that $J$ be negative semi-definite, where

$$
J = 
\begin{bmatrix}
\frac{\partial^2 \mathcal{K}^*}{\partial A^2} & \frac{\partial^2 \mathcal{K}^*}{\partial A \partial N} & \frac{\partial^2 \mathcal{K}^*}{\partial A \partial X} \\
\frac{\partial^2 \mathcal{K}^*}{\partial A \partial N} & \frac{\partial^2 \mathcal{K}^*}{\partial N^2} & \frac{\partial^2 \mathcal{K}^*}{\partial N \partial X} \\
\frac{\partial^2 \mathcal{K}^*}{\partial A \partial X} & \frac{\partial^2 \mathcal{K}^*}{\partial N \partial X} & \frac{\partial^2 \mathcal{K}^*}{\partial X^2}
\end{bmatrix}
$$

$$
0 \quad 0 \\
\lambda_1 \frac{\partial}{\partial N} [\text{mF} - \delta_N] \\
\lambda_3 \gamma \frac{\partial}{\partial N} [(K/N)g(F)H] \\
\lambda_2 \frac{\partial}{\partial N} [xH - \delta_X] \\
0 \\
\lambda_3 \gamma \frac{\partial}{\partial X} [(K/X)g(F)H]
$$

$$
0 \\
\lambda_1 \frac{\partial}{\partial X} [\text{mF} - \delta_X] \\
\lambda_3 \gamma \frac{\partial}{\partial X} [(K/X)g(F)H] \\
\lambda_2 \frac{\partial}{\partial X} [xH - \delta_X] \\
0 \\
\lambda_3 \gamma \frac{\partial}{\partial X} [(K/X)g(F)H]
$$

In particular, strong sufficient conditions for the concavity of $\mathcal{K}^*$ include the following: that, whenever $0 < H < 1$ and $F = 0$, $\lambda_1$, $\lambda_2$, and $\lambda_3$ are given and $\mathcal{K}$ is maximized,
\[ \frac{\partial Y}{\partial N} = (1-\alpha) \lambda_3 Y(K/N)^2 + \lambda_3 Y(K/N) \cdot \frac{\partial H}{\partial N} \leq 0 \]

\[ \frac{\partial Y}{\partial X} = -(1-\beta) \lambda_3 Y(K/X)^2 + \lambda_3 Y(K/X) \frac{\partial H}{\partial X} \leq 0 \]

From (8-2), the first order condition for \( H \), it follows that when \( 0 < H < 1 \) and \( F = 0 \),

\[ [-v''(1-H)] \frac{\partial H}{\partial N} = \lambda_3 Y \frac{X}{H} \frac{\partial H}{\partial N} + \left[ \lambda_2 x + \lambda_3 Y(K/X) \right] \frac{\partial H}{\partial X} \]

for given \( \lambda_1 \), \( \lambda_2 \), and \( \lambda_3 \). Hence, whenever \( 0 < H < 1 \), \( F = 0 \), and \( \lambda_1 \), \( \lambda_2 \) and \( \lambda_3 \) are fixed,

\[ \frac{\partial H}{\partial N} = \frac{\lambda_3 Y(K/N)}{[-v''(1-H)]} \]

\[ \frac{\partial H}{\partial X} = \frac{\lambda_2 x + \lambda_3 Y(K/X)}{[-v''(1-H)]} \]

Now substitute these expressions and (8-1) into the above pair of inequalities, and then rearrange terms and multiply by \( \frac{[-v''(1-H)]}{v'(1-H)} \cdot \frac{X^2}{\lambda_3 Y K} \) and \( \frac{[-v''(1-H)]}{v'(1-H)} \cdot \frac{X^2}{\lambda_3 Y K} \), respectively, to obtain the following strong sufficiency conditions:

\[ (1-\alpha) \frac{V'_H}{H} + (1-\beta) \frac{(y/p)K u'(c)}{v'(1-H)} - 1 \geq 0 \quad (9-1) \]

\[ (1-\beta) \frac{V'_H}{H} + (1-\beta) \frac{(y/p)K u'(c)}{v'(1-H)} - 1 \geq 0 \quad (9-2) \]

where \( \frac{V'_H}{H} = \frac{[-v''(1-H)]_H}{v'(1-H)} \) is the elasticity of the marginal utility of leisure.
with respect to work. (Note also that $\alpha = \frac{d \log W_p}{d \log N}$ is the elasticity of the potential wage with respect to training; that $1-\alpha = -\frac{d \log (\partial W_p/\partial N)}{d \log N}$

$= -\frac{d \log \alpha(K/N)}{d \log N}$ is the elasticity of the marginal "productivity" of training in raising the potential wage; that $\beta = \frac{d \log W_p}{d \log X}$ is the elasticity of the potential wage with respect to experience; and that $1 - \beta = -\frac{d \log (\partial W_p/X)}{d \log X}$

$= -\frac{d \log \beta(X/X)}{d \log X}$ is the elasticity of the marginal "productivity" of experience in raising the potential wage.) By (8-1) and (8-2), $V'(1-H)$

$\geq \lambda yK = (y/p)K \cdot U'(c)$ always when $F = 0$, so conditions (9) will always hold if

$$ (1 - \alpha)E_H^U - \alpha > 0 \quad \text{(9-1')} $$

$$ (1 - \beta)E_H^U - \beta > 0 \quad \text{(9-2')} $$

In other words, when $F = 0$ and $0 < H < 1$, a strong sufficient condition for the existence of an absolute maximum of $K$ is that the elasticity of the wage with respect to experience be less than the product of the elasticity of the marginal productivity of the wage and the elasticity of the marginal utility of leisure with respect to market time and similarly for training. These are exactly the sufficiency conditions obtained by Weiss (pp.1312-1315) for the "pure experience" case in which $a = n = 0$ and $K = X$. As Weiss remarks (p. 1303), a propos of his own model, "[This condition] is derived from the requirement that for a given shadow price of experience, the marginal gain from additional experience is decreasing..."
So much for sufficient conditions. Now, equations (8) imply that the individual's lifetime may be divided into four separate periods or stages, according to the values of the control variables F and H, as follows:

(i) a period of "schooling" (i.e., all non-leisure time is devoted to investment), in which $0 < H < 1$ and $F = 1$; (ii) a period of "on-the-job training" (i.e., some non-leisure time is devoted to work, and some is devoted to investment, presumably at work), in which $0 < F < 1$ and $0 < H < 1$; (iii) a period of "pure work" (i.e., all non-leisure time is devoted to work), in which $F = 0$ and $0 < H < 1$; and (iv) a period of "retirement" (i.e., all time is devoted to leisure), in which $H = 0$ and $L = 1$.\(^{1/}\)

To describe these states, and hence to describe the solution to the individual's maximization problem, I find it convenient to introduce three new variables, $\mu_1$, $\mu_2$, and $\mu_3$, where

\[
\begin{align*}
\mu_1 & = \lambda_1 N \\
\mu_2 & = \lambda_2 X \\
\mu_3 & = \lambda_3 X
\end{align*}
\]

Using these new variables, I shall now rewrite the necessary conditions (8-2) - (8-7) as follows:

\[
V'(1-H) \geq \mu_1 NF + \mu_2 X + \mu_3 X g(F)
\]  
\[(8-2')\]

with $> \rightarrow H = 0$

\(^{1/}\)I show later that under certain conditions Stage I precedes Stage II, which in turn precedes Stage III, etc. However, so far the numbering of the stages implies nothing about the chronological order in which they appear in the individual's life.
\[ \mu_1 n > \mu_2 v \left[ -g'(1) \right] \quad \Rightarrow \quad F = 1, \]

\[ \mu_1 n \leq \mu_2 v \left[ -g'(0) \right] \quad \Rightarrow \quad F = 0, \quad (8-3') \]

\[ \mu_1 n = \mu_3 v \left[ -g'(r) \right] \quad \text{otherwise} \]

\[ \dot{v} = \dot{\lambda}_1 n + \dot{\lambda}_2 N = \rho \mu_1 - \omega_2 y g(F) H \quad (8-6') \]

\[ \dot{\mu}_2 = \dot{\lambda}_2 X + \dot{\lambda}_2 X = \rho \mu_2 - 3 \mu_3 y g(F) C \quad (8-5') \]

\[ \dot{\mu}_3 = \lambda_3 K + \lambda_3 \dot{K} = \mu_3 \left[ c_{m} + H + B X H - \gamma \right], \quad (8-6') \]

where \( \gamma = r - \rho + \delta \)

\[ \delta = \alpha \delta_n + B \delta_X \]

\[ \mu_1(T) = \lambda_1(T) N(T) = \mu_2(T) = \lambda_2(T) X(T) = 0 \quad (8-7') \]

Thus the individual's lifetime optimization strategy reduces to two sets of equation systems: (8-1) and (8-8), giving consumption and asset behavior; and (8-2') - (8-7'), giving the behavior of L, H, and F. Indeed, since the utility function is separable in consumer goods and leisure time, we can analyze the two sets of equation systems independently of each other. Moreover, each set is linear and autonomous in either the two "old" costate variables, \( \lambda_1 \) and \( \lambda_3 \), or the three "new" state variables, \( \mu_1 \), \( \mu_2 \), and \( \mu_3 \). Hence neither set entails more than three dimensions, and analysis of either becomes a relatively simple task.\(^1\)

\(^1\)It should now be evident that adoption of the human capital function, (6), simplifies the analytical task considerably. Adoption either of other specific functions for the relationship between \( K \), \( N \), and \( H \) (e.g., a CES function, \( K = [a(N)^{-c} + b(X)^{-c}]^{-1/c} \)) or of the completely general function \( K = k(N,X) \) would either entail four or more state variables (which would force the analysis into four or more dimensions) or generate a system of complex equations which would be non-linear in the state variables, or both.

\(^2\)Note from (8-1) that \( \lambda_3 > 0 \) always, and from (6) and (7) that \( N, X, \) and \( K \) are also always positive. So \( \mu_3 = \lambda_3 K > 0 \) always. Also by (8-6') and (8-5'), when \( \mu_1 \) and \( \mu_2 \) are negative, they will never again be positive; but, in view of (8-7'), this means that \( \mu_1 \) and \( \mu_2 \) must be positive for \( t < T \). (Hence \( \lambda_3 \) and \( \lambda_2 \) must likewise be positive for \( t < T \).)
Before describing the individual's lifetime optimization process, I shall make several additional assumptions. First, I shall assume that the individual's subjective rate of discount ("rate of impatience") is less than the market rate of interest:

$$\rho - r < 0$$  \hspace{1cm} (10)

(10) of course implies that

$$\nu = r - \rho + \delta > 0$$  \hspace{1cm} (11)

In view of (10), it follows immediately from (2-1), (8-1) and (8-6) that C will rise monotonically over the life-cycle. To put the point a bit differently if (10) does not obtain, then, by (2-1),(8-1) and (8-6), C will necessarily fall over the life-cycle, a proposition which has little or no empirical support. Since the reverse -- C rising, not falling-- seems much closer to actual behavior, \(1/(10)\) seems a not unreasonable assumption to make, at least as a first approximation.

\(1/(\)In fact, while C seems to rise for most of the life-cycle, it also seems to fall towards the end of life. (See, e.g., Ghez and Becker, esp. Ch. 2) My model cannot possibly generate humped age-consumption profiles (unless it is assumed that \(\rho\) is a function of age and is less (greater) than \(r\) in young (old) age). Thus the model's predictions on consumption may not be in full agreement with actual behavior; but as my main objective is to analyze the allocation of time, I am fairly content with the consumption side of the model as a rough first approximation. Certainly, a model which did not assume a utility function separable in goods and time could easily generate more acceptable propositions on C; see Heckman, 1974, and Ghez and Becker. But while adoption of such a model would render the task of analyzing the allocation of time much more difficult, it would not lead to appreciably richer implications about the allocation of time over the life cycle.
Second, I shall assume that

\[ \rho \{ 1 + \left( \frac{\beta x}{\alpha n} \right) \left[ -g'(0) \right] \} \leq r + \delta \]  \hfill (12-1)

\[ \alpha n + \beta x \geq r + \delta \]  \hfill (12-2)

These assumptions are sufficient to ensure that a unique singular solution to the differential equation system \((8-4') - (8-6')\) occurs at positive \(\mu_1\), \(\mu_2\) and \(\mu_3\) with \(0 < F < 1\) and \(0 < H < 1\).

To see why \((12-1)\) and \((12-2)\) are sufficient, note first from \((8-4') - (8-6')\) that \(F\) must be less than 1 and \(H\) must be greater than zero for a singular solution \((\mu_1 = \mu_2 = \mu_3 = 0)\) to occur at positive \(\mu_1\), \(\mu_2\) and \(\mu_3\); and also that at a singular solution for which \(F = F^*\), where \(0 < F^* < 1\),

\[ \mu_2 = \frac{\beta}{\alpha} \mu_1 \]  \hfill (by \((8-4')\) and \((8-5')\), where \(\mu_1 = \mu_2 = 0, \mu_1 > 0, \mu_2 > 0\))

\[ \mu_1 n = \mu_3 y \left[ -g'(F) \right] \]  \hfill (by \((8-3')\))

\[ \alpha n F H + \beta n H = y \]  \hfill (by \((8-6')\) when \(\mu_3 = 0, \mu_3 > 0\))

and thus that, at the singular solution,

\[ \phi(F^*) = \frac{F^* - g'(F^*)}{g(F)} + \frac{\beta x}{\alpha n} \left[ -g'(F^*) \right] \equiv \frac{y}{\rho} \]  \hfill (by \((12-1)\))

Since \(\phi'(F) > 0\) and \(\phi(1) = \gamma \rho\), condition \((12-1)\), which states that \(\phi(0) = \left( \frac{\beta x}{\alpha n} \right) \left[ -g'(0) \right] < \gamma \rho\), will ensure that \(0 < F^* < 1\).

Now note from \((8-3') - (8-6')\) that, at this singular solution, \(H = H^*\), where
\[ H^* = \frac{\gamma}{\alpha n F^* + \beta x} = \frac{[-g'(F^*)]}{g(F^*)} \frac{n}{\alpha n} \]

and \( F^* \) is implicitly given by the condition that \( g(F^*) = \gamma/\rho \). Thus, if \( H^* < 1 \), it must be the case that \( \alpha n F^* + \beta x > \gamma \), i.e., it must be the case that \( F^* > (\gamma - \beta x)/\alpha n = Y \). (Since \( 0 < F < 1 \) and \( 0 < H < 1 \), and since \( \alpha n F H + \beta x H = \gamma \), clearly \( Y < 1 \).) To see under what conditions this will be true, define a new function, \( G(F^*) \), where

\[ G(F^*) = \frac{[-g'(F^*)]}{g(F^*)} \frac{\alpha n F^* + \beta n}{\alpha r} = \frac{\gamma}{\rho} \]

Clearly \( g'(F) > 0 \), so to determine the conditions under which \( F^* > Y \) (and thus \( H^* < 1 \)) one need simply determine the conditions under which \( G(F^*) > G(Y) \). If \( G(F^*) > G(Y) \), then

\[ \frac{\gamma}{\rho} > \frac{[-g'(Y)]}{g(Y)} \frac{\alpha n + \beta x}{\alpha n} = \frac{[-g'(Y)]}{g(Y)} \frac{\gamma}{\alpha n} \]

i.e.,

\[ g(Y) > \frac{[-g'(Y)]}{g(Y)} \frac{\alpha n + \beta x}{\alpha n} = \frac{[-g'(Y)]}{g(Y)} \]

\[ (12-2') \]

But \( g(Y) > (1 - Y)[-g'(Y)] \) for the \( g(\cdot) \) function.\(^1\) So a sufficient condition for \( H^* < 1 \) is that \( (r + \delta - \beta x)/\alpha n < 1 \), i.e., \( (12-2') \).

\(^1\) See, for example, Blinder and Weiss, who use exactly the same argument in the context of a discussion of a pure training model. The point is that since \( g(\cdot) \) is concave with \( g(0) = 1 \) and \( g(1) = 0 \), the first-order Taylor expansion of \( g(F) \) about the point \( F = Y = (\gamma - \beta x)/\alpha n < 1 \) will be less than \( g(1) = 0 \) -- in other words, \( g(Y) > (1 - Y)[-g'(Y)] \).
(Note also that since the \( g(.) \) function is assumed concave, with \( g(0) = 1 \) and \( g(1) = 0 \), it follows that, for the \( g(.) \) function, 

\[-g'(0)] < 1 < [-g'(1)]\]

a fact which will prove helpful later on.)

In economic terms, the notion that a singular solution occurs at positive \( \mu_1, \mu_2 \) and \( \mu_3 \) only with \( 0 < \gamma < 1 \) and \( 0 < \Delta < 1 \) means that the shadow values of the total stocks of training \( (\lambda_1 Y = \mu_1) \) of experience \( (\lambda_2 Y = \mu_2) \) and of human capital \( (\lambda_3 Y = \mu_3) \) can be positive and constant \( (\mu_1 = \mu_2 = \mu_3 = 0) \) only when some portion of total time is devoted to the market and some portion of market time is devoted to training. The sufficient conditions for this to occur, equations (12), may be interpreted as follows:

(i) \( \alpha + \beta \gamma \), the "full-time rate of marginal benefits" -- i.e., the rate of growth in the full-time wage \( (W_y = y K) \) during a period of full-time work \( (N = 1) \) in which all time is devoted to training \( (\gamma = 1) \)-- must exceed \( r + \delta \), the "rate of marginal costs" -- i.e., the rate of decrease in the value of investments in training and experience (a decrease which comes about because such investments pay off only in future periods and therefore are subject both to discounting, \( r \), and to decay, \( \delta = \alpha + \beta \cdot \delta \cdot \gamma \)). Otherwise, the equating of marginal benefits and marginal costs at the singular solution might entail \( N > 1 \).

(ii) \( \phi \), the "rate of impatience," should not be "too high" relative to the rate of decrease in the value of investment, \( r + \delta \). ("Too high" is expressed relative to the production parameters and wage-elasticities of
N and X, and to the marginal sacrifice of income consequent upon raising F from zero, \([-g'(0)]\). Otherwise, the equating of marginal costs and marginal benefits at the singular solution might require F = 0, i.e., the individual might be so "impatient" as to be unwilling to undertake any investment in training at all.

To sum up: for that portion of the model which refers to the allocation of time, there are several first-order or necessary conditions, (8-2') - (8-7'); some strong sufficienty conditions, (9-1') and (9-2'); an assumption on the magnitude of \(\sigma\), (11); and assumptions on the nature of the singular solution to the differential equation system (8-6') - (8-6'), i.e., (12-1) and (12-2). In using these to describe the behavior of hours of work, wages and earnings over the life cycle, I find it convenient, at least at first, to plot the life-cycle in the three-dimensional \(\mu_1, \mu_2, \mu_3\) space depicted in Figure 1. As noted earlier, the life-cycle may be divided into four stages, according to the values of F and H. In particular:

**Stage IV:** Here \(H = 0\) for all \(F\), so, by the first order condition (8-2'), \(V'(1) \geq \max_{0 \leq F \leq 1} \mu_1 nF + \mu_2 x + \mu_3 yg(F)\). When \(F = 1\), \(V'(1) \geq \mu_1 n + \mu_2 x\) and \(\mu_1 n \geq [-g'(1)]\mu_3 y\), i.e., Stage IV lies below the triangular plane ABE. When \(F = 0\), \(V'(1) \geq \mu_2 x + \mu_3 y\) and \(\mu_1 n \leq [g'(0)]\mu_3 y\), i.e., Stage IV lies below the triangular plane CDE. Finally, when \(0 < F < 1\), \(V'(1) \geq \max_{0 \leq F \leq 1} \mu_1 nF + \mu_2 x + \mu_3 yg(F)\) \(\geq \zeta(\mu_1, \mu_2, \mu_3)\), and Stage IV lies below the concave surface BCE. To see why BCE is concave, note that \(\zeta(\mu_1, \mu_2, \mu_3)\) is a convex function by virtue of the fact that, for any \(0 < \nu < 1\),
\[ v \left\{ \max_{\mathcal{S}_1} \left[ \mu_1 nF + \mu_2 x + \mu_3 y g(F) \right] \right\} + (1-v) \left\{ \max_{\mathcal{S}_1} \left[ \mu_1 nF + \mu_2 x + \mu_3 y g(F) \right] \right\} \]

\[ \geq \max_{\mathcal{S}_1} \left\{ v \left[ \mu_1 nF + \mu_2 x + \mu_3 y g(F) \right] + (1-v) \left[ \mu_1 nF + \mu_2 x + \mu_3 y g(F) \right] \right\} \]

Hence, in Figure 1a, Stage IV consists of the space lying below ABCDE.

**Stage I:** Here \( 0 < H < 1 \) and \( F = 1 \), and, by the first order conditions,
\[ V'(1-H) = \mu_1 n + \mu_2 x \text{ and } \mu_1 n \geq \mu_3 y [g'(1)] \]
Hence, in Figure 1, Stage I consists of the space lying above the plane labelled "\( F = 1 \)." Note that the \( F = 1 \) plane intersects the Stage IV region along \( BE \); coincides with the \( \mu_2 \) axis (for \( \mu_2 \geq \frac{V'(1)}{x} \) and \( \mu_1 = \mu_3 = 0 \)); is perpendicular to the \( \mu_1 \mu_3 \) plane; and intersects the \( \mu_1 \mu_2 \) and \( \mu_2 \mu_3 \) planes. Also, along the \( F = 1 \) plane,
\[ \frac{\partial \mu_1}{\partial \mu_2} = \frac{\partial \mu_3}{\partial \mu_2} = 0 \text{ and } \frac{\partial \mu_1}{\partial \mu_3} = [-g'(1)] > 0 \]

**Stage II:** Here \( 0 < H < 1, 0 < F < 1 \), and, by the first order conditions,
\[ V'(1-H) = \mu_1 n + \mu_2 x + \mu_3 y g(F) = \mu_2 x + \mu_3 y [g(F) + F [g'(F)]] \text{ and } \mu_1 n = \mu_3 y [g'(F)] \]
In Figure 1, then, Stage II consists of the space in between the \( F = 1 \) plane and the \( F = 0 \) plane, to be described in the context of the discussion of Stage III below. (Note from the discussion of Stage IV that Stage II lies above the concave surface EBC.)

**Stage III:** Here \( 0 < H < 1 \) and \( F = 1, \mu_1, \mu_3 \) by the first-order conditions,
\[ V'(1-H) = \mu_2 x + \mu_3 y \text{ and } \mu_1 n \leq \mu_3 y [g'(O)] \]
Hence, in Figure 1, Stage III consists of the space lying below the plane labelled "\( F = 0 \)." Note that the \( F = 0 \) plane intersects the Stage IV region along \( EC \); coincides with the \( \mu_2 \) axis (for \( \mu_2 \geq \frac{V'(1)}{x} \) and \( \mu_1 = \mu_3 = 0 \)); is perpendicular to the \( \mu_1 \mu_3 \) plane;
intersects the $\mu_1\mu_2$ and $\mu_2\mu_3$ planes; and (Since $g(1)=0, g(0)=1, g'(f)<0$ and $[-g''(f)]>0$, so that $[-g'(1)]>[-g'(0)]$) lies entirely above the $F = 1$ plane. Also, along the $F = 0$ plane, $\partial\mu_1/\partial\mu_2 = \partial\mu_2/\partial\mu_3 = 0$ and $\partial\mu_1/\partial\mu_3 = [-g'(0)]>0$.

I now use Figure 1 and several additional implications of the first order conditions (which will be introduced presently) to describe what may be called the "normal" life-cycle. By this is simply meant a life-cycle which starts (as most seem, in fact, to start) in Stage I, with $H > 0$ and $F = 1$, i.e., with a period of specialization in training (e.g., schooling). (The nature of "unusual" life-cycles -- those which start with $H = 0$ or with $H > 0$ and $F < 1" will be understood much more easily once the "normal" case has been analyzed.) Note first from (8-4') - (8-6') that in Stage IV, when $H = 0, \dot{\mu}_1 = 0$ and $\dot{\mu}_2 > 0$ always (except, of course, when $\mu_1 = \mu_2 = 0$); and that, in Stage I, when $F = 1, \mu_1 > 0$ and $\dot{\mu}_1 > 0$ and $\dot{\mu}_2 > 0$ as $\mu_2 \geq 0$ always. Moreover, by the transversality condition (8-7') the individual must by time $T$ arrive somewhere on the $\mu_3$ axis (i.e., have $\mu_1 = \mu_2 = 0$). And since all individuals must end life somewhere on the $\mu_3$ axis, it follows that all individuals who start life in Stage I must pass through both Stage II and Stage III (and may also pass through Stage IV as well), for it is now evident from Figure 1 and the fact that $\dot{\mu}_1 > 0$ in Stages IV and I that there is no other way by which someone in Stage I can get to the $\mu_3$ axis.

IIB. Analysis of Stages in the "Normal" Life Cycle

In view of the foregoing analysis, the normal life cycle must unfold as follows:
1. **Stage I:** Here, $0 < H < 1$ and $F = 1$, i.e., Stage I constitutes a period of specialization in training ("full-time schooling"). Also, by (8-2'), (8-4') and (8-5'), when $F = 1$ then

$$[-V''(1-H)]H = \mu_1 + x\mu_2 = \rho[\mu_1 + x\mu_2] = \rho V'(1-H) > 0$$

i.e., $H$ (which in this case is simply hours of schooling) rises as time (and schooling) goes on. Moreover, since $[-V''(1-H)] : H = \rho$,

$$\frac{[-V''(1-H)]}{V'(1-H)} H + \frac{[-V''(1-H)]}{V'(1-H)} = 0$$

Hence $H < 0$ as $\frac{d}{dH} \frac{[-V''(1-H)]}{V'(1-H)} < 0$. Now, for utility functions involving choices under risk, $\frac{[-V''(1-H)]}{V'(1-H)}$ is called the (negative of the) "degree of absolute risk aversion," and is generally thought to be an increasing function of $H$, i.e., $\frac{d}{dH} \frac{[-V''(1-H)]}{V'(1-H)} > 0$. While this model is completely deterministic and is not concerned with choices under risk, it seems reasonable to assume that the $V(.)$ function has the property of increasing absolute risk aversion. If so, then of course $H < 0$ in Stage I, i.e., $H$ will rise at a decreasing rate.

While $H$ is the only variable whose behavior is observable in Stage I, and is thus the only variable of direct relevance to empirical work, the behavior during Stage I of $K$, the full-time wage, is of some theoretical interest. Note that since

$$\frac{d}{dH} \frac{[-V''(1-H)]}{V'(1-H)} > 0$$

in Stage I, it follows that, if $\frac{d}{dH} \frac{[-V''(1-H)]}{V'(1-H)} > 0$, $\frac{d}{dH} \frac{[-V''(1-H)]}{V'(1-H)} > 0$.
then the rate of growth of potential wages will be an increasing and concave function of time, i.e.:

\[
\frac{d}{dt} (\dot{K}) = (\alpha_n + \beta x) H > 0, \quad \frac{d^2}{dt^2} (\ddot{K}) = (\alpha_n + \beta x) H' < 0
\]

2. Stage II: Here, \(0 < H < 1\) and \(0 < F < 1\), i.e., Stage II is a period of "on the job training" or "earning and learning." Also, by (8-2') and (8-3'), when \(0 < F < 1\),

\[
\begin{align*}
[-V''(1-H)]H & = \mu_1 F + \mu_2 x + \mu_3 y G(F) + F[\mu_1 n - \mu_3 y ] - g'(F)] \\
[-g''(F)]F & = \mu_1 n - \mu_3 y [-g'(F)]
\end{align*}
\]

or, by (8-2') - (8-6'),

\[
\begin{align*}
[-V''(1-H)]H & = \rho V'(1-H) - (r + \sigma) \mu_3 y G(F) \quad (13-1) \\
[-g''(F)]F & = [-g'(F)][(r + \sigma) \rho F] - \rho n H G(F) \quad (13-2) \\
\mu_3 & = \mu_3 [(\alpha F + \beta x) H - y] \quad (13-3)
\end{align*}
\]

I shall now use equations (13) to plot the stationary surfaces for \(H, F\) and \(\mu_3\) in \(H, F, \mu_3\) space, i.e., the surfaces along which \(H, F\) and \(\mu_3\) are zero.

Consider first the \(H = 0\) surface. By (13-1), \(H = 0\) where \(\rho V'(1-H) < (r + \sigma) \mu_3 y G(F)\). So, when \(H = 0\) on this surface,

\[
\mu_3 y G(F) = \frac{\rho}{r + \sigma} V'(1)
\]

i.e., when \(H = 0, H = 0\) and \(F = 0\) in Stage II,

\[
\mu_3 = \frac{\rho}{r + \sigma} \frac{V'(1)}{y}
\]
On the other hand, note by (2-2) that when $H = 0$, \( \lim_{H \to 1} \mu_3 y g(F) = \infty \).

Moreover, $\ddot{H} > 0$ whenever $F > 1$. Finally, note that in general, when $H = 0$,

\[
\begin{align*}
\frac{\partial H}{\partial \mu_3} &= g(F) \frac{(r + \delta)}{\rho} \frac{\mu_3}{[-V''(1-H)]} > 0 \\
\frac{\partial^2 H}{\partial \mu_3^2} &= -g(F) \frac{(r + \delta)}{\rho} \frac{V''(1-H)}{[-V''(1-H)]^2} < 0 \\
\frac{\partial H}{\partial F} &= \frac{[-g'(F)]}{g(F)} \frac{(r + \delta)}{\rho} \frac{\mu_3}{[-V''(1-H)]} < 0 \\
\frac{\partial^2 H}{\partial F^2} &= \frac{[-g''(F)]}{g(F)} \frac{(r + \delta)}{\rho} \frac{\mu_3}{[-V''(1-H)]^2} + \frac{[-g'(F)]}{g(F)} \frac{(r + \delta)}{\rho} \frac{\mu_3}{[-V''(1-H)]^2} V''(1-H) \frac{\partial H}{\partial F} < 0 \quad 2/ \\
\frac{\partial F}{\partial \mu_3} &= \frac{[-g'(F)]}{g(F)} \mu_3 > 0 \\
\frac{\partial^2 F}{\partial \mu_3^2} &= \frac{\partial}{\partial F} \frac{[-g'(F)]}{g(F)} + \frac{[-g'(F)]}{g(F)} \frac{\partial \mu_3}{\partial F} > 0
\end{align*}
\]

Hence, when drawn in $H, F, \mu_3$ space, the $\ddot{H} = 0$ surface will look somewhat as shown in Figure 2a. Note that $H$ is rising, constant or falling at points above or below the $H = 0$ surface.

Now consider the $\ddot{F} = 0$ surface. By (13-2), $\ddot{F} > 0$ when $[-g'(F)](r + \delta) - (\alpha F + \delta x) H > 0$. Note that \( \frac{\partial H}{\partial \mu_3} = \frac{\partial F}{\partial \mu_3} = 0 \) always on the $\ddot{F} = 0$ surface,

\[1/\text{This, of course, assumes that } V''(1-H) > 0, \text{ i.e., that the rate at which } V'(1-H) \text{ diminishes falls as } H \text{ increases (that the marginal utility of leisure diminishes at a diminishing rate).} \]

\[2/\text{See n. 1 above.} \]
and that, when $F = 0$,

$$\frac{\partial H}{\partial F} = \frac{-g''(F)\left[ -g'(F) \right]}{\left[ -g'(F) \right] + \left[ -g'(F) \right]} \left( \frac{\partial f}{\partial F} + \frac{\partial f}{\partial x} \right)$$

So that $\frac{\partial H}{\partial F} \neq 1$ as $F \neq 1$. Finally, when $F = 1$, $F = 0$ when $H = \hat{H} = \frac{r+\delta}{\alpha t + \beta x}$;

and when $F = 0$ when $H = \hat{H} = \frac{\left[ -g'(0) \right] (r+\delta)}{\alpha t + \left[ -g'(0) \right] \beta x}$ (where, since $[-g'(0)] < 1$, $H < \hat{H}$). So, when drawn in $H, F, \mu_3$ space, the $F = 0$ surface will look somewhat as shown in Figure 2b. Note that $F$ is rising, constant or falling at points below, on or above the $F = 0$ surface.

Now consider the $\mu_3 = 0$ surface. By (13-3) (or, equivalently, by (8-6'), \mu_3 = 0 when $(\alpha t + \beta x)H < \gamma$, where $\gamma = r - \beta + \delta$. (Of course, when $\mu_3 = 0, \mu_3 = 0$ always.) Note that $\frac{\partial H}{\partial \mu_3} = \frac{\partial F}{\partial \mu_3} = 0$ always (except for $\mu_3 = 0$) on the $\mu_3 = 0$ surface, and that, when $\mu_3 = 0, H > 0$ always and

$$\frac{\partial H}{\partial F} = - \frac{\partial f}{\partial H} < 0$$

$$\frac{\partial^2 H}{\partial F^2} = - \frac{\left( \frac{\partial f}{\partial H} \right) (\partial f/\partial F)}{\partial H^2} < 0$$

Finally, when $F = 1, \mu_3 = 0$ when $H = \hat{H} = \frac{\gamma}{\alpha t + \beta x}$, and when $F = 0, \mu_3 = 0$

when $H = \hat{H} = \frac{\gamma}{\beta x}$. (Clearly, $\hat{H} \geq \hat{H}$, \hat{H} > \hat{H}$ and \hat{H} > \hat{H}$.) So when drawn in $H, F, \mu_3$ space, the $\mu_3 = 0$ surface will look as shown in Figure 2c. Note that $\mu_3$ is rising, constant or falling at points above, on or below the $\mu_3 = 0$ surface.

$\frac{1}{1}$ I shall assume for convenience that $\beta > \gamma$, i.e., that $H$ is in fact a feasible value (less than unity) for $H$. But this is not really essential to any of what follows.
\[ \mu_3 = \text{surface}. \]

Figure 3 plots all three surfaces -- \( \mu_3 = 0, F = 0 \) and \( H = 0 \) -- in the same \( H, F, \mu_3 \) space, and may be used to describe the behavior of \( H \) and \( F \) during Stages II. As noted earlier, the individual must pass through Stage II and into Stage III in order to satisfy the terminal or transversal condition (8-7'). At the moment he leaves Stage I and enters Stage II, \( F = 1 \), i.e., the beginning of Stage II in the \( H, F, \mu_3 \) space is simply the \( F \mu_3 \) plane; at the moment he leaves Stage II and enters Stage III, \( F = 0 \), i.e., the end of Stage II in \( H, F, \mu_3 \) space is the \( H \mu_3 \) plane. Thus the individual begins Stage II on the \( F \mu_3 \) plane and leaves it on the \( H \mu_3 \) plane. Moreover, inspection of the \( F \mu_3 \) plane reveals that the individual will never have \( H < \bar{H} \) upon entering Stage II, for otherwise \( F = 1 \) and is increasing. Now, when \( F = 1 \) and \( H > H \), (i.e., at the beginning of Stage I), \( H \) and \( \mu_3 \) are increasing and \( F \) is decreasing. But given that \( H > \bar{H} \) at the beginning of Stage II, how will the rest of Stage I unfold? Figure 3 suggests a variety of possible sequences. One possible pattern for Stage II would be for \( H \) and \( \mu_3 \) to go on increasing and \( F \) to go on increasing, i.e., for the individual to remain above each of the stationary surfaces throughout Stage II, even at Stage II's end. Another possible pattern would be for \( H \) to rise and \( F \) to fall throughout Stage II, and for \( \mu_3 \) to rise and then fall, i.e., for the individual to remain above the \( F = 0 \) and \( H = 0 \) surfaces throughout Stage II, and to be above the \( \mu_3 = 0 \) surface at first but then to intersect it and ultimately to go below it. A third possible pattern is the following: for \( F \) to fall constantly and for both \( H \) and \( \mu_3 \) to rise at first but then to fall, i.e., for the individual to remain above the \( F = 0 \) surface always, and to be above the \( \mu_3 = 0 \) and \( H = 0 \) surfaces at first (but ultimately intersecting and then
going below them). These three patterns do not involve any "cycling," i.e., the F and H variables do not either (a) increase, decrease and then increase again or (b) decrease, increase and then decrease again; but Figure 3 seems to indicate that other patterns which do involve cycling may be possible. For example, inspection of Figure 3 reveals that either of the following "cycling" patterns might be possible:

(a) the individual stays above the $\dot{F} = 0$ surface (has $\dot{F} < 0$) throughout Stage II, but has $\dot{H} > 0$, $\dot{\mu}_3 > 0$ at first; then has $\dot{H} < 0$, $\dot{\mu}_3 > 0$; then has $\dot{H} < 0$, $\dot{\mu}_3 < 0$; and finally has $\dot{H} > 0$, $\dot{\mu}_3 < 0$ for the rest of Stage II. Here H "cycles" (rises, falls and then rises again) while F declines monotonically.

(b) the individual starts Stage II with $\dot{H} > 0$, $\dot{F} < 0$ and $\dot{\mu}_3 > 0$, and then proceeds as follows: (i) $\dot{H} < 0$, $\dot{F} < 0$, $\dot{\mu}_3 > 0$; (ii) $\dot{H} < 0$, $\dot{F} < 0$, $\dot{\mu}_3 < 0$; (iii) $\dot{H} < 0$, $\dot{F} > 0$, $\dot{\mu}_3 < 0$; (iv) $\dot{H} > 0$, $\dot{F} > 0$, $\dot{\mu}_3 < 0$; (v) $\dot{H} > 0$, $\dot{F} > 0$, $\dot{\mu}_3 > 0$; and finally (vi) $\dot{H} > 0$, $\dot{F} < 0$, $\dot{\mu}_3 > 0$ through the rest of Stage II. Here both H and F cycle: H rises, falls and then rises again, while F falls, rises and then falls again.

But the possibilities for cycling to which Figure 3 seems to point are in fact more apparent than real, as I now show. First, consider the possibility of cycling in F. For F to cycle, F must at first fall, then rise and ultimately (since F must be zero at the end of Stage II) fall again. That is, for F to cycle, then (since $a \equiv -g'(F)$ increases whenever F increases) there must be moments in time $t_0$, $t_1$ and $t_2$ such that the following conditions obtain:
at $t_0$, \[ a = 0 \text{ and } a > 0 \]

at $t_1$, \[ a > 0, a = 0 \text{ and } a < 0 \]

at $t_2$, \[ a = 0 \text{ and } a < 0 \]

By (13-2),

\[ [-g''(F)]_F = \dot{a} = a[(r + \delta) - (\alpha nF + \beta x)H] - \alpha nHg(F) \]

Thus, by (13-2) there must exist some point in time $t_1$ such that

\[ a = a[(r + \delta) - (\alpha nF + \beta x)H] - [a(\alpha nF + \beta x) + \alpha nHg(F)]H = 0 \]

\[ a = -2(\alpha nF + \beta x)aH - \alpha nH aF - [a(\alpha nF + \beta x) + \alpha nHg(F)]H < 0 \]

But note from the above that when $a = 0$,

\[ \dot{H} = \frac{\dot{a}[(r + \delta) - (\alpha nF + \beta x)H]}{a(\alpha nF + \beta x) + \alpha nHg(F)} = \frac{(\dot{a})^2 + \dot{a} \alpha nHg(F)}{a^2(\alpha nF + \beta x) + a \alpha nHg(F)} > 0 \]

so that, when $a = 0$, as at $t_1$, one has, upon taking the time derivative of $H$, \[ \dot{H} = \frac{-2(\alpha nF + \beta x)aH + \alpha nH aF}{a(\alpha nF + \beta x) + \alpha nHg(F)} \]

However, upon substituting this expression for $\dot{H}$ into the expression for $a = 0$, one finds that, when $a = 0$, $\dot{a} = 0$ also. In other words, cycling in $F$ is never possible in an optimal plan; $F$ will never fall, rise and then fall. Indeed, it is now clear that $F$ must always fall monotonically in Stage II, for if $F$ ever fell and then rose it would go on rising.
without limit. 1

Now consider whether cycles in \( H \) are possible. Since it is clear
that \( H \) must rise at first, it follows that for cycling to occur \( H \) must rise,
fall and then rise again. (It might then either go on rising or else fall
once more, but this is of no particular concern; the important point is
that it must rise, fall and then rise again at least once for a cycle to
occur.) That is, for \( H \) to cycle then (since \( b = V'(1-H) \) increases whenever
\( H \) increases) there must be moments in time \( t_3, t_4 \) and \( t_5 \) such that
the following conditions obtain:

at \( t_3 \), \( \dot{b} = 0 \) and \( \ddot{b} < 0 \)

at \( t_4 \), \( \dot{b} < 0 \), \( \ddot{b} = 0 \) and \( \dddot{b} > 0 \)

at \( t_5 \), \( \dot{b} = 0 \) and \( \dddot{b} > 0 \)

By (13-1),

\[
[-V''(1-H)]H = \dot{b} = \rho b - (r+\delta)g(F)\mu_3 y
\]

Thus, by (13-1), there must exist some point in time \( t_3 \) such that

\[
\dot{b} = \rho b + (r+\delta)[-g'(F)]\mu_3 y F - (r+\delta)g(F)\mu_3 y [(\alpha nF + \delta x)H - \gamma] = 0
\]

\[
\dddot{b} = (r+\delta)\mu_3 y [-g''(F)](y')^2 + 2[g'(F)][(\alpha nF + \delta x)H - \gamma]^2
\]

\[
+ [-g'(F)]F - g(F)[(\alpha nF + \delta x)H - \gamma] \dot{y}
\]

\[
- g(F)[(\alpha nF + \delta x)H + \sigma n H F] > 0
\]

1Lest the statement, "cycling in \( F \) is never possible," be misunderstood,
let me stress that this simply means that cycling in \( F \) could never be
optimal, for it would never permit the individual to satisfy one of the
conditions for a maximum, i.e., the transversality condition.
But note from the above that when $b = 0$,

$$
\begin{align*}
F &= \frac{(r+\delta)\mu_2 y}{(r+\delta)[-g'(F)]\mu_2 y} \\
&= \frac{(r+\delta)\mu_2 y}{(r+\delta)[-g'(F)]\mu_2 y}
\end{align*}
$$

So that, when $b = 0$, at $t_4$, one has, upon taking the time derivative of $F$,

$$
F_{b=0} = \frac{1}{[g'(F)]}
$$

$$
\times \left\{ [-g''(F)](F)^2 - 2[-g'(F)][(\alpha F + \beta x)H - \gamma]F \\
+ g(F)[((\alpha F + \beta x)H - \gamma]^2 + g(F)[(\alpha F + \beta x)H + \alpha HP] \right\}
$$

so that, on substituting the expression for $F_{b=0}$ into the expression for $b = 0$, one finds that, when $b = 0$, $b = 0$ also. In other words, cycling in $H$ is never possible in an optimal plan; $H$ will never rise, fall and then rise. The only possible patterns for $H$ in Stage II are for $H$ either to rise constantly or to rise and then fall.

Hence, while $H$ always rises at first and may either go on rising or else, after a while, fall, $F$ always falls -- from unity to zero -- in Stage II. This means that while $FH$ (total time devoted to training) may rise at first (if $H$ rises by more than $F$ falls), it must ultimately fall.

\textsuperscript{1}Note from Figure 3 that the stationary state -- where $H = \dot{F} = \dot{\mu}_2 = 0$ and the singular solution occurs, at $H = H^*$, $F = F^*$ -- will never be achieved, since $H$ and $F$ do not cycle and since trajectories in most regions of the $H$, $F$, $\mu_2$ space of Figure 3 lead away from this particular point. (For example, trajectories above all three stationary surfaces, i.e., in the upper part of Figure 3, point "up" (in the $H$ dimension) and "in" (in the $F$, $\mu_2$ plane), yet in this region $H$ would have to fall to reach the stationary state.)
Now consider the behavior of $H$ in more detail. By (13-1),
\[
\frac{-V'(1-H)}{V'(1-H)} + (r + \delta) g(F) \frac{\mu_3 Y}{V'(1-H)} = \rho
\]
So that
\[
\frac{d}{dH} \left[ \frac{-V''(1-H)}{V'(1-H)} \right] (H)^2 + \frac{-V''(1-H)}{V'(1-H)} H - (r + \delta) \frac{[-g'(F)] \mu_3 Y}{V'(1-H)^2} \\
[\frac{V'(1-H)^2}{V'(1-H)}]^2
\]
\[
= \frac{d}{dH} \left[ \frac{-V''(1-H)}{V'(1-H)} \right] (H)^2 + \frac{-V''(1-H)}{V'(1-H)} H - (r + \delta) \frac{[-g'(F)] \mu_3 Y}{V'(1-H)^2} \\
\frac{[\alpha \eta F + \beta x H - \gamma] V'(1-H) - \rho V'(1-H) + (r + \delta) \mu_3 Y g(F)}{[V'(1-H)]^2} = 0
\]
I.e.,
\[
\frac{d}{dH} \left[ \frac{-V''(1-H)}{V'(1-H)} \right] (H)^2 + \frac{-V''(1-H)}{V'(1-H)} H - (r + \delta) \frac{[-g'(F)] \mu_3 Y}{V'(1-H)^2} \\
\frac{(r + \delta) \mu_3 Y g(F)}{V'(1-H)^2} \left[ \alpha \eta F V'(1-H) + (r + \delta) \mu_3 Y g(F) + V'(1-H)[\beta x - (r + \delta)] \right] = 0
\]
Thus, if $\beta x > r + \delta$, then $H$ is necessarily always negative and the profile of $H$ in Stage II is concave; but if $\beta x < r + \delta$, it could in principle be zero or positive.\(^1\)
\(^1\)Recall that $F = 0$ always in Stage II and that I have assumed that $dH \left[ \frac{-V''(1-H)}{V'(1-H)} \right] / (\text{"increasing risk aversion")}.\)
\(^2\)If $\beta x > r + \delta$, then what I have called the "rate of marginal costs" (see Part IIA) is less than what may be called the "full-time" rate of marginal benefits of investment in experience only, $\beta x$. (Note from (12-2') above that $\beta x > r + \delta$ is also a strong sufficient condition for $H < 1$.\)
Now consider the behavior of wages in Stage II. Since
\[ W_0 = yKg(F) = yN^\alpha X^\beta g(F), \]
the percentage rate of change in the observed wage is
\[
\frac{\dot{W}_0}{W_0} = \alpha \frac{\dot{N}}{N} + \beta \frac{\dot{X}}{X} + \left[ \frac{\dot{g}(F)}{g(F)} \right] (\dot{F})
\]
\[ = \alpha nFH + \beta xH - \delta + \left[ \frac{\dot{g}'(F)}{g(F)} \right] (\dot{F}) \]
while the percentage rate of change in \( W_P = yK = yN^\alpha X^\beta \), the potential wage, is
\[
\frac{\dot{W}_P}{W_P} = \alpha nFH + \beta xH - \delta
\]
Since at the beginning of Stage II \( H = H = (r+\delta)/(\alpha n+\beta x) \) and \( F = 1 \), both \( W_0 \) and \( W_P \) must be rising at the beginning of Stage II. (Indeed, since \( F = 1 \).

At the beginning of Stage II, \( W_0 = 0 \) for the first instant of Stage II.)

Also, since \( \dot{F} > 0 \), the peak (if any) in observed wages must follow the peak (if any) in the potential wage (i.e., in human capital) -- simply because, even after net investments fall to zero and the potential wage begins to fall, the observed wage per hour at work can go on rising as the fraction of each working hour devoted to training goes on falling. However, it is also clear that \( W_0 \) and \( W_P \) could always rise during Stage II, i.e., observed and potential wages could always increase -- so long as \( H > \delta/\beta x \) throughout Stage II. (In a "pure training" model -- i.e., one which assumes no experience and thus that \( \beta = x = 0 \) -- \( \frac{\dot{W}_0}{W_0} = \frac{\dot{W}_P}{W_P} = -\alpha n/N < 0 \) at the end of Stage II, i.e., observed and full-time wages must fall by the end of Stage II. In the present model, both could rise throughout Stage II
if hours of work (and hence net investments in experience) are sufficiently large to offset the rate of decay in training.

Next, consider the behavior of earnings, \( Y = \gamma^{xg(F)}H \), during Stage II. Since \( \dot{Y} = \left( \dot{W}_O/W_0 \right) + \left( \dot{H}/H \right) \), at the beginning of Stage II, when \( W_0 \) and \( H \) are rising, \( Y \) is certainly rising; and \( Y \) could either (i) go on growing throughout the rest of Stage II or (ii) reach a peak and then decline.

Finally, consider the relationship between observed and potential wages, on the one hand, and work hours, on the other, during Stage II. From the above,

\[
\frac{\dot{W}_0}{W_0} = \alpha x H + \beta x H - \delta
\]

\[
\frac{\dot{W}_p}{W_p} = \alpha x H + \beta x H - \delta + \left[ \frac{f}{g(F)} \right] \left( \frac{\Delta g(F)}{g(F)} \right)
\]

\[
\left[ -\nu' (1-H) \right] \dot{H} = \rho v'(1-H) - (r+\delta) \mu_3 y \ g(F)
\]

Note that since, as Stage II starts, \( F = 1, g(F) = 0 \) and \( H > H = (r+\delta) / (\alpha x + \beta x) \), \( W_p, W_0 \) and \( H \) are all rising as Stage II starts. Now, the set of \( F, H \) combinations along which \( W_0 = 0 \) is given by \( \alpha x H + \beta x H = \delta \). In terms of the \( F, H, \mu_3 \) space of Figure 3, this \( W_0 = 0 \) surface has \( H = \delta / (\beta x) \) \( (\frac{\Delta H}{\delta} = \gamma / \beta x) \) when \( F = 0 \), and \( H = \delta / (\alpha x + \beta x) \) \( (\frac{\Delta H}{\delta} = \gamma / (\alpha x + \beta x)) \) when \( F = 1 \). Moreover, the surface is the same for all values of \( \mu_3 \), and has

\[\text{All that is necessary is that } (\dot{W}_0/W_0) + (\dot{H}/H) > 0; \text{ i.e., even if } H \text{ began falling, } W_0 \text{ could go on rising enough to keep } Y \text{ rising provided } H \text{ remained "sufficiently" above } \delta / \beta x.\]
\[
\frac{\partial H}{\partial F} = -\frac{\partial H_{\phi}}{\partial (\text{ln} F + \text{ln} X)} < 0 \quad \text{and} \quad \frac{\partial^2 H}{\partial F^2} = \frac{-\text{Sign} \left( \frac{\partial H_{\phi}}{\partial F} \right)}{(\text{ln} F + \text{ln} X)^2} > 0
\]

In other words, the \( W_P = 0 \) surface has a negative slope and lies entirely below the \( \mu_3 = 0 \) surface. In view of this and the results given above on the behavior of \( H \), it is evident that a wide variety of relationships between \( H \) and \( W_P \) could in principle hold, including, to list only a few, the following:

(i) \( H \) and \( W_P \) rise monotonically

(ii) \( H \) rises monotonically, while \( W_P \) rises and then falls

(iii) \( H \) and \( W_P \) rise; then \( H \) falls while \( W_P \) goes on rising; then \( H \) falls and \( W_P \) falls

(iv) \( H \) and \( W_P \) rise; then \( H \) and \( W_P \) fall

As Blinder and Weiss remark,\(^{1/} \) "It cannot be stressed too much that these contrasting behavior patterns have absolutely nothing to do with competing income and 'substitution' effects, although cross-sectional studies of labor supply might possibly confound the two phenomena."

While \( W_P = \gamma X \) is not observable, the fact that during Stage II \( W_P \) and \( H \) may change in a variety of ways, such as those noted above, is not without theoretical interest. Speaking of their model, Chen and Becker declare (p. 12):

\(^{1/}\) While made in the course of a discussion of the Stage III of a "pure training" model, their comment is quite germane to the present discussion.
time and an income effect away from it. The income effect is often supposed to dominate and cause a "backward-bending" supply curve of labor. In our analysis there is no income or wealth effect because all changes in wealth are perfectly foreseen [since they are all part of the individual's lifetime optimization strategy]. Hence a rise in wage rates with age generates only substitution effects, and the supply curve of labor would be positively sloped.

(A footnote adds, "This conclusion is not a negation of the observation that a parametric shift in the wage profile generates both income and substitution effects.") While these remarks refer to a variant of their model in which wages are exogenous and in which the interest rate, r, and rate of time preference, ρ, are zero, they later argue that exactly the same proposition holds for endogenous wages when \( r = ρ \). Specifically, they maintain that, in this case, for single persons,

\[
\frac{\dot{H}}{H} = a_1 \frac{\dot{K}}{K}, \quad a_1 > 0
\]

(see pp. 16, 22 and 93-95). In contrast, in the present model,

\[
\frac{\dot{K}}{K} = \frac{\dot{W}_p}{W_p} = \pi nPH + βxH - \delta
\]

\[
\frac{\dot{H}}{H} = \left( \frac{\rho}{K} \right)^{\gamma} - \left[ (r+δ)g(y)/H^\gamma (y/p)K u''(c) \right]
\]

i.e., the rates of growth of K and H depend on the levels of H and K.\(^{1/}\)

Hence, there is no reason to believe that the Ghez-Becker propositions

\(^{1/}\)Note that, in this model, these propositions do not depend on the assumption that experience is a form of human capital. A "pure training" version of this model would simply assume \( β = x = 0 \), which assumptions would not alter the propositions made here about the relationship(s) between \( \dot{H}/H \) and \( \dot{K}/K \). (For more on a "pure training" model of this kind, see Blinder and Weiss.)
on the relationship between \( W \) and \( H \) are particularly robust, or -- more generally -- that the notions of income and substitution effects, which are essentially concerned with displacements from equilibrium, are especially relevant to equilibrium dynamics, i.e., to the analysis of an equilibrium lifetime optimization strategy.

Next consider the relationship between observed wages, \( W_0 \), and hours of work, \( H \). The surface along which \( \dot{W}_0 = 0 \) is given by

\[
\alpha \dot{W}_0 + \beta x_2 \dot{H} - \delta + \frac{\left( -\mu'(\gamma) \right)}{\gamma} \left( -\dot{F} \right) = 0
\]

So in terms of Figure 3 the \( \dot{W}_0 = 0 \) surface is independent of \( \mu_3 \) and always lies below the \( \dot{W}_2 = 0 \) surface. Unless additional assumptions are introduced (e.g., on the magnitude of \( g'(\gamma) \)), not very much more can be said about the characteristics of \( \dot{W}_0 = 0 \) surface (e.g., its slope in the \( H, P \) plane), but this is of no great consequence, for it is already evident from the discussion of \( W \) and \( H \) that in principle a wide variety of relationship between \( W_0 \) and \( H \) might be observed during Stage II.

3. Stage III. So much for Stage II. In Stage III, \( F = 0 \) and \( 0 < H < 1 \); Stage III is a period of "pure work." Also, by (8-2'),

\[
[-\gamma'(1-H)] \dot{H} = \mu_2 x + \mu_3 y
\]

\[1\] In fairness to Ghez and Becker, whose book is full of interesting ideas and has, quite rightly, had a considerable impact on research in the area of life-cycle behavior, I should stress that their remarks on the relationship between \( H/H \) and \( K/K \) refer not to the kind of model presented here but rather to a somewhat different one in which (for example) \( K = N \) and \( N \) affects only \( W \) and does not affect \( H \). However, the reader is given the strong impression that as a general rule \( H/H \) and \( K/K \) (or \( N/H \)) will be positively related when \( r = p \) (See pp. 16, 22 and 93-95); and it is this proposition which I challenge.
or, by (8-2'), (8-5') and (8-6'),

\[ \dot{\mu}_2 = \rho \mu_2 - 3 \mu_3 y H \]  
(14-1)

\[ \dot{\mu}_3 = \mu_3 (3 \xi H - \gamma) \]  
(14-2)

\[ [-V''(1-H)]H = \rho V' (1-H) - (r+\delta) \mu_3 y \]  
(14-3)

in Stage III.

To discuss Stage III I find it convenient to represent in the \( \mu_1, \mu_2, \mu_3 \) space of Figure 5 the surfaces along which \( \dot{\mu}_1, \dot{\mu}_2, \dot{\mu}_3 \) and \( \ddot{H} \) are zero when \( F = 0 \) and \( 0 < H < 1 \). (Figure 5 reproduces from Figure 1 the area constituting Stage III)

Consider first the \( \dot{\mu}_1 = 0 \) surface. By (8-1'), in Stage III

\[ \dot{\mu}_1 = 0 \text{ as } \rho \mu_1 = \alpha \mu_3 y H. \]  
Hence, along the \( \dot{\mu}_1 = 0 \) surface, \( \mu_1 \to 0 \) as \( H \to 0 \)
or as \( \mu_3 \to 0 \) and (by (8-2')) as \( \mu_2 y + \mu_3 y = V'(1) \). Moreover, along the \( \dot{\mu}_1 = 0 \) surface,

\[ \rho d\mu_1 = \alpha \mu_3 y dH + \alpha y H d\mu_3 \]

\[ [-V''(1-H)]dH = x d\mu_2 + y d\mu_3 \]

So, along the \( \dot{\mu}_1 = 0 \) surface,

\[ \rho d\mu_1 = \alpha \frac{\mu_3 y x}{(-V''')} d\mu_2 + \alpha y \left[ H + \frac{\mu_3 y}{(-V''')} \right] d\mu_3 \]

So, on this surface,
\[
\frac{\partial \mu_1}{\partial \mu_2} = \frac{\mu_3 \gamma x}{\rho (-\gamma' y)} > 0; \quad \frac{\partial \mu_1}{\partial \mu_3} = \frac{\alpha y}{\rho} \left[ H + \frac{\mu_3 \gamma y}{(-\gamma' y)} \right] > 0;
\]

\[
\frac{\partial \mu_2}{\partial \mu_3} = -\frac{\rho (-\gamma' y) \mu_3 y}{\mu_3 x} < 0
\]

Finally, at the boundary between Stage III and Stage II, \( \mu_1 = y \mu_3 [\gamma' (0)] \) and (when \( \mu_1 = 0 \)) \( \mu_1 = \alpha y \mu_3 y H \). That is, when the \( \mu_1 = 0 \) surface intersects the Stage III - Stage II boundary,

\[
\mu_1 = \alpha y \mu_3 y H
\]

\[
\mu_1 = \gamma' (0) \mu_3 y
\]

\[
\therefore H = \frac{\gamma}{\alpha} = \frac{\rho (-\gamma' (0))}{\alpha y}
\]

and so, along this intersection,

\[
V'(1-H) = \mu_2 x + \mu_3 y
\]

So the \( \mu_1 = 0 \) surface in Stage III looks as shown in Figure 4a.

Next consider the \( H = 0 \) surface. \( H \leq 0 \) as \( H' = (1-H) \leq (r+\delta) \mu_3 y \) or (since \( V'(1-H) = \mu_2 x + \mu_3 y \) in Stage III) as \( \mu_2 x \leq \mu_3 y \). Hence, in the \( \mu_1, \mu_2, \mu_3 \) space of Figure 4b, the \( H = 0 \) surface is independent of \( \mu_1 \) and has a slope in the \( \mu_2, \mu_3 \) plane of \( \frac{\partial \mu_2}{\partial \mu_3} = \frac{\gamma y}{\rho x} \). (When \( H = 0 \) and \( H = 0 \),

\[
\mu_3 = \frac{2}{(r+\delta)} \frac{V'(1)}{y} \quad \text{and} \quad \mu_2 = \frac{y}{r+\delta} \frac{V'(1)}{x}
\]

Now consider the \( \mu_2 = 0 \) surface. By (11-1), in Stage III \( \mu_2 \leq 0 \) as
\[ p_{\mu_2} \geq \beta_{\mu_3} yH; \] so, along the \( \mu_2 = 0 \) surface, \( \mu_2 \rightarrow \infty \) as either \( \mu_3 \) or \( H \rightarrow 0 \) and (by (8-2')) as \( \mu_2 x + \mu_3 y = V'(l) \). Also, along the \( \mu_2 = 0 \) surface, with \( 0 < H < 1 \),

\[ \rho \mu_2 = \beta_{\mu_3} y \ dH + \beta yH \ d\mu_3 \]

\[ [\mu_2 V'(1-H)]^H = x \ d\mu_2 + y d\mu_3 \]

So, along the \( \mu_2 = 0 \) surface,

\[ [\rho - \beta_{\mu_3 yx} \left( \frac{V'}{V'} \right) ] \ d\mu_2 = \beta y c \left( \frac{\mu_3 y}{V'(1-H)} \right) + H \] \ d\mu_3

or

\[ [\frac{V'(1-H)\beta_{\mu_3 y} y}{-V'(1-H)}]^{\mu_2} \left[ \frac{V'}{V'(1-H)} + (x/p)X \ U'(c) - 1 \right] d\mu_2 \]

\[ = \beta y \left[ \frac{\mu_3 y}{V'(1-H)} \right] + H \] \ d\mu_3

So, along the \( \mu_2 = 0 \) surface, \( \frac{\partial \mu_2}{\partial \mu_1} = \frac{\partial \mu_3}{\partial \mu_1} = 0 \) and, by the sufficiency conditions (9), \( \frac{\partial \mu_2}{\partial \mu_3} > 0 \). So in Stage III the \( \mu_2 = 0 \) surface looks as shown in Figure 4c.

Now consider the \( \mu_3 = 0 \) surface by (14-2), \( \mu_3 \geq 0 \) as \( H \geq \beta x \)

(of course, \( \mu_3 = 0 \) wherever \( \mu_3 = 0 \)). Since \( V'(1-H) = \mu_2 x + \mu_3 y \), it follows that, along the \( \mu_3 = 0 \) surface, \( \mu_2 x + \mu_3 y = V'(1-H) \). So the \( \mu_3 = 0 \) surface is independent of \( \mu_1 \) and so, in the \( \mu_2 \mu_3 \) plane,

\[ \frac{\partial \mu_2}{\partial \mu_3} = \frac{V}{x} \]
when $\mu_3 = 0$ in Stage III. Thus the $\mu_3 = 0$ surface looks as shown in Figure 4d.

Now consider Figure 5, which brings together the $\mu_1 = 0$, $\mu_2 = 0$, $\mu_3 = 0$ and $H = 0$ surfaces in Stage III. It is evident that, as the individual passes from Stage II to Stage III, $\mu_1$ must be falling. (Geometrically, the explanation is that the only way to go "down" below the $F = 0$ line is for $\mu_1$ to fall; in economic terms, the explanation is that the only reason one would have for moving from a situation of on-the-job training ($0 < F < 1$) to one of "pure work" ($F = 0$) would be that the shadow value of the stock of training ($\mu_1$) was falling.) Hence, at least at first, $\mu_1$ is falling in Stage III. Moreover, it is evident that if $\mu_1$ is ever rising in Stage IV the individual will leave Stage III, either for Stage IV or Stage II. I shall refer to this kind of behavior as "cycling," since (after the individual has left Stage III) it must ultimately entail an increase in $F$. For the time being, I shall assume that no cycling of this kind occurs. But if no such cycling of this kind occurs -- if the individual does not go through Stage IV (or once again) through Stage II -- then in view of (8-7') the individual's life cycle must end somewhere on the $\mu_3$ axis, with $\mu_1 = \mu_2 = 0$, $\mu_3 \geq \frac{V'(1)}{y}$, and $H > 0$.\(^{2}\) Hence, under the

\(^{1}\)Note that since $\mu_1 > 0$ always in Stage IV (except when $\mu_1 = \mu_2 = 0$, or when $\mu_1 = \mu_2 = 0$ and $\mu_3 = \frac{V'(1)}{y}$) anyone who enters Stage IV will ultimately go to Stage I, with $H > 0$ and $F = 1$.

\(^{2}\)Actually, if the individual ever has $\mu_1 = \mu_2 = 0$ and $\mu_3 = \frac{V'(1)}{y}$, then, by (8-2') and (8-4') -- (8-6'), $H = 0$, $\mu_1 = 0$, $\mu_2 = 0$ and $\mu_3 = -\mu_2 < 0$. In this case $H$ will remain at zero and the individual will enter Stage IV, moving towards the origin along the $\mu_2$ axis. At the origin, $H = 0$ and $\mu_1 = \mu_2 = \mu_3 = \mu_2 = \mu_3 = 0$, i.e., there is a singular solution to (8-4') -- (8-6') with zero values of the state variables. This behavior may appropriately be called "retirement," for in this kind of behavior $H$ falls to zero and stays at zero ever after.
assumption of no cycling from Stage III to Stage II or Stage IV (i.e., the assumption that \( \dot{\mu}_1 < 0 \) always in Stage III), one may analyze Stage III by looking only at equations (14), and can ignore \( \dot{\mu}_1 \) and \( \dot{\mu}_3 \).

With this in mind, I now turn to Figure 6, which reduces Stages III to two dimensions, i.e., the \( \mu_2, \mu_3 \) plane. (Recall that the \( \dot{\mu}_2 = 0, \dot{\mu}_3 = 0 \) and \( H = 0 \) surfaces depend only on \( \mu_2 \) and \( \mu_3 \), and that it is assumed that \( \dot{\mu}_1 < 0 \) always. Now, prior to time \( T, \mu_1 > 0 \); but since the life-cycle terminates on the \( \mu_3 \) axis with \( \mu_3 = \frac{y' (l)}{y} \), all trajectories must go to that axis, i.e., \( \mu_1 = 0 \) at time \( T \). Figure 6 must of course be read with this in mind.)

Figure 6 reveals that if the individual leaves Stage II (enters Stage III) with \( H > 0 \), he must eventually have \( \dot{H} = 0 \) and then \( H < 0 \) at some point in Stage III even if he has \( H > 0 \) at first in Stage III, i.e., he must eventually cross the \( H = 0 \) line (and thus have \( H < 0 \) at the end of Stage III) in order to satisfy the terminal condition (6-7'). Likewise, individuals who have \( H < 0 \) on leaving Stage II may simply go directly to the \( \mu_3 \) axis, and thus have \( H < 0 \) throughout Stage III. However, Figure 6 also suggests that cycles in \( H \) may be possible. For example, an individual might start in part 6 of Stage III (i.e., the area above the \( \dot{\mu}_3 = 0 \) line, below the \( \dot{\mu}_3 = 0 \) line and to the right of the \( \mu_3 \) axis), with \( H < 0 \), and then move counter-clockwise through parts 1-5, terminating on the \( \mu_3 \) axis in part 5 (i.e., above the \( \dot{\mu}_3 = 0 \) axis and above the \( \dot{H} = 0 \) axis). Here \( H \) falls (while the individual is in parts 6 and 1 of Stage III), then rises (while the individual is in parts 2-4 of Stage III) and ultimately falls once again (in part 5). Indeed, if the individual starts Stage III in part 1,
his labor supply must cycle, i.e., he must cross the $H = 0$ locus at least twice in order to get to the $\mu_3$ axis.

I have shown that cycling in $H$ or $F$ within Stage II is never possible in an optimal plan, and have been assuming that cycling from Stage III to Stage II or Stage IV does not occur. Is the implication of Figure 6, to the effect that cycling within Stage III may occur, in fact correct? It is evident that for cycling to occur within Stage III, the individual's trajectory must cross the $H = 0$ locus at least twice, once when $H$ has been falling and then when $H$ has been rising. Now, $a = v'(1-H)$ is rising whenever $H$ is rising. Thus, if cycles in $H$ are to occur within Stage III, there must exist three points in time $t_7$, $t_8$ and $t_9$ such that

- at $t_7$, $\dot{a} = 0$ and $a > 0$
- at $t_8$, $\ddot{a} > 0$, $\dot{a} = 0$ and $a < 0$
- at $t_9$, $\dot{a} = 0$ and $a < 0$

Hence, if cycles are to occur, there must exist some point in time $t_8$ such that, by (14-3),

$$a = p v'(1-H) - (r + \delta) \mu_3 y > 0$$

$$\ddot{a} = p \ddot{a} - (r + \delta) \dot{\mu}_3 y = p \ddot{a} - (r + \delta) \mu_3 y \left[ \frac{p}{x_1 H} v - v \right] = 0$$

$$\dot{a} = p \dot{a} - (r + \delta) \mu_3 y$$

$$= -(r + \delta) \left[ \mu_3 y \frac{p}{x_1 H} \dot{v} + \mu_3 y \left( \frac{p}{x_1 H} - v \right)^2 \right] < 0$$
But since $H > 0$ when $\dot{a} > 0$, it certainly seems possible to have $a < 0$ when $\dot{a} > 0$ and $\ddot{a} = 0$, i.e., cycles in $H$ within Stage III cannot be ruled out a priori. That is, $H$ may well cycle within Stage III; certainly, none of the assumptions introduced thus far appears to preclude such behavior.

This result -- to the effect that $H$ may cycle in Stage III -- makes an interesting contrast with the Blinder-Weiss (pure training) and Weiss (pure experience) models, for in both of these cycles in $H$ are impossible in Stage III. In the Blinder-Weiss model, cycling is impossible at least in part, it seems, because theirs is a pure training model, i.e., it assumes that $\beta = \alpha = 0$ and thus (by (8-6')) that the equivalent of $\mu_3$ is always negative and equal to $\gamma \mu_3$ in Stage III. Hence $\dot{\alpha}$ would always be positive if $\alpha$ were ever positive. (Indeed, when $x = 0$, then $V'(1-H) = \mu_3 y$ by (8-2'); and since $\mu_3$ always falls, then, as Blinder and Weiss show, $H$ must always fall also.) In the Weiss pure experience model cycling appears to be impossible at least in part because in contrast with the assumption adopted here that $\dot{X} = X(H - \dot{H})$, Weiss assumes that $\dot{X} = xH - \dot{X}$. This, together with the sufficiency condition (9-2)$\dagger$, turns out to imply$\ddagger$ that $H$ can never cycle when $F = 0$ (as it always will be in a pure experience model, which in effect assumes that $\sigma = n = 0$).

Now, whether or not cycling in $H$ occurs is of course an empirical question; and as Weiss has stressed, cycles in $H$ do seem to occur towards

$\dagger$Which is exactly the same in Weiss' model as it is in this one.
$\ddagger$See Weiss, pp. 1312-1315.
the end of life, which may be equated with Stage III. Thus the fact that the present model allows for such cycling seems to be a strength, not a weakness.

So much for the behavior of \( H \) in Stage III. To determine the behavior of wage rates, note first that since \( F = 0 \) in Stage III, the levels and rates of change in the observed wage and the full-time or potential wage are equal, i.e.,

\[
W_p = W_0 = yK
\]

\[
\dot{W}_p = \dot{W}_0 = \frac{\dot{K}}{K} = \beta xH - \delta
\]

But for \( \mu_3 > 0 \), \( \mu_3 \gtrless 0 \) as \( \beta xH \gtrless \gamma = r - \rho + \delta \). So in terms of Figure 6 the locus of \( W_p, \mu_3 \) combinations along which \( \dot{W}_p = \dot{W}_0 = \dot{K} = 0 \) lies parallel to and above and to the right of the \( \mu_3 = 0 \) locus; and \( W_p \) (or \( W_0 \) or \( K \)) is rising, constant or falling on points above, on or below this \( W_p = W_0 = K = 0 \) locus. Clearly, in Stage III as in Stage II, \( H \) and \( W_p \) (or \( W_0 \)) could move either in the opposite direction or in the same direction.

IID. Empirical Predictions of the Model for a "Normal" Life-Cycle

It is now time to sum up the results concerning the observable variables \( H, W_0 \) and \( Y \) which have been obtained thus far. In what follows, I shall restrict my attention to a "normal" life-cycle, which in view of the foregoing discussion I define as follows:

\( \frac{1}{2} \) For a brief discussion of empirical evidence on cycling, see Weiss, p.1311, esp. n. 1. For some cross-section age-labor supply profiles which appear to constitute evidence of cycling, see Ghez and Becker, pp. 85-90.

\( \frac{2}{2} \) Of course, the fact that cycling cannot occur in the Blinder-Weiss model means that it contains a strong, testable proposition. But while it is in this sense "strong," it is also "weak" in the sense that cycles do seem to be observed empirically.

\( \frac{3}{2} \) The following behavior patterns are "normal" in that they seem to characterize fairly well the behavior of men over the life cycle. But they certainly do not characterize the behavior of women very well.
(i) a period of specialization in training (i.e., full-time schooling, with \( F = 1 \)), followed by a period of on-the-job training with \( 0 < F < 1 \) and then a period of "pure work" (i.e., a period with \( F = 0 \)); and

(ii) no cycling from Stage III to Stage II or (except for "retirement," with \( H = 0 \) from some \( t_2 < T \) to \( t = T \)) Stage IV; and at most one cycle within Stage III.  

(Recall that cycling in either \( F \) or \( H \) within Stage II can never be part of an optimal plan in this model; and that any life-cycle which begins in Stage I with a period of full-time schooling must be followed by both a Stage II and a Stage III.)

Note that, in Stage II, when \( 0 < F < 1 \),

\[ \mu_1 n = \mu_3 y [-g'(F)] \]

Since \( \mu_3 \) is strictly positive and \( \mu_1, \mu_3 \) and \([-g'(F)]\) are continuous, \( F \) is also continuous both within Stage II and across the Stage I - II and II - III boundaries, i.e.,

\[ \lim_{F \to 1} (\mu_1 n / \mu_3 y) = [-g'(1)] \]

\[ \lim_{F \to 0} (\mu_1 n / \mu_3 y) = [-g'(0)] \]

Also, within Stage II,

The life-cycle behavior of women seems to involve several periods during which \( H > 0 \) and then \( H = 0 \), which might be interpreted as cycling from Stage III to Stage II. See, for example, Mincer and Solomon Polacheck.
\[-g''(F)\] \[F = [-g'(F)] [(r + \delta) - (\omega_n + \beta x) H] - \omega H g(F)\]

I show below that \(H\) is a continuous function of time. In view of this and the fact that \(F, g(F), [-g'(F)]\) and \([-g''(F)]\) are continuous, \(F\) is also continuous within Stage II. However,

\[
\lim_{F \to 1} F = \frac{[-g'(1)]}{[-g''(1)]} [(r + \delta) - (\omega_n + \beta x) \lim_{F \to 1} H] \leq 0
\]

since \(H = (r + \delta)/(\omega_n + \beta x)\) when \(F = 1\) in Stage II, so \(F\) may not be continuous at the beginning of Stage II. Likewise,

\[
\lim_{F \to 0} F = \frac{[-g'(0)]}{[-g''(0)]} [(r + \delta) - (\omega_n + \beta x) \lim_{F \to 0} H] + \frac{\omega_n}{[-g'(0)]} \lim_{F \to 0} H
\]

which may or may not equal zero; so \(F\) may not be continuous at the end of Stage II. Finally, any statements about the concavity of the time profile of \(F\) within Stage II must, obviously, depend on special assumptions on (e.g.) the sign and magnitude of \([-g''(F)]\); in other words, nothing can be said about \(F\) on the basis of the assumptions made thus far.

I now consider the implications of the model for the behavior of \(H, W_0\) and \(Y\) over the life cycle.

1. "Labor Supply,\(^{17}\) H: As noted in the discussion of Stage I, \(H > 0\) and \(H < 0\) throughout Stage I, i.e., time devoted to schooling increases (at a diminishing rate) as time and educational attainment increase. Since \(V'(1-H) = \mu_1 n, \mu_1\) and \(V'(1-H)\) are continuous functions, and \([-V''(1-H)] > 0\), \(H\) is a continuous function of time in Stage I. Also, since \(F = 1, Y = 0\) throughout Stage I, i.e., persons who are in school full-time have zero earnings.
It is also continuous as the individual (a) leaves Stage I and enters Stage II, and (b) leaves Stage II and enters Stage III. To see this, note that \( H \) will be continuous across Stages I and II and Stages II and III if

\[
\lim_{F \to 1} \mu_1^F + \mu_2^F + \mu_3^F g(F) = \mu_1^0 + \mu_2^0; \quad \text{and}
\]

\[
\lim_{F \to 0} \mu_1^F + \mu_2^F + \mu_3^F g(F) = \mu_2^0 + \mu_3^0,
\]

respectively. These follow immediately from the assumptions on \( g(F) \), i.e.,
that \( g(0) = 1, g(1) = 0 \) and \( F \) and \( g(F) \) are continuous. Moreover, \( H \) is continuous when the individual leaves Stage III and enters Stage IV (if indeed he does so). To see this, note that \( H \) will be continuous across the III - IV switchpoint if

\[
\lim_{t \to R} H(t) = 0, \quad \text{where} \ R \ \text{is the age of retirement,} \ H(R) = 0, \ \text{and} \ \mu_1 = \mu_2 = 0, \ \mu_3 = \frac{V'(1)}{y} \ \text{at the beginning of "retirement".}
\]

Since, at the beginning of retirement \( V'(1) = V'[1 - H(R)] = \mu_3(R) \), condition (iii) may also be written as

\[
\lim_{t \to R} \mu_3(R) = \frac{V'(1)}{y}, \quad \text{which must also hold since} \ \mu_3 \ \text{is a continuous function.}
\]

So much for the continuity of \( H \). As regards \( \dot{H} \), recall that

\[
[-V'(1-H)]H = \rho \left[ \mu_1^0 + \mu_2^0 \right] = \rho V'(1-H) \ \text{in Stage I;}
\]

\[
[-V'(1-H)]H = \rho V'(1-H) - (r + \delta) \mu_3^0 g(F) \ \text{in Stage II; and}
\]

\[
[-V'(1-H)]H = \rho V'(1-H) - (r + \delta) \mu_3^0 y \ \text{in Stage III.}
\]
Continuity within each stage follows from the fact that $F$, $Y'(1-H)$, 
$[\gamma'](1-H)$, $g(F)$, and $\mu_1$, $\mu_2$ and $\mu_3$ are continuous. Continuity across 
Stages I, II and III, is implied by the fact that

\[
\lim_{F \to 1} \frac{\rho Y'(1-H) - (r + \delta)\mu_3 Y g(F)}{\gamma} = \rho Y'(1-H) \quad \text{(for the switch from I to II)}
\]

\[
\lim_{F \to 0} \frac{\rho Y'(1-H) - (r + \delta)\mu_3 Y g(F)}{\gamma} = \rho Y'(1-H) - (r + \delta)\mu_3 Y \quad \text{(for the switch from II to III)}
\]

However, $H$ is discontinuous across the switch from Stage III to "retirement" 
(i.e., Stage IV) if such a switch occurs, for 
\[
\lim_{t \to R} \frac{\rho Y'(1)}{[\gamma'](1)} - \frac{(r + \delta)\mu_3(R)}{[\gamma'](1)} < 0.
\]

So much for the continuity of $H$ and $Y$. Now review the behavior of $H$ 
in Stages I - III in the "normal" life cycle. As noted earlier,

(i) In Stage I, $H > 0$ and $H < 0$ always.

(ii) In Stage II, $H$ may either rise monotonically, or else rise at first 
and then fall. In either case, if $\delta x > r + \delta$, $H$ will certainly be negative 
($H$ could easily be negative even if $\delta x < r + \delta$).

(iii) In Stage III, $H$ may either fall monotonically (if, e.g., $H$ was 
falling at the end of Stage II); rise and then fall (if, e.g., $H$ was rising 
at the end of Stage II); or cycle (e.g., fall, rise and fall again; rise, 
fall, rise and then fall again; etc.). If cycles do occur, they must always 
end with $H$ falling. Also, since $[-\gamma'](1-H) = \rho Y'(1-H) - (r + \delta)\mu_3 Y$ in 
Stage III,

$\gamma$, I stress that the behavior described below is a stylized portrait of 
"normal" or typical life-cycle behavior of men, but has rather little to do 
with that of women.
\[
\frac{[-V''(1-H)]}{V'(1-H)} \dot{H} + \frac{d}{dh} \left( \frac{[-V''(1-H)]}{V'(1-H)} \right) (h)^2 \\
+ (r + \delta) \left[ y' (1-H)[(kxH - (r + \delta)] + (r + \delta)x \mu_3 y \right] = 0
\]

So if one assumes "increasing risk aversion" then \( H \) is certainly negative whenever \( H > (r + \delta)/\beta x \), but could be positive for "sufficiently" small \( H \).

Figures 7(A) - 7(D) depict several possible "normal" life cycle patterns for \( H \), as implied by the above analysis. Figures 7(A) - 7(C) assume that at \( t = T \), \( \mu_1 = \mu_2 = 0 \) and \( \mu_3 = V'(1)/y \), and thus that the individual's life cycle ends in Stage III; Figure 7(D) assumes that at \( t = R < T \), \( \mu_1 = \mu_2 = 0 \) and \( \mu_3 = [-V''(1)]/y \), and thus that at \( t = R \) the individual enters "retirement" (Stage IV), which lasts from \( t = R \) to \( t = T \). As was apparent earlier, a variety of possible "normal" life cycle patterns for \( H \) can arise in this model. Since \( H > 0 \) throughout Stage I and \( H < 0 \) for \( t \) close to \( T \), all profiles must have at least one peak; but cycles are possible (at least, in Stage III), and the age-labor supply profile need not be concave everywhere.

2. Observed wage rates, \( W_0 \): As noted in the discussion of Stage I, when \( F = 1 \) the individual described here may be said to be engaged in full-time schooling; thus, \( W_0 \) is not observed in Stage I. In Stage II, since \( W_0 = y \lambda_g(F) \) and since \( F \) and \( g(F) \) are continuous, \( W_0 \) is continuous also, both within stages and across stages. At the very beginning of Stage II, \( W_0 = 0 \) since \( F = 1 \) and \( g(1) = 0 \). Thereafter, \( W_0 > 0 \), with

\[
\frac{\dot{W}_0}{W_0} = \alpha \mu F \dot{H} + \beta x H - \delta + \left( \frac{[g'(F)]/g(F)}{\dot{F}} \right) \]
in Stage II, and

\[ \frac{\dot{W}_0}{\dot{W}_0} = \beta x \delta - \delta \]

in Stage III.

At the beginning of Stage II, \( F = 1 \) and \( H > (r + \delta)/(\alpha N + \beta x) \), so, as noted earlier, \( W_0 \) is certainly rising at the beginning of Stage II. \( W_0 \) may go on rising throughout Stage II and some of Stage III -- indeed, it will always rise provided \( \delta \leq \frac{1}{2} \) and/or \( \frac{H}{x} > \frac{\beta x}{\beta x} \). (Of course, when \( H < \delta/\beta x \), with \( \delta > 0 \), then observed wages are falling.) (Since \( F \) may be discontinuous at the Stage II-III boundary, \( \dot{W}_0/W_0 \) may be discontinuous also.)

Figures 8(A) - 8(C) depict several possible life-cycle patterns for \( W_0 \). As with \( H \), a variety of patterns is possible for \( W_0 \); in particular, there is no reason to suppose that \( W_0 \) and \( H \) will be positively correlated over time (due to substitution effects or any other reason), as Geisz and Becker seem to suggest (see above, Part IIIB). The same applies to the relationship between \( W_p(w; k) \) and \( H \).2 Rather, the present model implies, the rate of growth of \( W_0 \) (and \( W_p \)) will be positively correlated with the level of \( H \) and, in Stage II, \( F \) (due to the fact that net investments in \( K \) are determined by the level of \( H \) and, in Stage II, \( F \)).

1/ In other words, if technical progress and/or "learning by living" effect offset \( \alpha_N + \beta x \), the wage elasticity-weighted sum of the decay rates of \( N \) and \( X \).

2/ Curiously, while Geisz and Becker argue in their Ch. 1 that \( W_p \), the potential wage, and \( H \) will be positively correlated over the life cycle, their Ch. 3 tests this proposition by regressing \( \log(H) \) on \( \log(W_0) \), the observed wage, even though \( W_0 = W_p \) only in Stage III -- regardless of whether \( g(F) \) is concave, as I have assumed, or linear and equal to \( 1-F \), as Geisz and Becker assume.
3. Earnings, \( Y \): In Stage I, \( Y = 0 \) since \( F = 1 \) and the individual is engaged in full-time schooling. In Stage II, \( Y > 0 \) and

\[
\frac{\dot{Y}}{Y} = \frac{\dot{W}_0}{W_0} + \frac{\dot{H}}{H} = \omega H + \delta x H - \delta + \left[ \frac{-g'(F)}{g(F)} \right] (\dot{F}) + \frac{\dot{H}}{H}.
\]

In Stage III, \( Y > 0 \) (except when \( H = 0 \)) and

\[
\frac{\dot{Y}}{Y} = -\frac{\dot{W}_0}{W_0} + \frac{\dot{H}}{H} = \frac{K}{K} + \frac{\dot{H}}{H} = \delta x H - \delta + \frac{\dot{H}}{H}.
\]

Since \( F \) may be discontinuous on the boundary between Stage II and III, \( \dot{W}_0/W_0 \) and hence \( \dot{Y}/Y \) may be also. If \( H > \delta/\delta x \), or if \( \delta > 0 \), \( \dot{Y}/Y \) will rise so long as \( \dot{H}/H > 0 \). As the beginning of Stage II, \( \dot{H}/H > 0 \) and \( H > \delta/\delta x \), so here \( Y \) is certainly rising; presumably \( Y \) will fall eventually later in life, e.g., in Stage III.

Figures 9(A) - 9(C) depict several possible patterns for the life cycle behavior of \( Y \).

So much for the model's predictions for \( H, W_0 \) and \( Y \) over the life cycle. Note that they are in broad agreement with observed patterns of \( H, W_0 \) and \( Y \) as revealed in cross-sectional data on males by age,\(^1\) and that the cycles in \( H \) which may occur in the model have indeed been observed empirically.\(^2\) One final remark on the model's predictions for \( H \) seem in order: the model equates "the end of life" with the \( \mu_3 \) axis, yet as my discussion of the life-cycle behavior of \( H \) indicates, the individual may "end life" at \( F \) with \( H > 0 \) (i.e., arrive at the \( \mu_3 \) axis with \( \mu_3 > V'(1)/y \)). This seems to mean that the individual goes on working literally until he dies; yet in most cases people retire well before death, and one observes a discontinuity in \( H \) as \( H \) falls immediately from some positive number (e.g., 2000 hours per year) to zero.

Does this mean that the model's treatment of retirement is inadequate?

\(^1\) For such data on \( W_0 \) and \( H \), see Ghez and Becker, pp.85-92, for such data on \( Y \), see Mincer (1974), esp. Ch. 5.

My answer is somewhat equivocal. On the one hand, one could argue that the availability of pensions, Social Security benefits, etc., late in life -- a factor which of course is not represented in this model -- accounts both for retirement before death and for the sudden discontinuous change in \( N \) from a fairly large positive number to zero which typically occurs late in life. Of course, such public and private old-age programs could (at least if they are actuarially fair with no load, which they usually are not) be regarded as a form of saving, i.e., as a form of nonhuman asset accumulation, and thus as having been captured adequately in the model; but in this case it is evident that the model's predictions certainly do not, in general, agree with observed events. Clearly, more attention should be paid to the effects of old-age programs on labor supply late in life; this is an important area for future research.

On the other hand, one might argue that "\( T \)" should be regarded as the length of the individual's working life and that the age of retirement is exogenously determined. In this case, one could add, the individual should be understood to have reached retirement age (not the end of life) when he reaches the \( \mu_3 \) axis, and to have "retired early" if he reaches the \( \mu_3 \) axis with \( \mu_3 = \nu'(1)/\gamma \) at some \( t_r < T \). (Of course, this begs the question of why the working life is \( T \), i.e., of why retirement age is fixed; indeed, one would think that the presence of persons who wish to work more than \( T \) years would give employers an incentive to allow for a variety of working lives. But perhaps the "market for working lifetimes" operates so slowly that, as a first approximation, the assumption of an exogenously-fixed working life will be useful.) In this case, the model treats retirement in a perfectly adequate way.
III. Econometric Issues

The principal testable propositions\(^{1/}\) contained in the model described in Part II are as follows: that

\[
\dot{H}(t) = \frac{p}{R(H(t))} - (r + \delta) \frac{U'(C(t))}{[-V''(1-H(t))]} \frac{W_0(t)}{p} \tag{II-1}
\]

\[
\dot{W}_0(t)/W_0(t) = -\delta + \alpha \circ H(t) + \beta x H(t) + E(F(t))[-F(t)/F(t)] \tag{II-2}
\]

in Stage II; and that

\[
\dot{H}(t) = \frac{p}{R(H(t))} - (r + \delta) \frac{U'(C(t))}{[-V''(1-H(t))]} \frac{W_0(t)}{p} \tag{III-1}
\]

\[
\dot{W}_0(t)/W_0(t) = -\delta + \beta x H(t) \tag{III-2}
\]

in Stage III, where

\[
R(H(t)) = \frac{[-V''(1-H(t))]'}{V'(1-H(t))}, \text{ the (negative of the) "degree of absolute risk aversion"}
\]

\[
E(F(t)) = F(t)[-g'(F(t))]/g(F(t)), \text{ the elasticity of the } g(F) \text{ function}
\]

\(^{1/}\)(II) and (III) are "testable" in that they relate mainly to variables such as \(H\) and \(W_0\), variables which not only are observable in principle but also have in fact been observed. (Of course, once the model's parameters have been estimated, the estimates can be used to "work backwards," i.e., to construct measures of unobservable variables such as \(W_0\). And once such measures have been constructed, it will be possible to draw inferences about whether the model's implications about these variables have been supported by the data.)
\(t_1\) is age at which the switch from Stage I to Stage II occurs
\(t_2\) is age at which the switch from Stage II to Stage III occurs
\(R\) is retirement age

(II-1) is derived from (13-1) on p. 39, the definition \(\mu_3 = \lambda_3 K\), the fact that \(W_0 = yKg(F)\) in Stage II and the first-order condition (8-1) that \(\lambda_3 = \text{U}'(C)/p\). (III-1) is derived from (14-3) on p. 53, the definition \(\mu_3 = \lambda_3 K\), the fact that \(W_0 = yK\) in Stage III and the first-order condition (8-1). Since \(W_0 = Y/H = yKg(F), K = x^\alpha, 0 < \alpha < 1\) in Stage II and \(F = 0\) in Stage III, equations (II-2) and (III-2) follow immediately from equations (7) and differentiation with respect to time of the equation \(W_0 = yKg(F)\).

In principle, the parameters governing these functional relationships can be estimated using econometric methods, and the estimates can then be evaluated in the light of the model's predictions. (For example, the model assumes that \(\alpha, \beta, \beta\) and \(x\) are all positive.) But a great many questions must be answered before estimation may proceed. I consider some of the questions, grouping them under three heads. Some questions refer to problems which would be encountered if one were using data on the behavior with which the model is specifically concerned, namely, a time series for an individual over the individual's entire life-cycle. A second group of questions refers to problems which would be encountered if one

\[\text{Of course, the list of problems discussed below hardly constitutes a complete enumeration. For example, I ignore several problems of specification, e.g., how to capture the effects of taxes and transfers or of exogenous changes (due, say, to changes in the macroeconomy) in variables such as } y, p, \text{ etc., during the life cycle. The discussion below, then, attempts to show how one might mitigate some, but by no means all, of the important problems which will arise in the course of attempts to investigate equations such as (II) and (III).}\]
attempted to apply the model to longitudinal data such as the National Longitudinal Survey covering (typically) only a portion of the life-cycle of different individuals. The third group of questions refers to problems which would be encountered if one attempted to apply the model to the kind of data which are most readily available, namely, cross section data such as the 1/1000 U.S. Census sample covering different individuals observed as of one point in time.

III.A. Estimation using life-cycle data on one individual

Complete life-cycle data on one individual are almost never available. But since the problems which one would encounter using these data will also be encountered using longitudinal or cross-section data, it is nevertheless useful to ask how equations (II) and (III) might be estimated using data on one individual’s life cycle.

First, it is clear that the functions in these equations -- $R(H)$, $U'(C)$, $[-V''(1-H)]$, and $E(F)$ -- must be specified; and that the behavior of the unobservable variable $F$ (or, equivalently, of $(-\dot{\bar{F}}/\bar{F})$) must also be specified. The assumptions introduced in Part II are of some assistance; they imply, for example, that

$$R'(H) > 0$$
$$U'(C) > 0; \lim_{C \to 0} U'(C) = \infty$$
$$F < 0; E(0) = 0, E(1) = \infty, E'(F) > 0$$
$$[-V''(1-H)] > 0, V'''(1-H) > 0$$

But these assumptions do not take things very far, for obviously a great range of alternative specifications can obey these restrictions. Thus,
estimation may -- indeed, should -- use a variety of alternative assumptions about each of these functions. (For example, Mincer (1974) uses a variety of assumptions about what is, in effect, my F; see pp. 85-88.)

A second problem relating to estimation of (II) and (III) from individual life-cycle data arises from the fact that $t_2$, the date at which the individual switches from Stage II (i.e., equations (II)) to Stage III (i.e., equations (III)), is not known or observable a priori. Many approaches to this problem are, in principle, possible. For example, one might use the method of "switching regressions" (see, for example, Richard Quandt), which in a very loose sense is a technique for choosing the livelihood-maximizing value of $t_2$. On the other hand, a simpler approach would be to adopt an iterative procedure, trying different values of $t_2$ in turn and adopting that value for which the estimates of the parameters $\delta$ and $\delta x$ in the two different equations (II-2) and (III-2) most nearly coincide. (In contrast, Rosen (1975) chooses $t_2$ more or less arbitrarily, in the context of his attempt to estimate the parameters of a pure training model which assumes income, rather than utility, maximization.)

A third problem which will arise in the course of attempts to estimate (II) and (III) from individual life-cycle data concerns the fact that, while the model in effect assumes that the individual it describes is quite without any family ties or relationships, most individuals are, of course, married. Since most individuals are married, should equations (II) and (III) be modified in any way? The answer depends critically on one's assumptions about the way in which the state of being married affects one's $
abla_{t_1}$ is observable; it is simply the last period for which $Y = 0$ (since $F = 1$ at $t_1$).
economic activity.

For example, the model might be modified as follows:

1. assume a household utility function, given utility at any moment of time as \( U(C) + \sum \psi_i(L_i) \), with \( i = h \) or \( w \) (for husband or wife). (This, one would postulate that family utility depends on total family consumption, \( C \), and on the leisure times of both its members, \( L_h \) and \( L_w \); and is separable in each of its arguments.)

2. assume pairs of expressions like those given by equations (7), giving (for example) the rate of change of \( N, X \), etc., for both the husband and the wife.

In this case, the Hamiltonian would be expanded to

\[
\mathcal{H} = e^{-\rho t} [(U(C(t)) + \sum \psi_i(L_i(t))] + \sum \lambda_{ii} \dot{N}_i(t) [n_i \dot{F}_i(t)h_i(t) - \delta_{N_i}]
\]

\[
+ \sum \lambda_{ii} \dot{X}_i(t) [x_i \dot{H}_i(t) - \delta_i] + \lambda_3 [rA(t) + \sum \beta_i y_i \dot{N}_i(t) x_i \dot{H}_i(t) - \rho c(t)]; i = h \text{ or } w
\]

It can be shown that in this case the analogue to equations (II) and (III) is a pair of equations (one for each family member) which are identical to equations (II) and (III). In other words, the assumption that family utility is separable in the leisure times of its members turns out to mean that

\[
H_i(t) = \frac{\rho}{R(H_i(t))} \left( r + \delta \right) \frac{U_i(C(t))}{[-\psi_i(1-H_i(t))]} \frac{W_{0i}(t)}{\rho} \quad (\text{II-11})
\]

\[
\frac{W_{0i}(t)}{W_{0i}(t)} = -\delta_i + \alpha_i \dot{F}_i(t)H_i(t) \beta_i \dot{X}_i \dot{H}_i(t)
\]

\[
+ \sum _i (\dot{F}_i(t)) [\dot{F}_i(t)/F_i(t)] \quad (\text{II-21})
\]
in Stage II, and

\[ H_1(t) = \frac{\rho}{R(H_1(t))} - \left( r + \delta_1 \right) \frac{U'(C(t))}{-\nu'(1-H_1(t))} \frac{W_{G1}(t)}{p} \]  

\[ t_{III} \leq t < \infty, \quad (III-11) \]

\[ W_{G1}(t)/W_{G1}(t) = -\delta_1 + \beta_1 x_1 H_1(t) \]

in Stage III, where

\[ R(H_1(t)) = \frac{-\nu''(1-H_1(t))}{\nu'(1-H_1(t))} \]

\[ E_{\alpha_1}(t) = F_1(t) \frac{-\gamma'_1(F_1(t))}{\gamma_1 F_1(t)} \]

\[ t_{II} \equiv \text{age at which i switches from i's Stage I to i's Stage II} \]

\[ t_{III} \equiv \text{age at which i switches from i's Stage II to i's Stage III} \]

\[ R \equiv \text{retirement age} \]

In this case, then, the system of equations describing hours of work and the observed wage for the husband and the system describing hours of work and the observed wage for the wife are independent of each other. (This should not be very surprising; after all, the utility function is also assumed to be separable in consumer goods and leisure time; and the system of equations governing the behavior of the former is independent of the system governing the latter.) Hence, under this assumption no adjustment or allowance for the fact of family membership need be made -- at least in the sense that equations (II) and (III) need not be modified.  

\[ 1/ \text{This is not to say that married people and unmarried people are identical, however; presumably, these two types of people have different tastes (i.e., } U(.). \text{ and } V(.). \text{ functions), and may also have had different values of } N(0), X(0) \text{ and } A(0). \text{ So estimation should allow the parameters of the } R(N), \text{ etc., functions to be different for these different groups.} \]
On the other hand, the model might instead be modified as follows:

(1) assume a household utility function, giving utility at any moment of time as \( U(C) + V(L_h, L_w) \). (Thus, one would postulate that family utility depends on \( C, L_h, \) and \( L_w \); that family utility is separable in \( C \), on the one hand, and \( L_h \) and \( L_w \), on the other; and that family utility is not separable in \( L_h \) and \( L_w \).)

(2) assume pairs of expressions like those given by equations (7), giving (for example) the rate of change in \( N, X, \) etc. for both the husband and the wife.

In this case, the Hamiltonian would be expanded to

\[
\mathcal{H} = e^{-\rho t} [U(C(t)) + V(L_h, L_w)] + \sum_i \lambda_{1i}(t) N_i(t) [n_i P_i(t) H_i(t) - \delta_{M_i}]
\]

\[+ \sum_i \lambda_{2i}(t) X_i(t) [x_i H_i(t) - \delta_{X1}]
\]

\[+ \lambda_3 [rA(t) + \sum_1 \alpha_1(t) a_{i1}(t) x_i(t) + \alpha_2(t) H_i(t) - pC(t)]; \quad i = h \text{ or } w\]

It can be shown that in this case the analogue to equations (II) and (III) is two pairs of equations, one pair for each family member, as follows:

\[
H_1(t) = \frac{\partial}{\partial t} \frac{H_1(t)}{V_1(t)} - (r + \delta_1) \frac{U'(C(t))}{V_1(t)} \frac{W_{01}(t)}{p}
\]

\[
\text{for } t_{ii} < t < t_{III}\]

\[
\dot{W}_{01}(t)/W_{01} = -\delta_1 + \alpha_1 n_{i1}(t) H_i(t) + \beta_1 x_i H_i(t)
\]

\[+ \delta_1 [F_1(t)/p] \]

\[\text{for } t_{ii} < t < t_{III}\]
in Stage II, and

\[
\dot{H}_1(t) = \frac{-\delta_1 \cdot \frac{W'(C(t))}{-V_{11}(t)} \cdot \frac{W_{01}(t)}{p}}{\left[-\frac{V_{11}(t)}{V_{11}(t) \cdot H_1(t)} \right]} \quad (III-1i')
\]

\[
\dot{W}_{01}(t)/W_{01} = -\delta_1 + \beta_1 x_{11} \cdot H_1(t) \quad (III-21')
\]

in Stage III, where

\[
V_i(t) = \delta \cdot \nabla \left[ (1 - H_1(t)) \cdot (1 - H_j(t)) \right]/\partial H_1(t)
\]

\[
V_{11}(t) = \delta^2 \cdot \nabla^2 \left[ (1 - H_1(t)) \cdot (1 - H_j(t)) \right]/\partial H_1(t)^2
\]

\[
V_{1j}(t) = \delta^2 \cdot \nabla^2 \left[ (1 - H_1(t)) \cdot (1 - H_j(t)) \right]/\partial H_j(t) \cdot \partial H_1(t)
\]

\[
E_i(F_i(t)) = F_i(t) \cdot \left[ -g_i'(F_i(t)) \cdot g_i'(F_i(t)) \right]/g_i(F_i(t))
\]

\[t_{IIi} \equiv \text{age at which } i \text{ switches from Stage II to Stage III of } i\]

\[t_{IIIi} \equiv \text{age at which } i \text{ switches from Stage II to Stage III of } i\]

\[i = h \text{ or } w; \quad j = w \text{ or } h; \quad R \equiv \text{retirement age}\]

Note that since in general \(V_i(t), V_{11}(t)\) and \(V_{1j}(t)\) depend on \(H_1(t)\) and \(H_j(t)\), and since \(H_1(t)\) depends on \(H_j(t)\), the absolute levels and rates of change of the labor supplies of husband and wife affect each other.

In other words, the assumption that family utility is not separable in the leisure times of its members turns out to mean that the time rate of
change of labor supply for any given member depends, in part, on the level and time rate of change of labor supply of the other member. So (II-1) and (III-11') are systems of simultaneous equations and must obviously be estimated as such.

A fourth problem which will arise in the course of attempts to estimate equations such as (II) and (III) is mathematical and econometric rather than theoretical in nature. It occurs because the dependent variable \( H(t) \) and \( W_0(t)/W_0(t) \) are instantaneous time-rates of change. Now, since the only kinds of observable changes are discrete, one must somehow translate the instantaneous rates of change \( \dot{H}(t) \) and \( \dot{W}_0(t)/W_0(t) \) into something measurable, i.e., discrete changes. Here there are several possibilities.

On the one hand, one might simply measure \( H(t) \) by the discrete absolute change \( H(t+1) - H(t) \), i.e., use the difference between \( H(t+1) \) and \( H(t) \) as the dependent variable; and, likewise, measure \( \dot{W}_0(t)/W_0(t) \) by the discrete percentage change \( [W_0(t+1) - W_0(t)]/W_0(t) \). However, note that this would put \( H(t) \) on both sides of equations such as (II-1) or (III-1); and this could bias estimates of such equations in two ways. First, errors of measurement in \( H(t) \) would be repeated on both sides of (II-1) or (III-1) leading to spurious correlation between the dependent and independent variables; and, second, since the stochastic term in such equations is in effect specified as affecting \( H(t) \), it will necessarily be correlated with the independent variables in (II-1) or (III-1) (for these are also functions

1/ Equations (II-21) and (III-21) seem to suggest that wage rates of husband and wife are independent of each other. Strictly speaking, this is certainly true; but they are nevertheless related in an indirect sense; for since \( H(t) \) is affected by \( H(t) \) and \( H(t) \), \( W_0(t+1) \) is also. More technically, (II-2X) and (III-21) are pairs of recursive systems.

2/ The discussion which follows also applies to estimation of equations (III1) and (III1'), and (III') and (III1').

3/ This is in effect how Heckman (1975) instantaneous time-rates of change. (My comments on bias are not directly relevant to his work, however, since he estimates equations quite different from (II) or (III).
of \( H(t) \). Similar problems will arise in the estimation of equations for \( \dot{W}_0(t) \) if, as is usually the case, \( W_0(t) \) is not directly observed but, rather, must be derived by dividing \( Y(t) \) by \( H(t) \).

Another approach to this problem would be to use \( H(t+1) - H(t) \) as the measure of \( H(t) \) and use \( [W_0(t+1) - W_0(t)]/W_0(t) \) as the measure of \( \dot{W}_0(t) \); move \( H(t) \) to the right-hand side of equations such as (II-1) and (III-1), and multiply by \( W_0(t) \) both sides of equations such as (II-2) and (III-2); and then (a) constrain the estimated coefficient on \( H(t) \) in (II-1) and (III-1) to be unity, and (b) constrain the estimated constant term in (II-2) and (III-2) to be zero. In this way, one might avoid many of the bias problems which would arise if one used the (discrete) changes in \( H \) and \( \dot{W}_0 \) as dependent variables.

However, the notion that an instantaneous time-rate of change can be measured by a discrete change is valid only if time-rates of change are always constant; this is certainly a very strong assumption and, indeed, seems quite untenable.\(^1\) This suggests another, more appealing (and also more difficult) approach to the measurement of \( H(t) \) and \( \dot{W}_0(t)/W_0(t) \), i.e., to develop expressions approximating these instantaneous rates of change, e.g., by Taylor's series approximations. A fairly simple way of approaching this task is as follows. In general, equations (II) and (III) take the form

\[
\dot{I}(t) = \sum_i a_{i1} \dot{Y}_i(t)
\]

\(^1\)For example, the analysis of Part II suggests that, under suitable assumptions about \( R(H(t)) \), \( H \) will be negative, at least in Stage III.
where \( I(t) \) and \( J_i(t) \) are variables and the \( a_i \) are constants. (For example, in (II-1), \( \dot{I}(t) = \dot{H}(t) \), \( a_1 = \rho, a_2 = (r + \delta) \), \( J_i(t) = [H(H(t))]^{-1} \) and \( J_i(t) = \frac{[U'(C(t))W_0(t)]}{1-H(t)} \).) Integration over time from \( t \) to \( t+1 \) yields

\[
I(t+1) - I(t) = \sum a_i \bar{J}_i(t)
\]

where \( \bar{J}_i(t) = \frac{\int_t^{t+1} J_i(t) \, dt}{t+1} \), i.e., \( \bar{J}_i(t) \) is the average value of \( J_i(t) \) in the interval \( (t, t+1) \). Thus, to replace equations such as (II) and (III), which involve instantaneous changes, with equations which involve discrete (and therefore measureable) changes, it will be necessary to develop explicit expressions for each \( J_i(t) \) so that the \( \bar{J}_i(t) \) corresponding to it may be obtained by integration over time. In some cases this will be quite easy; for example, one \( J_i(t) \) in (II-1) and (III-1) is simply \( H(t) \). In other cases, of course, this will be much more difficult; for example, \( [U'(C(t))W_0(t)]/[1-H(t)] \) is a complex function which is nonlinear in \( C(t) \) and \( H(t) \). In either case, however, (the approximation to) a \( \bar{J}_i(t) \) will usually depend in part on the values both at \( t \) and at \( t+1 \) of its arguments (e.g., \( C, H, \) etc.), which means that the potential for biases (due, e.g., to spurious correlation and to correlation between the stochastic term and the independent variables) may very well creep in.

This being the case, one would be hard put to say a priori whether the third approach to the treatment of the problem posed by the fact that the \( I(t) \) are infinitesimal rather than discrete is superior to the second.

The second approach entails biases because \( \dot{I}(t) \neq I(t+1) - I(t) \) whenever \( \ddot{I}(t) \neq 0 \); while, as just suggested, biases may creep into the third approach
because of the way in which the $\overline{J}_1(t)$ will have to be derived. In effect, the $I(t)$ will be measured inaccurately under the second approach, while the $\overline{J}_1(t)$ will be measured inaccurately under the third approach. Since both approaches seem to be subject to biases to an unknown (and probably unknowable) degree, and since the second approach is manifestly less complicated than the third approach, I shall in general use it rather than the third approach.

IIIB. Estimation using panel data

So much for some of the major problems associated with attempts to estimate equations such as (II) and (III) from data on an individual life-cycle. Of course, complete data on an individual life-cycle are simply not available, and in any case would probably not be of great interest, for one is usually much more interested in "average" behavior than in the behavior of any given individual. Thus, use of panel or longitudinal data seems much more attractive as a means of estimating the parameters which govern life-cycle behavior -- but attempts to estimate (II) and (III) from such data must confront not only the problems enumerated in Section IIIB, but others as well.

First, and in many ways most important, is the obvious but extremely significant fact that panel data are concerned with different individuals, who typically start life with different endowments of assets, training, and experience, possess different subjective rates of discount, face different interest rates, have different tastes (i.e., have different $U(.)$, $V(.)$ and $B(.)$ functions), etc. Presumably estimation must allow for such differences, but this raises what in a formal sense may be called questions
of comparative dynamics, i.e., questions relating to the effects on life-cycle paths of changes in parameters (or functional forms). Obviously, if differences of this kind across individuals are important -- and there certainly seems to be no strong a priori case for accepting the notion that they are not -- then estimates which ignore such differences will presumably be biased. While I shall presently discuss a rough-and-ready way of allowing for these effects, I shall postpone to a later date a thorough discussion of ways by which estimation might allow for such factors.

A second point is in some ways related to the first: panel data are concerned with different persons, and since different persons work in different occupations, industries and locations, non-pecuniary factors (e.g., the pleasantness or unpleasantness of work across industries, occupations and locations) are likely to vary across the sample. If all persons were identical (e.g., had the same initial endowments, subjective rates of discount, tastes, etc.) non-pecuniary factors could probably be ignored; but since all persons are obviously not identical, non-pecuniary factors will presumably generate differences in behavior (e.g., so-called "equalizing differentials" in observed wage rates). For example, the \( V(.) \) function should be regarded as a vector, \( V^A(1-H^A) + \ldots + V^N(1-H^N) \), where there are \( n \) different kinds of work, by occupation/industry/location, which the individual might do. Thus equations (II) and (III), especially (II), will presumably be different for individuals in different occupations, industries and locations.

While I shall postpone to a later date a thorough discussion of ways by which estimation might allow for the influences of comparative-dynamics and non-pecuniary effects which are likely to be present in panel data,
I shall nevertheless note one possible way to adjust for these effects — albeit in an extremely rough-and-ready fashion. This is simply to add to equations such as (II) and (III) as many predetermined variables as can be measured — both by themselves and in interaction with the variables which do appear in these equations (e.g., $X$, $W_0$, etc.) — such as sex, race and educational attainment. The intuitive rationale for this procedure as a way of adjusting for comparative dynamics effects is simply (a) that the variation in values of $y$, $N(0)$, tastes, etc., within groups for which such predetermined variables are the same is likely to be much smaller than the variation across groups with different values for these predetermined variables; and, much more important, (b) that (variations in) the predetermined variables are in large measure proxies — both as causes and as effects — for (variations in) the parameters and taste variable such as $y$, $N(0)$, $X(0)$, etc. In effect, then, this is a rationale for the almost universal practice² of stratifying samples by educational attainment for purposes of estimating life-cycle equations such as (II) and (III).

The suggestion made here is that the use of additional predetermined variables would seem, at least intuitively, to represent a still better — although still quite rough-and-ready — means of adjusting for comparative dynamics effects, taste differences and the like.

One final problem associated with the use of panel data as a means of estimating equations such as (II) and (III) is the following: panel data almost always cover appreciably less than a complete life-cycle. (For example,

²Followed, for example, by Ghez and Becker, Mincer (1974) and Rosen (1975).
the National Longitudinal Survey, one of the most ambitious attempts at collecting such data, covers at most only a ten-year span in the life of any given individual.) This presents particularly serious difficulties as regards estimation of equations such as (II-2) and (III-2): these two equations are relevant only for \( t < t_{II} \) and \( t > t_{II} \) respectively, yet \( t_{II} \) is not known a priori, and may occur at an age not included in the panel data. Hence equations such as these two may have to be estimated not from panel data on actual cohorts but from cross-section data on "synthetic" cohorts, i.e., on cross-section observations arranged by age, to a consideration of which I now turn.

IIIC. Estimation using cross-section data ("synthetic cohorts")

Attempts to estimate equations such as (II) and (III) from cross-section data ("synthetic cohorts") must confront virtually all of the problems noted above which would be encountered in the context of estimation using panel data (actual cohorts) or data on individual life-cycles.\(^1\) Essentially, the crudest version of the notion underlying use of cross-section data to estimate life-cycle behavior is that, when ranked in order of age, the observations in cross-section data constitute a set of life-cycles, a "synthetic cohort": that observations on a 50-year old in 1975 convey essentially what one would find if one were to able to observe the behavior in the year 2000 of someone who was 25 years old in 1975.

\(^1\)The obvious exception to this statement is that most cross-sections cover virtually the entire range of ages while, as just noted, panel data often do not. Thus, at least in this respect, cross-sections contain "more" information than panel data; but, as I shall argue in a moment, this information is quite different from that in panel data.
This is certainly a very ingenious idea, and is not without a degree of intuitive appeal as a means of using what is often available (cross-section data) as a substitute for what is almost never available (data on one or more complete life cycles). However, it is clear that synthetic cohorts are likely to be rather poor substitutes for actual cohorts for at least one reason: so-called "vintage effects." In other words, macroeconomic changes, economic growth, advances in technology, knowledge and the like, all have the effect of making calendar time as well as age an important determinant of behavior; each "vintage" in the "synthetic" cohort may differ. How, then, may such vintage effects be controlled for? Rosen (1975, esp. pp. 215-217) does so by introducing rate-of-growth parameters into his model, i.e., by assuming that the notion of vintage effects means, essentially, that certain variables such as $N(t)$ grow secularly and that such secular changes can be allowed for, or at least "smoothed," by the assumption that they occur at given rates of change. A more general (even if still very rough-and-ready) procedure might attack the problems on two fronts. First, one would assume that when equations such as (II) and (III) are estimated from synthetic cohorts, all parameters are functions of calendar time -- which, in such cohorts, simply means age. This would imply that age would appear as an independent variable in (II) and (III), both along and in interaction with other variables. Second, one could also use the values of certain macroeconomic variables (e.g., the real interest rate) as of time "0" for each "vintage." Indeed, it seems that at least some "vintage" effects are in principle very little different from the comparative dynamics effects to which I referred earlier;
and while I shall postpone to a later date a detailed discussion of ways by which either kind of effect might be controlled for, it seems that the same kind of approximations (e.g., use of predetermined or exogenous variables which could be used to allow for comparative dynamics effects might also provide a means of mitigating some of the worst problems caused by vintage effects.

The final problem associated with the use of synthetic cohorts which I shall discuss here is far and away the most important: the cross-sections from which the synthetic cohorts must be constructed contain data on individuals as of one point in time (say, t) and thus do not contain information on variables such as W_0(t + 1) or H(t + 1) for different individuals. Thus, data for individuals at each age will usually have to be averaged\(^1\) which shrinks the effective sample size enormously and thus makes it much more difficult to control for such factors as vintage effects. Of course, some cross-sections do contain some information for each individual on (certain) variables at t + 1 as well as information on (other) variables at t, so experimentation with these measures may to some extent be useful. (For example, the 1/1000 Census Sample contains data for each individual on "hours worked last week" and "weeks worked last year"; and perhaps the former (divided by 7 x 24 = 168) might be used by a measure of H(t+1) while the latter (divided by 52) might be used as a measure of H(t).) But it is clear that cross-sections, i.e., synthetic cohorts, can yield only a rather small amount of information for use in estimation, much smaller

\(^1\) See, e.g., Ghez and Becker and Heckman (1975; forthcoming). In otherwords, the (geometric) mean across all individuals aged t + 1 of some variable is treated as that variable's value at t + 1; etc.
(at least as regards effective sample size) then is available in panel data.

However, note that there are in principle no serious obstacles to estimating equations such as (II-1) and (III-1) from panel data, as these two equations are in fact identical for all $t < R$. The real problem associated with panel data is that $t_{II}$ may occur at an age not covered by these data, preventing accurate estimation of (II-2) and (III-2); in other words synthetic cohorts are indispensable only for estimation of (II-2) and (III-2). Perhaps, then, a satisfactory (though by no means ideal) strategy would be to proceed as follows:

1. estimate (II-1) and (III-1) directly from panel data

2. use estimates of (II-2) and (III-2) based on synthetic cohorts to get a rough estimate $\hat{t}_{II}$ of $t_{II}$; and then, if possible, use panel data to estimate (II-2) and (III-2) once again, trying alternative values of $t_{II}$ in the neighborhood of $\hat{t}_{II}$.

/\ i.e., if $\hat{t}_{II}$ as obtained from the synthetic cohort estimates is an age covered by the panel data.
Fig. 1: Stages of the Life Cycle in $\mu_1$, $\mu_2$, $\mu_3$ Space
Figure 1a: Stage IV of the Life Cycle in $\mu_1$, $\mu_2$, $\mu_3$ Space
Figure 2a: $H = 0$ Surface in Stage II

Figure 2b: $F = 0$ Surface in Stage II

Figure 2c: $\mu_3 = 0$ Surface in Stage II (Includes $H$, $F$ Plane)
Figure 3: Stage II in $H, F, \mu_3$ Space
Figure h: $y = 0$ Locus in Stage III

Figure h: $z = 0$ Locus in Stage III
Figure la: \( \theta = 0 \) Locus in Stage II

Figure lb: \( \theta = 0 \) Locus in Stage II
Figure 5: Stage III in $\mu_1$, $\mu_2$, $\mu_3$ Space
Figure 6: Stage III in $\mu_2, \mu_3$ Space (Assumes $\dot{\mu}_1 < 0$ Always)
Figure 7: Life-Cycle Patterns for H

Figure 7(A)

\[ \frac{(r + \delta)}{\beta x} \]

\[ \frac{(r + \delta)}{(\alpha_{m} + \beta x)} \]

Figure 7(B)

\[ \frac{(r + \delta)}{\beta x} \]

\[ \frac{(r + \delta)}{(\alpha_{m} + \beta x)} \]

Figure 7(C)

\[ \frac{(r + \delta)}{\beta x} \]

\[ \frac{(r + \delta)}{(\alpha_{m} + \beta x)} \]

Figure 7(D)
Figure 8: Life-Cycle Patterns for $W_0$

Figure 8(A)

$W_0$

$0 \quad \quad I \quad II \quad III \quad t$

$(\delta \leq 0 \text{ and/or } H(t) > \delta/2x \text{ for all } t)$

Figure 8(B)

$W_0$

$0 \quad \quad I \quad II \quad III \quad t$

$(\delta > 0 \text{ and } H(t) < \delta/2x \text{ for some } t)$

Figure 8(C)

$W_0$

$0 \quad \quad I \quad II \quad III \quad t$

$(\delta > 0, H(t) < \delta/2x \text{ for some } t, \text{ and cycles in } H)$
Figure 9: Life-Cycle Patterns for Y

(assumes cycles in $R$)

Figure 9(A)

(assumes "retirement" prior to $T$)

Figure 9(C)
REFERENCES

Gary S. Becker (1974),

Yoram Ben-Porath (1970),

Alan S. Blinder and Yoram Weiss (forthcoming),

William G. Bowen and T. Aldrich Finegan (1969),

Gilbert Shez and Gary S. Becker (1975),

Malcolm R. Fisher (1971),

James J. Heckman (1975),

James J. Heckman (forthcoming),

Jacob Mincer (1970),

Jacob Mincer (1974),

Jacob Mincer and Solomon Polachek (1974),

Richard Quandt (1971),
Sherwin Rosen (1972),

Sherwin Rosen (1975),

Paula Stephan (1975),

Yoram Weiss (1972),