NBER Working Paper Series

EDUCATION AND SELF SELECTION

Robert J. Willis
Stanford University and NBER

and

Sherwin Rosen*
University of Chicago and NBER

Working Paper No. 249

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR
AND SOCIAL INSTITUTIONS

National Bureau of Economic Research, Inc.
204 Junipero Serra Boulevard, Stanford, CA 94305

June 1978

*This research was supported by the National Bureau of Economic Research. Thanks are due to Sean Becketti for excellent research assistance and to Lung Fei Lee for advice on statistical issues. The order of the authors' names was selected by a random device.

A draft of this paper was presented at the UK/US Conference on Human Capital and Income Distribution, Kings College, Cambridge, England, March 19-23, 1978.

NBER Working Paper 249
June 1978

EDUCATION AND SELF SELECTION

Robert J. Willis and Sherwin Rosen

Abstract

A structural model of the demand for college attendance is derived as a selection problem in the theory of comparative advantage, in which individuals are endowed with different kinds of talents. Some talents and abilities are more valuable in the types of work associated with college education and others are more valuable for work associated with high school education. For example, mechanical abilities are less important for lawyers than for plumbers, whereas verbal abilities are much more valuable for lawyers. The market tends to sort people into work activities for which they have a comparative advantage, as indexed by expected earnings in each activity. The structural model also allows for the influence of parental background in the selection process.

Estimates are based on NBER-Thorndike data and support the theory. There is marked heterogeneity in the population and expected financial gains from college attendance are distributed with substantial variance. Nevertheless, those with greater expected gains have a much larger probability of attending college. The elasticity of demand for college attendance with respect to the permanent college-high school earnings differential is 2.0. Parental background factors also influence demand. The data support the comparative advantage theory: Those who did not attend college would have earned less as college graduates than those who actually chose to attend. More surprisingly, those who attended college would have earned less as high school graduates than did those who actually chose high school. There is no "ability bias" in these data.

Address and phone number for both authors, until September 1, 1978:

National Bureau of Economic Research
204 Junipero Serra Boulevard
Stanford, CA 94305
415/326-7160

9/1/78:

Robert J. Willis
Department of Economics
State University of New York
at Stony Brook
Stony Brook, New York 11794

Sherwin Rosen
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
312/753-4503

Revised
May 1, 1978

EDUCATION AND SELF SELECTION

Robert J. Willis and Sherwin Rosen*

I. INTRODUCTION

In this paper we specify and estimate a model of the demand for college education derived from its effect on expected lifetime earnings compared with its cost. Attention is focused on specifying the role of earnings expectations in the derived demand for schooling, which are found to be important determinants of the decision to attend college. In addition to including financial incentives, the model allows for a host of selectivity or sorting effects in the data that are related to "ability bias," family effects and tastes that have occupied other researchers. Background and motivation are presented in Section II. The structure of the model is a variant of a simultaneous equations problem involving discrete choices and is derived in Section III. The estimates, based on data from the NBER-Thorndike sample, appear in Section IV. Some implications and conclusions are found in Section V.

II. NATURE OF THE PROBLEM

Estimates of rates of return to education have been controversial because they are based on ex post realizations and need not reflect structural parameters necessary for correct predictions. For example, it is well understood that college and high school graduates may have different abilities

so that income foregone during college by the former is not necessarily equal to observed earnings of the latter. Our objective here is two-fold: One is to estimate life earnings conditioned on actual school choices that are purged of selection bias. The other is to determine the extent to which alternative earnings prospects as distinct from family background and financial constraints influence the decision to attend college.

One need go no further than straightforward comparisons of earnings outcomes among school classes for structural rate of return estimates if educational wage differentials were everywhere equalizing on the direct, opportunity and interest costs of schooling. For then the supplies of graduates (or "demands" for each level of education) would be nearly elastic at the equalizing wage differentials and the distribution of human wealth would be approximately independent of the distribution of schooling.[1] However, recent evidence on the structure of life earnings based on panel data strongly rejects this as a serious possibility. Total variance of earnings among people of the same sex, race, education and market experience is very large, and more than two-thirds of it is attributable to unobserved components or person-specific effects that probably persist over much of the life cycle.[2] The panel evidence therefore suggests that supply elasticities are substantially less than completely elastic at unique wage differentials, and that there are inframarginal "ability rents." Put in another way, observed rates of return are not wholly supply determined and depend on interactions with relative demands for graduates as well.

A natural approach has been to incorporate measures of ability into the statistical analysis, either directly or as indicators of unobserved factors, in order to, in effect, impute ability rent. But merely partitioning observed

earnings into schooling and ability components does not use any of the restrictions imposed on the data by a school stopping rule and that decision embodies all the economic content of the problem. Some of that additional structure is incorporated here.

Economic theories of education, be they of the human capital or signaling varieties, are based on the principle of maximum capital value: Schooling is pursued to the point where its marginal (private) internal rate of return equals the rate of interest. It is easy to show that this leads to a recursive econometric model in which (i) schooling is related to a person's ability and family background; and (ii) earnings are related to "prior" school decisions and ability. Earnings gains attributable to education do not appear explicitly in the schooling equation. Instead, the cost benefit basis of the decision is embedded in cross equation restrictions on the overall model, because the earnings equation is a constraint for the maximum problem that determines education attainment.[3] There are many estimates of recursive models in the literature, but very few have tested the economic (wealth maximizing) hypothesis.[4]

We begin with the assumption of marked heterogeneity and diversity in the population implicit in the unobserved component approach to panel data mentioned above. Costs and benefits of alternative school completion levels are assumed to be randomly distributed among people according to their capacities to finance education, tastes, perceptions, expectations, and an array of talents that affect performance in work activities associated with differing levels of schooling. Some of these things are observed while others are unobserved. Individuals are sorted into educational classes according to the interaction of a selection criteria (such as maximum present value) and the

underlying joint distribution of tastes, talents, expectations and parental wealth. The selection rule determines a partition of the underlying joint density, which corresponds to the realized educational distribution. The supply function of graduates at any level of schooling is "swept out" of the joint taste, talent, parental wealth distribution as increased wage differentials enlarge the subset of the partition relevant for that class.

Let Y_{ij} represent the potential lifetime earnings of person $\underline{i}$ if schooling level $\underline{j}$ is chosen, X_i a vector of observed talent or ability indicators of person $\underline{i}$ and τ_i an unobserved talent component relevant for person $\underline{i}$. Similarly, split family background and taste effects into an observed vector Z_i and an unobserved component ω_i. Let V_{ij} denote the value of choosing school level $\underline{j}$ for person $\underline{i}$. Then a general school selection model is:

$$Y_{ij} = y_j(X_i, \tau_i) \qquad j = 1,\ldots,n \tag{2.1}$$

$$V_{ij} = g(y_j, Z_i, \omega_i) \tag{2.2}$$

$$i \text{ belongs to } j \text{ if } V_{ij} = \max \{V_{i1},\ldots,V_{in}\} \tag{2.3}$$

$$(\tau,\omega) \sim F(\tau,\omega) \tag{2.4}$$

Equation (2.1) shows how potential earnings in any given classification vary with talent and ability.[5] The earnings function differs among school classes because work activities associated with alternative levels of education make use of different combinations of talent. Equation (2.2) translates the earnings stream from choice $\underline{j}$ into a scaler such as present value and is conditioned on family background to reflect tastes and financial barriers to extending schooling. Equation (2.3) is the selection rule: The person chooses the classification that maximizes value, and is observed in one and only one of the $\underline{n}$ possibilities open to him. Equation (2.4) closes

the model with a specification of the distribution of unobservables. Since observed assignments of individuals to schooling classes are selected on (X,Z,τ,ω), earnings observed in each class may be nonrandom samples of population potential earnings because those with larger net benefits in the class have a higher probability of being observed in it.

This formulation is suggested by the theory of comparative advantage.[6] It allows for a rather eclectic view of the role of talent in determining observed outcomes, since the X's may affect earning capacity differently at different levels of schooling [See equation (2.1)] and covariances among the unobservables are unrestricted. Indeed, there may be negative covariance among talent components. For example plumbers (high school graduates) may have very limited potential as highly schooled lawyers, but by the same token lawyers may have much lower potential as plumbers than those who actually end up choosing that kind of work. This contrasts markedly with the one factor ability-as-IQ specifications in the literature which assume that the best lawyers would also be the best plumbers and would imply strictly hierarchical sorting in the absence of financial constraints. In effect an IQ-ability model constrains the unobserved ability components to have large positive covariances--an assumption that is probably erroneous and not necessary for our methods. Note also that population mean "rates of return" among alternative schooling levels have no significance as guides to the social or private profitability of investments in schooling. For example, a random member of the population might achieve a negative return from an engineering degree; yet those with appropriate talents who choose engineering will obtain a return on the time and money costs of their training which is at least equal to the rate of interest.

There are difficult estimation problems associated with selectivity models. In brief, the unobservables impose distinct limits on the amount of structural information that can be inferred from realized assignments in the data. For example, it would be very desirable to know the marginal distribution of talents in (2.4), since it would then be possible to construct the socially efficient assignment of individuals to school classes, defined as the one that maximizes overall human wealth. Then the dead weight losses due to capital market imperfections could be computed by comparing optimal with observed assignments. However, the marginal density is not itself identified, since both unobserved financial constraints and talent jointly determine observed outcomes. These issues will be made precise shortly, but roughly speaking we do not necessarily know if a person chose college education because he had talent for it or because he was wealthy. What can and will be done is to map out the joint effects of the unobservables embedded in the actual demand curve for college attendance, which embodies all constraints inherent in the actual market, but which nevertheless is a valid structural basis for prediction. Selectivity or "ability bias" in unadjusted rate of return computations that do not take account of the sorting by talent inherent in observed assignments can also be computed.

A few limitations to these methods must be noted at the outset. It is crucial to the spirit of the model, based as it is on human diversity, that few covariance restrictions be placed on the distribution of unobservables. This practically mandates the assumption of joint normality, since no other non-independent multivariate distribution offers anything close to similar computational advantages. While the general selection rule specified below is likely to emerge from a broad class of economic models of school choice,

it is not known how sensitive the results are to the normality assumptions. In addition, nonindependence forces some aggregation in the number of choices considered for computational feasibility, even though the statistical theory can be worked out for any finite number.[7] This rules out of consideration other selection aspects of the problem that should be considered, such as choice of school quality.[8] All people in our sample have at least a high school education and we have chosen a dichotomous split between choice of high school and more than high school (college attendance). Some internal diagnostic tests help check on the validity of this aggregation. Experiments with a college completion or more compared with high school graduation or some college classification yielded results very similar to those reported below.

III. THE MODEL

Specification of the econometric model is tailored to the data at our disposal. More detail will be given below, but at this point the important feature is that earnings are observed at two points in the life cycle for each person, one point soon after entrance into the labor market and another point some twenty years later. The earnings stream is parameterized into a simple geometric growth process to motivate the decision rule. This is a reasonable approximation to actual life earnings patterns for the period spanned by the data. Two levels of schooling are considered, labeled A (for more than high school) and B (for high school).

If person i chooses A, the expected earnings stream is

$$y_{ai}(t) = 0 \qquad 0 < t \leq S$$

$$y_{ai}(t) = \bar{y}_{ai}\exp(g_{ai}t) \qquad S \leq t < \infty \qquad (3.1)$$

where S is the incremental schooling period associated with A over B and $\underline{t-S}$ is market experience. If alternative B is chosen, the expected earnings stream is

$$y_{bi}(t) = \bar{y}_{bi}\exp(g_{bi}t) \qquad 0 \leq t < \infty \qquad (3.2)$$

Thus earnings prospects of each person in the sample are characterized by four parameters: initial earnings and rates of growth in each of the two alternatives. Diversity is represented by a random distribution of the vector $(\bar{y}_a, g_a, \bar{y}_b, g_b)$ among the population.[9]

Equations (3.1) and (3.2) yield convenient expressions for present values. Assume an infinite horizon, a constant rate of discount for each person, r_i, with $r_i > g_{ai}, g_{bi}$, and ignore direct costs of school. Then the present value of earnings is

$$V_{ai} = \int_S^\infty y_{ai}(t)\exp(-r_i t)dt = \bar{y}_{ai}/(r_i-g_{ai}) \exp(g_{ai}-r_i)S \qquad (3.3)$$

if A is chosen and is

$$V_{bi} = \int_0^\infty y_{bi}(t)\exp(-r_i t)dt = \bar{y}_{bi}/(r_i-g_{bi}) \qquad (3.4)$$

if B is chosen. These are likely to be good approximations since the consequences of ignoring finite life discount corrections and nonlinearities in earnings paths toward the end of the life cycle are lightly weighted for non-negligible values of r.

Selection Rule.

Assume that person i chooses A if $V_{ai} > V_{bi}$ and chooses B if $V_{ai} \leq V_{bi}$. Define $I_i = \ln(V_{ai}/V_{bi})$. Substitution from (3.1) - (3.4) yields

$$I_i = \ln \bar{y}_{ai} - \ln \bar{y}_{bi} + (g_{ai} - r_i)S - \ln(r_i - g_{ai}) + \ln(r_i - g_{bi})$$

A Taylor series approximation to the nonlinear terms around their population mean values $(\bar{g}_a, \bar{g}_b, \bar{r})$ yields

$$I_i = \alpha_o + \alpha_1 (\ln\bar{y}_{ai} - \ln\bar{y}_{bi}) + \alpha_2 g_{ai} + \alpha_3 g_{bi} + \alpha_4 r_i \tag{3.5}$$

with

$$\begin{aligned} \alpha_1 &= 1 \\ \alpha_2 &= \partial I/\partial g_a = 1/(\bar{r} - \bar{g}_a) + S > 0 \\ \alpha_3 &= \partial I/\partial g_b = -1/(\bar{r} - \bar{g}_b) < 0 \\ \alpha_4 &= \partial I/\partial r = -\boxed{S + (\bar{g}_a - \bar{g}_b)/(\bar{r} - \bar{g}_a)(\bar{r} - \bar{g}_b)} \end{aligned} \tag{3.6}$$

Hence the selection criteria is

$$\begin{aligned} \Pr(\text{Choose A}) &= \Pr(V_a > V_b) = \Pr(I > 0) \\ \Pr(\text{Choose B}) &= \Pr(V_a \leq V_b) = \Pr(I \leq 0) \end{aligned} \tag{3.7}$$

Earnings and Discount Functions.

Let X_i represent a set of measured characteristics that influence a person's lifetime earnings potential and let $u_{1i},\ldots,u_{4i}$ denote permanent person-specific unobserved components reflecting unmeasured factors influencing earnings potential.[10] Specify structural (in the sense of population) earnings equations of the form

$$\ln \bar{y}_{ai} = X_i\beta_a + u_{1i}$$

$$g_{ai} = X_i\gamma_a + u_{2i} \qquad \text{if A is chosen} \qquad (3.8)$$

and

$$\ln \bar{y}_{bi} = X_i\beta_b + u_{3i}$$

$$g_{bi} = X_i\gamma_b + u_{4i} \qquad \text{if B is chosen} \qquad (3.9)$$

The variables on the left hand sides of (3.8) and (3.9) are to be interpreted as the individual's expectation of initial earnings and growth rates at the time the choice is made. In order to obtain consistent estimates of $(\beta_a, \gamma_a, \beta_b, \gamma_b)$ from data on realizations it is assumed that expectations were unbiased. Hence forecast errors are assumed to be independently normally distributed with zero means.

Let Z_i denote another vector of observed variables that influence the schooling decision through their effect on the discount rate.

Then

$$r_i = Z_i\delta + u_{5i} \qquad (3.10)$$

where u_5 is a permanent unobserved component influencing financial barriers to school choice. The vector (u_j) is assumed to be jointly normal with zero means and variance-covarience matrix $\Sigma = \{\sigma_{ij}\}$. Σ is unrestricted.

Reduced Form.

The structural model is (3.5), (3.8), (3.9), and (3.10). A reduced form of the selection rule is obtained by substituting (3.8) - (3.10) into (3.5):

$$I = \alpha_o + X[\alpha_1(\beta_a - \beta_b) + \alpha_2\gamma_a + \alpha_3\gamma_b] + \alpha_4 Z\delta + \alpha_1(u_1 - u_3) + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_5 u_5$$

$$\equiv W\pi - \varepsilon \tag{3.11}$$

with $W = [X,Z]$ and $-\varepsilon = \alpha_1(u_1 - u_3) + \alpha_2 u_2 + \alpha_3 u_4 + \alpha_5 u_5$. Thus an observationally equivalent statement to (3.5) and (3.7) is

$$\Pr(A \text{ is observed}) = \Pr(W\pi > \varepsilon) = F\left(\frac{W\pi}{\sigma_\varepsilon}\right) \tag{3.12}$$

where $F(\cdot)$ is the standard normal c.d.f. (3.12) is a probit function determining sample selection into categories A or B, to be estimated from observed data.[11]

Selection Bias and Earnings Functions

The decision rule selects people into observed classes according to largest expected present value. Hence the earnings actually observed in each group are not random samples of the population, but are truncated nonrandom samples instead. The resulting bias in observed means may be calculated as follows. Note that $\Pr(\text{observing } y_a(t)) = \Pr(I > 0) = \Pr(W\pi > \varepsilon)$. Therefore, from (3.8)

$$E(\ln\bar{y}_a \mid I > 0) = X\beta_a + E(u_1 \mid W\pi > \varepsilon)$$

Define $\rho_1 = \rho(u_1/\sigma_1, \varepsilon/\sigma_\varepsilon) = \sigma_{1\varepsilon}/\sigma_1\sigma_\varepsilon$. Then

$$E(\ln\bar{y}_a \mid I > 0) = X\beta_a + \sigma_1\rho_1 E\left(\varepsilon/\sigma_\varepsilon \mid \frac{\varepsilon}{\sigma_\varepsilon} < \frac{W\pi}{\sigma_\varepsilon}\right)$$

$$= X\beta_a + \sigma_1\rho_1[-f(W\pi/\sigma_\varepsilon)/F(W\pi/\sigma_\varepsilon)]$$

where F is the cumulative normal density and f is its p.d.f. Define

$$\lambda_a \equiv -f(W\pi/\sigma_\varepsilon)/F(W\pi/\sigma_\varepsilon) \tag{3.13}$$

as the truncated mean (with truncation point $W\pi/\sigma_\varepsilon$) of the normal density due to selection. Making use of the definition of ρ_1 and λ_a yields

$$E(\ln\bar{y}_a \mid I > 0) = X\beta_a + \frac{\sigma_{1\varepsilon}}{\sigma_\varepsilon}\lambda_a \tag{3.14}$$

A parallel argument for g_a, $\bar{y}_b$ and g_b yields

$$E(g_a \mid I > 0) = X\gamma_a + \frac{\sigma_{2\varepsilon}}{\sigma_\varepsilon}\lambda_a \tag{3.15}$$

$$E(\ln\bar{y}_b \mid I \leq 0) = X\beta_b + \frac{\sigma_{3\varepsilon}}{\sigma_\varepsilon}\lambda_b \tag{3.16}$$

$$E(g_b \mid I \leq 0) = X\gamma_b + \frac{\sigma_{4\varepsilon}}{\sigma_\varepsilon}\lambda_b \tag{3.17}$$

with

$$\lambda_b = E(\varepsilon/\sigma_\varepsilon \mid \frac{\varepsilon}{\sigma_\varepsilon} > \frac{W\pi}{\sigma_\varepsilon}) = f(W\pi/\sigma_\varepsilon)/[1 - F(W\pi/\sigma_\varepsilon)] \tag{3.18}$$

and

$$\sigma_{k\varepsilon} = -[\alpha_1(\sigma_{1k} - \sigma_{3k}) + \alpha_2\sigma_{2k} + \alpha_3\sigma_{4k} + \alpha_5\sigma_{5k}],\ k = 1,\ldots,4 \tag{3.19}$$

Note from (3.13) that $\lambda_a \leq 0$. Therefore the observed (conditional) means of initial earnings and rates of growth among persons in A is greater or less than their population means as $\sigma_{1\varepsilon}$ (and $\sigma_{1\varepsilon}$) $\lessgtr 0$, from (3.14) and (3.15). Conversely, $\lambda_b \geq 0$ [See (3.18)], and there is positive or negative selection bias in initial earnings and growth rates for people observed in class B according to $\sigma_{3\varepsilon}$ (and $\sigma_{4\varepsilon}$) $\gtrless 0$. Since σ_{ij} is unrestricted, $\sigma_{k\varepsilon}$ is also unrestricted and selection bias can go in either way. In particular, it is possible that the bias is positive in both groups, consistent with the comparative advantage argument sketched above. Positive bias in A and negative

bias in B would be consistent with a single factor (hierarchical) interpretation of ability. Of course neither finding yields a definitive "ability" interpretation because of the presence of expectational errors and financial factors (σ_{5k}) in (3.19): The assignments are based on talent, expectations and wealth, not on talent alone.

Estimation

Consider the following regressions applied to observed data:

$$\begin{aligned} \ln\bar{y}_a &= X\beta_a + \beta_a^*\lambda_a + \eta_1 \\ g_a &= X\gamma_a + \gamma_a^*\lambda_a + \eta_2 \\ \ln\bar{y}_b &= X\beta_b + \beta_b^*\lambda_b + \eta_3 \\ g_b &= X\gamma_b + \gamma_b^*\lambda_b + \eta_4 \end{aligned} \tag{3.20}$$

Equations (3.14) - (3.17) suggest that β_a^* estimates $\sigma_{1\varepsilon}/\sigma_\varepsilon$, γ_a^* estimates $\sigma_{2\varepsilon}/\sigma_\varepsilon$ and so on. Including λ_a or λ_b in the regressions along with X corrects for truncation and selectivity bias, and $E(\eta_{ij}) = 0$ for $j = 1,4$. In addition $E(\eta_{ij}^2)$ is heteroskedastic (See below) because the observations are truncated and at different points for different people. (3.20) cannot be implemented directly because λ_a and λ_b are not known. However, it can be shown[12] that consistent estimates of (3.20) are obtained by replacing λ_a and λ_b with their values predicted from the reduced form probit equation (3.12). These values are

$$\begin{aligned} \hat{\lambda}_{ai} &= -f(W_i\hat{\pi}/\sigma_\varepsilon)/F(W_i\hat{\pi}/\sigma_\varepsilon) \\ \hat{\lambda}_{bi} &= f(W_i\hat{\pi}/\sigma_\varepsilon)/[1 - F(W_i\hat{\pi}/\sigma_\varepsilon)] \end{aligned} \tag{3.21}$$

and are entered as least squares regressors along with X_i. Estimation of (3.20) with λ_i replaced by $\hat{\lambda}_i$ corrects for selectivity bias in the observations. What is more interesting for the economic theory of educational choice is that these estimates provide a basis for estimating the structural selection rule or structural probit function (3.5) and (3.7). The structural probit is

$$\Pr(\text{Choose A}) = \Pr[(\alpha_o + \alpha_1(\ln\bar{y}_a - \ln\bar{y}_b) + \alpha_2 g_a + \alpha_3 g_b + \alpha_4 Z\delta)/\sigma_\varepsilon > \varepsilon/\sigma_\varepsilon] \quad (3.22)$$

from (3.5), (3.7), and (3.10). Use the consistent estimates of structural earnings and growth described above to predict earnings gains for each person in the sample according to

$$\begin{aligned} \widehat{\ln(\bar{y}_{ai}/\bar{y}_{bi})} &= X_i(\hat{\beta}_a - \hat{\beta}_b) \\ \hat{g}_{ai} &= X_i\hat{\gamma}_a \\ \hat{g}_{bi} &= X_i\hat{\gamma}_b \end{aligned} \quad (3.23)$$

where $\hat{\beta}$ and $\hat{\gamma}$ are estimated by the method above.[13] These predicted values are inserted into (3.22) and estimated by the usual probit method to test the economic restrictions (3.6).[14]

Other Tests

Alternative estimates are available to serve as an internal consistency check on the model. In particular, the model can be specified using the observed level of earnings at time $\bar{t}$ and earnings growth instead of initial earnings. From (3.1) and (3.2) it follows that

$$\ln y_a(\bar{t}) = X_i(\beta_a + \gamma_a \bar{t}) + u_1 + \bar{t}u_2$$

$$\ln y_b(\bar{t}) = X_i(\beta_b + \gamma_b \bar{t}) + u_3 + \bar{t}u_4 \qquad (3.24)$$

Substitute for the level equations in (3.8) and (3.9) and this model also can be estimated as described above. However, now the structural probit is of the form

$$\Pr(A \text{ is chosen}) = \Pr[(\theta_o + \theta_1(\ln y_a(t) - \ln y_b(t)) + \theta_2 g_a + \theta_3 g_b)/\sigma_\varepsilon > \varepsilon/\sigma_\varepsilon] \qquad (3.25)$$

Since $\ln y_a(\bar{t}) - \ln y_b(\bar{t}) = \ln\bar{y}_a - \ln\bar{y}_b + (g_a - g_b)\bar{t}$

the following restrictions are implied:

$$\theta_1 = \alpha_1$$

$$\bar{t}\theta_1 + \theta_2 = \alpha_2$$

$$-\bar{t}\theta_1 + \theta_3 = \alpha_3$$

Hence we have a check on the validity of the model. Of course its main validation is the power to predict behavior and assignments on independent data.

Identification

Two natural questions regarding identification arise in this model.

(i) Estimation of the selection rule or structural probit equation is possible only if the vectors X and Z have elements that are not in common. If X and Z are identical, the predicted values of $\ln\bar{y}_a - \ln\bar{y}_b$, g_a, and g_b are co-linear with the other explanatory variables in (3.22) and its estimation is precluded. Note however that even if X and Z are identical, the

reduced form probit (3.12) is estimable and it still may be possible to estimate initial earnings and growth rate equations and selection bias. The reason is that although the $\hat{\lambda}$ corrections in (3.20) are functions of the same variables that enter the $X\beta$ or $X\gamma$ parts of these equations, they are nonlinear functions of the measured variables. Structural earnings equations might be identified off the nonlinearity, though in any particular application there may be insufficient nonlinearity if the range of variation in $W\pi$ (See (3.11)) is not large enough.[15]

In the general discussion of Section II, X was tentatively associated with measured abilities and Z with measured financial constraints (and tastes), corresponding to the Beckerian distinction between factors that shift the marginal rate of return to investment schedule and those that shift the marginal supply of funds schedule. Evidently, if one takes a sufficiently broad view of human investment and in particular of the role of child care in the new home economics, easy distinctions between the content of X and of Z become increasingly difficult, if not impossible to make. If X and Z are indistinguishable, the economic theory of school choice has no empirical content. In the empirical work below a very strong dichotomy with no commonalities is maintained: X is specified as a vector of ability indicators and Z as a vector of family background variables. This hypothesis is maintained for two reasons: First it provides a test of the theory in its strongest form. Certainly if the theory is rejected in this form there is little hope for it. Second, there have been no systematic attempts to find empirical counterparts for the things that shift marginal rate of return and marginal cost of fund schedules that cause different people to choose different amounts of schooling. The validity

of the theory rests on the possibility of actually being able to find an operational set of indicators, and this distinction is the most straightforward possibility.

(ii) Given resolution of problem (i), not all parameters in the model can be estimated. Some are overidentified and some are underidentified. The selectivity bias corrected structural earnings equations (3.20) directly estimate β_a, β_b, γ_a, γ_b, and the structural probit (3.22) provides estimates of $(\alpha_1/\sigma_\varepsilon, \alpha_2/\sigma_\varepsilon, \alpha_3/\sigma_\varepsilon, \alpha_4\delta/\sigma_\varepsilon)$. Furthermore, from the approximations in (3.6), the coefficient on $\ln(y_a/y_b)$ in (3.22) estimates $1/\sigma_\varepsilon$ (given that $\alpha_1 = 1$) so that it is also possible to estimate population average real rates of interest. In addition there are 15 parameters in the unobserved component variance-covariance matrix Σ. Following a development similar to the one leading to (3.14) - (3.17) it can be shown that the variances of residuals in (3.20) are

$$\text{var}(\eta_{ij}) = \sigma_{jj} + \frac{\sigma_{j\varepsilon}}{\sigma_\varepsilon}\left[\frac{W_i\pi}{\sigma_\varepsilon}\lambda_{ai} - \lambda_{ai}^2\right], \quad j = 1, 2$$

$$\text{var}(\eta_{ij}) = \sigma_{ij} + \frac{\sigma_{j\varepsilon}}{\sigma_\varepsilon}\left[\frac{W_i\pi}{\sigma_\varepsilon}\lambda_{bi} - \lambda_{bi}^2\right], \quad j = 3, 4. \tag{3.27}$$

Similar expressions hold for covariances between η_{i1} and η_{i2} and between η_{i3} and η_{i4}. Hence it is possible to estimate the own population variances σ_{jj} for $j = 1,\ldots,4$, two within group covariances, and four covariances $\sigma_{j\varepsilon}$ for $j = 1,\ldots,4$. These along with the estimate of σ_ε provide only 11 statistics to estimate 15 parameters. Evidently all the covariance terms in Σ cannot be estimated without additional zero or other restrictions because we never

observe the path not taken. This is the basis for the statement above that dead weight losses from assignments based jointly on wealth and talent rather than on talent alone cannot be imputed. The demand function for college attendance implicit in (3.22) reflects the joint density of talent, wealth, tastes, and expectations and their separate effects cannot be disentangled.

IV. ESTIMATION

The model has been estimated on a sample of 3611 respondents to the NBER-Thorndike-Hagen survey of 1968-71.[16] These data refer to male WWII veterans who applied for the Army Air Corps. They do not come from a random sample of the population since the military screening criteria were based on certain aspects of ability and physical fitness. Therefore it is not possible to extrapolate these results to the population at large. However, the sample's advantages more than compensate for this. First, it covers more than twenty years of labor market experience, far longer than any other panel of comparable size, and most appropriate for measuring lifetime earnings effects of educational choice as the theory requires. Second, it contains extensive information on family background and talent. While several other panels are as good on family background, virtually none compare in their range of talent and ability indicators most appropriate to the theory of comparative advantage.

The sample actually used is a subset of 5085 total respondents. Forty-two observations were dropped for not responding to the age question, another 480 persons were deleted because they were pilots, had extended military service, or did not report a job in 1969, and 952 were dropped because they

did not report both initial ($\bar{y}$) and latest ($y(\bar{t})$) earnings required for structural estimation. Definitions of variables are given in Table I. Individuals were put into two categories: Group A represents those who entered college and Group B, those who stopped school after high school graduation. Not all members of Group A completed college and a substantial fraction completed more than a college education. They are labeled college attendees hereafter. Descriptive statistics also appear in Table I. Notice that more than 75% of the sample chose to attend college for some period, reflecting the unusual ability distribution in the sample and eligibility for a liberal college subsidy (the G.I. Bill). However, the presence of the G.I. Bill is common to both college attendees and high school graduates.

There are some obvious differences between the two groups. Both mean and relative variance of earnings in both years is smaller for high school graduates, as tends to be true in other samples. In addition, high school graduates had smaller earnings growth over the period, had more siblings and were lower in birth order than college attendees, and were more likely to have taken vocational training in high school. Their fathers had less schooling and were more likely to be blue-collar workers as well. Four ability measures have been chosen for analysis, out of some 16 indicators available in the data. Math and reading scores are related to IQ-type of ability (in fact it is known that math score is highly correlated with IQ score in these data), while the other two are more associated with manual skills. The four together seem well suited to the comparative advantage logic underlying the formulation of the model. High school graduates tend to score lower in the math and reading comprehension tests, about the same in manual dexterity and somewhat better on mechanical ability. In line with the previous discussion all ability measures

in Table I are assigned to X, while the family background measures--reflecting financial constraints, tastes and perceptions--are assigned to Z. Experience, school completion dummies (for Group A) and year of reported earnings are used exclusively as controls in structural earnings equations.

The first columns in Table II present estimated coefficients and asymptotic t-statistics of the reduced form probit selection into Group A--equation (3.12). These effects more or less parallel the summary of Table I given above. Math score has a particularly strong positive effect and mechanical score a strong negative effect on the college attendance decision. The effect of mother's working is somewhat unexpected. Mother's home time when the respondent was five years old or younger has virtually no effect on college attendance, whereas the respondent was more likely to go to college if his mother worked when he was 6-14 years of age. This is more supportive of market investment through relaxation of financial constraints rather than of home investments in kind.[17]

Structural estimates of earnings and growth equations corrected for selection are found in Table III. These are somewhat different from the typical earnings equations found in the literature, because they include a much sparser set of regressors. For example, we know respondents' unemployment experience, weeks worked, weeks ill, marital status and so forth but have not included them in the regressions. The logic of this lies in the model itself: At the time the college attendance decision was made, there is no reason to expect that respondents knew the outcomes of such variables. It is more in the spirit of the choice framework of the model to allow these "current" events to be captured indirectly via their correlations with included variables in order to estimate expected or anticipated values relevant to the structural probit.[18]

The problem is more difficult in the case of school completion differences among members of Group A in Table III, and in truth raises an unresolvable aggregation problem. The anticipations argument above suggests that school completion differences within Group A may not enter the earnings equations, so that included variables pick up average completion experience in the sample. Alternatively it can be argued that the level of schooling achieved within Group A should be controlled by including school completion dummies. This latter specification is reported in Table III and is the one used to estimate the structural probit in Table II. Of course we do not switch on the school completion dummies to estimate the earnings advantages of college attendance, since that would clearly stack the deck in favor of finding strong financial effects. Earnings and structural probit equations were also estimated with school dummies deleted and the results were very similar to those reported here. However, it is clear that this issue only can be resolved by going into a more disaggregated model with multiple classifications.

With the exception of experience, most of the variables have little effect on initial earnings in either A or B (See columns 1 and 2 of Table III).[19] Experience effects are the strongest and are known to be most important at early and late stages of career patterns, fact borne out in these data since experience has little effect on later (surveyed around 1969) earnings. The ability measure that has the largest effect on initial earnings is math score for college attendees. Ability indicators are more important for earnings growth (columns 3 and 4) and later earnings (columns 4 and 5). Dexterity and reading scores have positive effects on g_b and $y_b(\bar{t})$, while math and reading scores have positive effects on $\ln y_a(\bar{t})$, but exhibit much weaker effects on earnings growth. Interestingly enough, the effect on mechanical

score is negative in all cases, raising obvious questions about what it is that this test supposedly measures (recall, however, the sample truncation on high ability-military personnel). Even so, it seems to have a more important negative effect for members of Group A. This along with the results for dexterity and math scores lend support to the comparative advantage hypothesis.

Selectivity biases are particularly interesting in that regard. The coefficients of λ_b show no selectivity bias for initial earnings of high school graduates, but positive bias for growth rates. Therefore, observed earnings patterns of high school graduates show higher rates of growth compared with the pattern that would have been observed for the average member of this sample had he chosen not to continue school. On the other hand the coefficients of λ_a show positive selection bias for initial earnings of college attendees and negative bias for earnings growth. The latter is due to the fact that there is no selection effects for late earnings. Thus observed earnings patterns among members of Group A is everywhere higher than the population mean pattern would have been, and converges toward the population mean late earnings level. Positive selection among both A and B also lends support to comparative advantage.

The most striking empirical results are the structural probit estimates in Table II, which show how anticipated earnings gains affect the decision to attend college. The predicted earnings variables are statistically significant except for g_b in (3.22) and g_a in (3.26).[20] More striking however, is the agreement of the sign patterns predicted by the theory (See equation (3.6) and recall that the structural probit coefficients are normalized by σ_ε, from (3.22) and (3.25)). The model passes two internal

consistency checks. The first is restrictions (3.26). Working backwards to normalized α estimates from directly estimated θ's in column five of Table II yields[21] a predicted $(\alpha/\sigma_\varepsilon)$ vector of (5.15, 117.62, -38.06) which is remarkably similar to the direct estimates in column three of (5.15, 138.39, -44.27). Working forwards from actual estimates of normalized α to predicted estimates of θ gives prediction (5.15, 28.46, 65.65) compared with actual (5.15, 7.66, 71.90). Certainly these comparisons would not be so close if the two parameter approximation to earnings patterns in (3.1) and (3.2) was not reasonably good. Second, equations (3.12) and (3.22) indicate that estimated coefficients on the Z variables in structural and reduced from probits should be the same. Direct comparison of coefficients of Z in Table II shows extremely close similarity of $\alpha_4\delta$ in all three equations. In sum, the results give direct, internally consistent evidence on the validity of the economic theory of the demand for schooling derived from its (private) investment value. The economic hypothesis is strongly accepted.

V. CONCLUSIONS

The structural probit estimates of Table II support the economic hypothesis that expected gains in life earnings influence the decision to attend college. They also show important effects of financial constraints and tastes working through family background indicators, a finding in common with most other studies of school choice.[22] Availability of the G.I. Bill might well be expected to dull the observed monetary effects, but they remain strong enough to persist for a significant fraction of the sample.

The estimates also show positive sorting or positive selection bias in observed earnings of both high school graduates and college attendees.

High school graduates had larger lifetime earnings than the average college graduate would have experienced had he chosen to stop school after high school. Conversely the average college graduate's observed earnings were greater than what the typical high school graduate would have earned had he continued school. It is difficult to tell a coherent story for the observed pattern of selection in these data based on the usual discussion of financial constraints and/or hierarchical ("one factor") ability considerations. The most attractive and simplest interpretation is the theory of comparative advantage because hierarchical assignments are not observed. While the results are consistent with comparative advantage, they do not prove the case because life persistent luck and random extraneous opportunities could have played just as important roles in the observed assignments as differential talents did. For all we know those who decided to stop school after high school may have married the boss' daughter instead, or made better career connections in the military, and so forth. The important point is that their prospects in B were higher than average.

As noted above, the population average rate of discount, $\bar{r}$, is an identifiable statistic in the model. Estimates are obtained by applying restrictions (3.6) to the estimates in Table II. Maintain the hypothesis that $\alpha_1 = 1$. Then the estimated coefficient of $\ln(\bar{y}_a/\bar{y}_b)$ in Table II estimates $(1/\sigma_\varepsilon)$, from equation (3.22). Since all the equations of the structural probit are normed by σ_ε this estimate provides a basis for estimating the population parameters in (3.6).

Straightforward computations using the observed mean schooling increment of 4.11 years yields

$$(\bar{r} - \bar{g}_a) = .0439$$
$$(\bar{r} - \bar{g}_b) = .1163 \qquad (5.1)$$

Observed growth rates do not estimate population means because of truncation and selection. However, (3.15) and (3.17) imply

$$\tilde{g}_a = \bar{g}_a + \frac{\sigma_{\varepsilon}}{\sigma_{\varepsilon}} \cdot \tilde{\lambda}_a$$
$$\tilde{g}_b = \bar{g}_b + \frac{\sigma_{4\varepsilon}}{\sigma_{\varepsilon}} \cdot \tilde{\lambda}_b \qquad (5.2)$$

where $\tilde{g}$ and $\tilde{\lambda}$ are observed sample means and the estimated coefficients $(\sigma_{k\varepsilon}/\sigma_{\varepsilon})$ are from Tables III and IV.

Apply formula (5.2) to the data in Table I to obtain corrected mean estimates of $\bar{g}_a = .0555$ (compared with an observed mean of .0535) and $\bar{g}_b = .0175$ (compared with .031 observed). The population mean discount rate, $\bar{r}$, is over-identified. The first equation of (5.1) yields an estimate of $\bar{r} = .0994$ while the second yields $\bar{r} = .1338$. Two more estimates of $\bar{r}$ are implied by the structural probit that uses the late earnings difference rather than the initial earnings difference. Using restrictions (3.26) and the corrected estimates of $\bar{g}_a$ and $\bar{g}_b$ above yields $\bar{r} = .1089$ and of $\bar{r} = .1578$. Even if the precise derivation and specification of the model in Section III strains the reader's credulity, it is nonetheless clear that the structural specification is consistent with more casual derivations and the estimated sign patterns in the structural probit, if not the precise restrictions among coefficients,

would be predicted by virtually any economic model.

The positivity of earnings selection effects in both groups also implies that selection bias in simple rate of return estimates could go in either direction. The following procedure gives a rough and ready indication in this sample. First the 2-parameterization of earnings in (3.1) and (3.2) implies that the average internal rate of return, i, is estimated by

$$(\bar{y}_a/\bar{y}_b)\ [(i - g_b)/(i - g_a)]\exp(g_a - i)S = 1.0$$

where i is the rate of discount that equates average present values. The unadjusted estimate uses sample means in Table I for the $\bar{y}$'s and the g's and a schooling increment of 4.11 years. It is 11 percent. An adjusted rate is obtained by using the corrected mean values of the g's above and a similar correction applied to the $\bar{y}$'s. The resulting estimate of the population mean internal rate of return is 12 percent, which is actually larger, not smaller than the observed mean rate of return.

Predictions

The model passes the test of empirical verification of its structural restrictions. How well does it do in predicting assignments on independent data? The sample used is not a random drawing of the U.S. population and for this reason cannot be extrapolated to the population at large. However, only a subset of the NBER-Thorndike-Hagen sample was used to estimate it and the remaining remnant is more likely to be a suitable group for prediction purposes. The remnant refers to those who did not report initial earnings. For this reason it may not be a random sample of the relevant population either. And while there is no reason to suppose that the censoring of initial

earnings was systematically related to the selection mechanism of the model, it should be noted that a somewhat smaller proportion of these individuals (66% of them) chose to attend college than in the sample used for structural estimation.

One indirect test of the model's predictive content has been calculated. First, the reduced form probit was re-estimated for the remnant, which does not involve extrapolations, since the sample selection between A and B and the content of $W = [X,Z]$ is known for these people. Results appear in Table AI. While there is some conformity with Table II, there are also many differences between reduced form estimates in the two samples. In short, family background coefficients are not too stable.

The second experiment involves an extrapolation. Both initial earnings differences and growth rates were predicted for members of the remnant sample from the structural earnings estimates of Table III and then used to re-estimate the structural probit of this group (No t-statistics are reported for structural probit coefficients because of the large expense of doing so). The results also appear in Table AI. The sign reversals on family background indicators carry over to these estimates too. Indeed, the coefficients and signs of the Z variables in the structural estimates are very close to those found in the reduced form estimates in Table AI. However, the coefficients on the earnings differences and growth rates for the remnant sample are very close to those estimated for the original sample of Table II.

Enrollment Functions

Perhaps the simplest and most useful summary of the results is obtained from the demand function for college attendance implicit in the structural probit estimates. Recalling the definition of the index function in (3.5), the

probability of attending college is given by

$$\Pr(\text{A is chosen}) = F(I/\sigma_\varepsilon)$$

where F is the standard normal cdf. Let m denote the size of the relevant population and let N represent the number choosing to attend college. Then the number enrolled in college is given by

$$N = mF(I/\sigma_\varepsilon) \tag{5.3}$$

This would be equivalent to a supply function of graduates were it not for the aggregation involved in Group A. The supply of graduates is somewhat different since we do not know how long people outside the sample would stay in school. The normality assumptions imply that the enrollment function (5.3) follows the cumulative normal curve. It therefore has zero elasticities at its extremes and positive elasticities in between. The major point of interest here is responsiveness of enrollments to earnings opportunities near the sample mean. From the definitions of present value in Section III, note that $d\ln(V_a/V_b)/d\ln(\bar{y}_a/\bar{y}_b) = 1$. A one percent change in relative initial earnings changes relative capital values by one percent. To clarify a possible point of confusion on this conceptual experiment, $d\ln(\bar{y}_a/\bar{y}_b)$ represents a permanent-- not a transitory change in lifetime prospects, because it increases relative differences between potential earnings in A compared with B not only initially but forevermore [See (3.1) and (3.2)]. Differentiating (5.3) yields an elasticity formula

$$d\ln N/d\ln(\bar{y}_a/\bar{y}_b) = [F'(I/\sigma_\varepsilon)(\alpha_1/\sigma_\varepsilon)]/F(I/\sigma_\varepsilon)$$

where I/σ_ε is evaluated at the desired sample proportion. For example, the

elasticity evaluated at a sample proportion of .5 (half in A and half in B) is 4.1. On the other hand the initial earnings elasticity at the observed sample proportion is 1.94, still a substantial response given the presence of marked diversity in the population. By way of comparison, an increment of father's education of 1.59 years (the difference in means of father's schooling between groups in Table I) elicits a relative response of .0337.

Footnotes

* Thanks are due to Sean Becketti for excellent research assistance and to Lung Fei Lee for advice on statistical issues. This research was supported by the National Bureau of Economic Research, but is not an official NBER publication. The order of the authors' names was selected by a random device.

1. The equalizing difference model originates with Friedman and Kuznets (1954). Jacob Mincer (1974) has developed it most completely in recent years.

2. See Lillard and Willis (forthcoming) for additional detail and confirmation of these remarks. Related studies have reached similar conclusions, e.g. Lillard and Weiss (1976). Of course, it is conceivable but unlikely that educational wage differentials are exactly equalizing for each individual although considerable lifetime income inequality exists among individuals. This possibility is rejected in the empirical findings presented below.

3. The basic model is discussed in Becker (1975). See Rosen (1977) for an elaboration of this argument, and a survey of the relevant literature. Blaug (1976) also stresses the need for estimating structural demand for schooling relationships, and Griliches (1977) discusses the difficulty of doing so in conventional models. Part of Griliches' discussion is pursued in Griliches, Hall and Hausman (1977). The model elaborated here is conceptually distinct from that work, though some of the statistical techniques are similar. A similar remark applies to the work of Kenney, Lee, Maddala and Trost (1977).

4. There is aggregate time series evidence that earnings are important determinants of professional school enrollment (See Freeman [1968] and numerous subsequent studies by the same author); but there is virtually no micro evidence even though such data has been most often studied in the human capital and signaling framework.

5. Actually, expository convenience dictates a more restrictive formulation than is necessary. X and Z need not be orthogonal. They may have some elements in common, but identification requires that they not have all elements in common. See below.

6. Roy (1951) gives a surprisingly modern and rigorous treatment of a selection problem based on the theory of comparative advantage. See Rosen (forthcoming) for extensions and elaboration on this class of problems. Heckman (1976), Lee (1976) and Maddala (1976) develop the appropriate estimation theory.

7. The problem is that the aggregates are sums of distributions that are themselves truncated and selected. Therefore the distributions underlying the aggregate assignments are not necessarily normal. We are unaware of any systematic analysis of this kind of aggregation problem.

8. Methods such as conditional logit have been designed to handle high dimensioned classifications (MacFadden [1973]) but require independence and other (homogeneity) restrictions that are not tenable for this problem. Hausman and Wise (1978) have worked out computational methods on general normal assumptions for three choices. Note also that maximum likelihood methods are available, but are extremely expensive

because multiple integrals need be evaluated. Hence we follow the literature in using consistent estimators.

9. Wise (1974), Lazear (1975), and Zabalza (1977) have used initial earnings and growth of earnings to study life earnings patterns. The distribution of potential earnings and growth is not constrained in our model, thus for example allowing the possibility that $\bar{y}_a$ and g_a are negatively correlated (and similarly for $\bar{y}_b$ and g_b) as in Mincer (1975). On this see Hause (1977).

10. The τ's of Section II are related to $(u_1,\ldots u_4)$ by a set of implicit prices that vary across school classifications, as in Mandelbrot (1960). See Rosen (forthcoming) for the logic of why these differences in valuation can be sustained indefinitely and cannot be arbitraged.

11. For completeness, $-\varepsilon$ should be redefined to take account of deviations between realizations and expectations at the time school decisions were made. Thus, let $\ln\bar{Y}_{ai} = \ln\bar{y}_{ai} + v_{1i}$, where $\bar{Y}_{ai}$ is realized initial earnings, $\bar{y}_{ai}$ is expected initial earnings and v_{1i} is normally distributed forecast error. Similarly, forecast errors v_{2i}, v_{3i}, and v_{4i} are defined for g_{ai}, $\ln\bar{y}_{bi}$ and g_{bi}. Then the complete definition of $-\varepsilon$ is obtained from replacing u_{ji} with $(u_{ji} + v_{ji})$, $j = 1,\ldots,4$, in (3.11). Clearly this has no operational significance for the model, given the assumption of unbiased expectations.

12. See Heckman (1976) and Lee (1976).

13. This method is due to Lee (1976) who used it to study unionization status.

Our model differs somewhat in that there is more than one structural equation in each classification.

14. Heckman (1977) and Lee (1977) show that OLS estimates of the standard errors of β_a, γ_a, β_b and γ_b in (3.20) are biased if $\sigma_{k\varepsilon}/\sigma_\varepsilon \neq 0$ when estimated values of λ_b are used in place of their true values. Lee also shows that the usual estimates of standard errors for the structural probit (3.22) are biased when estimated values of $\ln(\bar{y}_a/\bar{y}_b)$, g_a and g_b are used in place of their true values and derives exact assymptotic distributions for these parameters. We use Lee's results to compute consistent estimates of standard errors below.

15. Heckman (1976) raises some subtle issues regarding specification error in selection models. Elements of Z may be incorrectly specified in X and can be statistically significant in least squares regressions because of truncation. Conversely, coefficients on selection bias variables λ_a and λ_b can be significant because variables are incorrectly attributed to selection when they more properly belong directly in X. For example, some might argue that family background belongs in structural earnings equations and our selectivity effects work (See below) because family background comes in the back door through its indirect effect on $\hat{\lambda}$. However, a reversal of the argument suggests that family background might have significant estimated direct effects on earnings merely because they work through selection and resulting truncation. There is no statistically satisfactory way of resolving this problem. In any event, we cannot be "agnostic" about specification because both the economic and statistical theory require certain nontestable zero

identifying restrictions. The problem is even more complicated in the present context because the theory is based on unobserved talent and financial constraint shifters and must have observable counterparts to be operational. Evidently choice among alternative specifications ultimately must rest on predictive performance outside the sample.

16. These data have been extensively analyzed by other investigators, especially Taubman (1975), who also discovered them. For complete documentation see NBER (1973).

17. Recall that female labor force participation during the war increased. The normalized category for mother's work classifications is nonresponse. We do not know how many did not respond because no mother was in the home.

18. A related and thorough discussion of this issue appears in Hanoch (1967), to which the reader is referred. It has not escaped our attention that current variables such as hours of work and unemployment experience might serve as indicators of an unobserved "taste for leisure" component, but we have not experimented with that possibility.

19. Initial earnings is recall data from the 1955 Thorndike survey and refers to a period as much as nine years prior to that survey date. Late earnings is closer to the NBER survey date and probably has less recall error in it. The low R^2 statistics in Tables III and IV are due to the fact that we are looking at within-group variation, whereas most results in the literature get a lot of mileage out of current variables and explanation of between-group mean variation. It is also worth noting

that the standard errors in the earnings and growth equations computed from the exact asymptotic distribution reported in the table are virtually identical to those estimated by OLS.

20. Recall (Footnote 14) that the t-statistics for the structural probit in Table II are based on consistent estimates of the standard errors, as suggested by Lee (1977). The t-statistics on background variables are not very different from the biased values computed by a standard probit algorithm. However, the t-statistics on the predicted earnings and growth variables are substantially reduced when corrected for bias; e.g., the standard probit estimates of t-values for $\ln(\bar{y}_a/\bar{y}_b)$, g_a and g_b in (3.22) are [10.8 , 8.15, -4.81] compared with the unbiased values of [2.25, 1.83, -1.28] in Table II.

21. The value of $\bar{t}$ used in these computations is 21.35, which is the total sample mean time between reporting initial earnings and late earnings (the weighted difference in means of Year 48 and Year 69 variables in Table I).

22. See Kohn, Manski and Mundel (1976) and Radner and Miller (1970) for logit models of college choice. These models contain more detail in personal and college attributes but do not make any attempt to assess the effects of anticipated earnings in college attendance decisions. See Abowd (1977) for another approach to the selection problem focusing on school quality.

REFERENCES

Abowd, J.M. "An Econometric Model of the U.S. Market for Higher Education," Ph.D. thesis, University of Chicago, 1977.

Becker, G.S. Human Capital, 2nd ed., New York: NBER, 1975.

Blaug, M. "The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey," Journal of Economic Literature, 1976.

Freeman, R. The Labor Market for College Trained Manpower, Cambridge, Harvard University Press, 1971.

Friedman, M. and Kuznets, S. Income From Independent Professional Practice, New York, NBER, 1954.

Griliches, Z., "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica, 1977.

Griliches, Z., Hall, B., and Hauseman, J. "Missing Data and Self-Selection in Large Panels," Harvard University, 1977.

Hanoch, G. "An Economic Analysis of Earnings and Schooling," Journal of Human Resources, 1976.

Hause, J. "The Fine Structure of Earnings and the On-the-Job-Training Hypothesis," University of Minnesota, 1977.

Hausman, J. and Wise, D. "A Conditional Probit Model for Qualitative Choice," Econometrica, 1978.

Heckman, J. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables," Annals of Economic and

Social Measurement, 1976.

_____. "Sample Selection Bias as a Specification Error" in J.P. Smith, ed., Female Labor Supply: Theory and Estimation. Rand Corporation, 1977.

Kenny, L., Lee, L., Maddala, G.S., and Trost, R. "Returns to College Education: An Investigation of Self-Selection Bias in Project Talent Data," University of Florida, 1977.

Kohn, M.G., Manski, C.F., and Mundel, D.S. "An Empirical Investigation of Factors Which Influence College-Going Behavior," Annals of Economic and Social Measurement, 1976.

Lazear, E. "Age, Experience and Wage Growth," American Economic Review, 1976.

Lee, Lung Fei "Estimation of Limited Dependent Variables Models by Two-Stage Methods," Ph.D. thesis, University of Rochester, 1976.

_____. "Unionism and Wage Rates," International Economic Review (forthcoming).

_____. "On the Asymptotic Distributions of Some Two-Stage Consistent Estimators: Unionism and Wage Rates Revisited," University of Minnesota, 1977.

Lillard, L. and Weiss, Y. "Analysis of Longitudinal Data: American Scientists 1960-70," NBER, 1976.

Lillard, L. and Willis, R. "Dynamic Aspects of Earnings Mobility," Econometrica (forthcoming).

McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.) Frontiers in Econometrics, North-Holland, 1973.

Maddala, G.S. "Self-Selectivity Problems in Econometric Models," University of Florida, 1976.

Mandelbrot, B. "Paretian Distributions and Income Maximization, Quarterly Journal of Economics, 1960.

Mincer, J. Schooling, Experience and Earnings, New York, NBER, 1974.

National Bureau of Economic Research, "The Comprehensive NBER-TH Tape Documentation," March, 1973.

Radner, R. and Miller, L.S. "Demand and Supply in U.S. Higher Education," American Economic Review, Vol. 60, May, 1970.

Rosen, S. "Human Capital: Relations Between Education and Earnings," in M. Intilligator (ed.) Frontiers of Quantitative Economics, Vol. IIIB, North-Holland, 1977.

_____. "Substitution and Division of Labor," Economica (forthcoming).

Roy, A.D. "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers, 1951.

Wise, D. "Academic Achievement and Job Performance," American Economic Review, 1975.

Zabalza, A. "The Determinants of Teacher Supply," London School of Economics, 1977.

TABLE I: DESCRIPTIVE STATISTICS[a]

	High School (Group B)		More Than High School (Group A)	
Variable	Mean	Standard Dev	Mean	Standard Dev
Father's ED	8.671	2.966	10.26	3.623
Father's ED^2	83.99	55.53	118.4	78.09
DK ED	.0999		.0464	
Manager	.3628		.4954	
Clerk	.1239		.1450	
Foreman	.2238		.1695	
Unskilled	.1492		.0819	
Farmer	.1062		.0720	
DK Job	.0177		.0124	
Catholic	.2933		.2138	
Jew	.0405		.0617	
Old Sibs	1.143	1.634	.9035	1.383
Young Sibs	.9381	1.486	.8138	1.266
Mother Works:				
Full 5	.0468		.0486	
Part 5	.0392		.0504	
None 5	.7168		.7507	
Full 14	.0822		.0936	
Part 14	.0708		.0851	
None 14	.6384		.6713	
H.S. Shop	.2592		.0908	
Read	20.57	10.17	24.06	11.63
NR Read	.0291		.0128	
Mech	59.24	18.27	58.88	18.96
NR Mech	.0253		0.	
Math	18.13	11.82	28.94	17.17
NR Math	.0683		.0188	
Dext	50.04	9.359	50.68	9.811
NR Dext	0.		.0071	
Exp	29.33	2.439	24.54	2.907
Exp^2	866.1	147.1	610.4	147.4
S13-15			.3106	
S16			.3993	
S20			.0823	
Year 48	46.62	1.584	48.05	1.869
Year 69	69.11	.3691	69.08	.3437
ln $\bar{y}$	8.365	.4107	8.526	.3871
ln $y(\bar{t})$	9.326	.4573	9.639	.4904
g	.0309	.0251	.0535	.0283
λ_a	-.5651	.2873	-.3193	.2256
λ_b	1.138	.3763	1.605	.5212
No. observations	791		2820	

a. Definition of Variables:

Father's ED	Father's years of school. Nonresponse assigned mean.
Father's ED^2	Square of Father's ED.
DK ED	Dummy variable: 1 if respondent didn't know father's education.
Manager	Dummy variable: 1 if father was a businessman, manager or professional.
Clerk	Dummy variable: 1 if father had white collar occupation other than those in Manager.
Foreman	Dummy variable: 1 if father was a foreman, supervisor or skilled craftsman.
Unskilled	Dummy variable: 1 if father was semiskilled operative or unskilled laborer.
Farmer	Dummy variable: 1 if father was a farmer.
DK Job	Dummy variable: 1 if respondent didn't know father's occupation.

TABLE I (continued)

a. Definition of Variables (continued):

Catholic	Dummy variable: 1 if respondent is Catholic
Jew	Dummy variable: 1 if respondent is Jewish
Old Sibs	Number of older siblings.
Young Sibs	Number of younger siblings.
Mother's Work:	
Full 5	Dummy variable: 1 if mother worked full time when respondent was less than 6 years of age.
Part 5	Dummy variable: 1 if mother worked part time when respondent was less than 6 years of age.
None 5	Dummy variable: 1 if mother did not work when respondent was less than 6 hears of age.
Full 14	Dummy variable: 1 if mother worked full time when respondent was 6-14 years of age.
Part 14	Dummy variable: 1 if mother worked part time when respondent was 6-14 years of age.
None 14	Dummy variable: 1 if mother did not work when respondent was 6-14 years of age.
H.S. Shop	Dummy variable: 1 if respondent majored in vocational courses in high school.
Read	Raw score on college undergraduate level reading comprehension test. Continuous variable, nonrespondents assigned mean.
NR Read	Dummy variable: 1 if reading score not reported.
Mech	Raw score on pictorial representation of mechanical problem test. Continuous variable, nonrespondents assigned mean.
NR Mech	Dummy variable: 1 if mechanical score not reported.
Math	Raw score on mathematics test (performance in advanced arithmatic, algebra and trignometry). Continuous variable with nonrespondents assigned mean.
NR Math	Dummy variable: 1 if math score unreported.
Dext	Score on test of finger dexterity. Continuous variable, nonrespondents assigned mean.
NR Dext	Dummy variable: 1 if dexterity score not reported.
Exp	Continuous variable: Age - Schooling - 6.
Exp^2	Square of Exp.
S13-15	Dummy variable: 1 if respondent received 13-15 years of school.
S16	Dummy variable: 1 if respondent received 16 years of school.
S20	Dummy variable: 1 if respondent received 20 or more years of school.
Year 48	Year in which initial postwar earnings are reported. Continuous variable.
Year 69	Year in which earnings at time of NBER survey was reported. Continuous variable.
Ln $\bar{y}$	Log of earnings on first job after finishing school, in 1967 prices.
Ln y(t)	Log of earnings at time of NBER survey in 1967 prices.
g	(Ln earn 69 - Ln earn 48) ÷ (Year 69 - Year 48) percentage rate of growth between the two observations.
λ_a	See equation (3.13), based on estimates in Table II, Column 1.
λ_b	See equation (3.18), based on estimates in Table II, Column 1.

TABLE II. COLLEGE SELECTION RULES: Probit Analysis

	Reduced Form (3.12)		Structure (3.22)		Structure (3.25)	
	Coefficient	t^a	Coefficient	t^a	Coefficient	t^a
Constant	.0485	.20	.1512	.22	.1030	.17
BACKGROUND						
Father ED_2	-.0145	-.41	-.0168	-.54	-.0152	-.49
Father ED	.0037	2.05	.0038	2.26	.0037	2.26
DK ED[b]	-.4059	-3.96	-.3924	-2.79	-.4001	-2.91
Manager	.1897	2.17	.1825	2.13	.1871	2.21
Clerk	.0556	.54	.0561	.59	.0554	.59
Foreman	.0182	.19	.0210	.23	.0200	.22
Unskilled	-.0910	-.85	-.0948	-.89	-.0928	-.87
Farmer	-.2039	-2.12	-.2256	-2.27	-.2094	-2.14
DK Job[b]	-.0413	-.19	-.0629	-.29	-.0609	-.28
Catholic	-.1144	-1.91	-.0982	-1.51	-.1083	-1.66
Jew	-.0293	-.23	.0143	.12	-.0158	-.14
Old Sibs	-.0162	-.93	-.0162	-.93	-.0161	-.93
Young Sibs	.0122	.63	.0096	.49	.0112	.57
Mother Work:						
Full 5	.1039	.66	.1168	.81	.1104	.76
Part 5	.2179	1.42	.2106	1.52	.2156	1.56
None 5	.0655	.63	.0677	.65	.0661	.64
Full 14	.2898	2.29	.2884	2.30	.2888	2.33
Part 14	.2709	2.20	.2768	2.02	.2693	2.03
None 14	.1980	1.91	.1990	1.92	.1966	1.92
H.S. Shop	-.4411	-6.14	-.4397	-3.74	-.4379	-3.90
ABILITY						
Read	.0047	1.67				
NR Read[c]	-.2575	-1.41				
Mech	-.0070	-4.29				
NR Mech[c]	-3.0236	-1.04				
Math	.0244	12.34				
NR Math[c]	-.7539	-5.75				
Dext	.0019	.72				
NR Dext[c]	2.2797	.47				
EARNINGS						
$\ln(\bar{y}_a/\bar{y}_b)$			5.1486	2.25		
g_a			138.3850	1.83	7.6632	.11
g_b			-44.2697	-1.28	71.8981	2.34
$\ln y_a(t)/y_b(t)$					5.1501	2.57
Observations	3611		3611		3611	
Limit OBS	791		791		791	
Nonlimit OBS	2820		2820		2820	
-2 ln (likelihood Ratio)	579.5		568.8		576.6	
χ^2 degree freedom	28		23		23	

NOTE: a = t is asymptotic t-statistic.

b = DK: Don't know, dummy variable.

c = NR: No response, dummy variable.

TABLE III: STRUCTURAL EARNINGS ESTIMATES:
EQUATIONS (3.20) AND (3.24), OLS (t-values in parenthesis)

Regressor	DEPENDENT VARIABLE					
	(1)	(2)	(3)	(4)	(5)	(6)
	$\overline{\ln y}_a$	$\overline{\ln y}_b$	g_a	g_b	$\ln y_a(\bar{t})$	$\ln y_b(\bar{t})$
Constant	8.7124 (16.51)	2.8901 (1.37)	.1261 (3.90)	.2517 (2.11)	10.3370 (5.52)	7.5328 (2.08)
Read	.0009 (1.21)	-.0019 (-1.17)	.0001 (1.11)	.0003 (3.20)	.0027 (2.80)	.0057 (3.28)
NR Read	.0791 (1.24)	.0506 (.58)	-.0034 (-.76)	.0046 (-.89)	.0033 (.04)	-.0402 (-.42)
Mech	-.0002 (-.48)	-.0005 (-.54)	-.0001 (-2.16)	-.0001 (-1.13)	-.0021 (-3.59)	-.0017 (-1.73)
NR Mech		.1969 (.69)		.0002 (.01)		.2196 (.68)
Math	.0015 (2.02)	-.0013 (.74)	.0001 (1.18)	-.0000 (-.20)	.0030 (3.31)	-.0019 (-1.00)
NR Math	-.1087 (-1.94)	.0562 (.83)	.0015 (.38)	.0006 (.15)	-.0877 (-1.24)	.0712 (.96)
Dext	.0008 (1.03)	-.0019 (-1.21)	-.0000 (-.78)	.0003 (2.77)	.0002 (.16)	.0036 (2.19)
NR Dext	.0751 (.28)		-.0004 (-.02)		.1466 (.43)	
Exp	-.0523 (-1.49)	.4260 (3.10)	-.0028 (-1.11)	-.0154 (-1.93)	-.0129 (-.29)	.0776 (.53)
Exp^2	.0015 (2.22)	-.0067 (-2.95)	.0000 (.21)	.0002 (1.82)	-.0000 (-.01)	-.0012 (-.49)
Year 48	-.0020 (-.48)	-.0156 (-1.72)				
Year 69					-.0067 (-.26)	.0039 (.09)
S13-15	.1288 (5.15)		-.0062 (-3.49)		.0168 (.52)	
S16	.0760 (3.82)		.0026 (1.79)		.1095 (4.26)	
S20	.1318 (4.10)		.0049 (2.13)		.2560 (6.15)	
λ_a	-.1069 (-3.21)		.0058 (2.45)		.0206 (.49)	
λ_b		-.0558 (-.66)		.0118 (2.39)		.2267 (2.48)
R^2	.0750	.0439	.1578	.0513	.0740	.0358

TABLE A1: COLLEGE SELECTION RULES: Probit Analysis

(Independent Subsample of Individuals with no Report on Initial Earnings)

	Reduced Form (3.12)		Structure (3.22)	Structure (3.25)
	Coefficient	t[a]	Coefficient	Coefficient
Constant	-.4424	-.986	-.1170	-.1514
BACKGROUND				
Father ED_2	-.0183	.27	.0131	.0123
Father ED^2	.0020	.61	.0023	.0023
DK ED[b]	-.2645	-1.69	-.2548	-.2608
Manager	.2009	1.50	.1689	.1768
Clerk	.1664	.92	.1523	.1490
Foreman	-.1276	-.83	-.1359	-.1369
Unskilled	-.3118	-1.79	-.3298	.3260
Farmer	.1353	.75	.1174	-.1332
DK Job	-.3515	-1.04	-.3133	-.3426
Catholic	-.0887	-.80	-.0847	-.1024
Jew	-.2169	-.95	-.1879	-.2159
Old Sibs	.0335	1.02	.0343	.0336
Young Sibs	.0191	.56	.0170	.0176
Mother Work:				
Full 5	-.6039	-2.06	-.6080	-.6080
Part 5	-.0470	-.18	-.0409	-.0351
None 5	-.0200	-.11	-.0345	-.0248
Full 14	.1656	.67	.1747	.1764
Part 14	-.1248	-.58	-.1258	-.1310
None 14	-.0581	-.31	-.0360	.0448
H.S. Shop	-.5387	-3.95	-.5436	-.5395
ABILITY				
Read	.0056	1.07		
NR Read[c]	.2393	.74		
Mech	-.0480	-1.64		
NR Mech[c]	-----	----		
Math	.0251	6.80		
NR Math[c]	-.4775	-2.15		
Dext	.0050	1.03		
NR Dext[c]	-----	----		
EARNINGS				
$\ln(\bar{y}_a/\bar{y}_b)$			4.9674	
g_a			122.1460	-1.8761
g_b			-34.8393	76.4555
$\ln[y_a(t)/y_b(t)]$				4.8837
Observations	952		952	952
Limit OBS	321		321	321
Nonlimit OBS	631		631	631
-2 ln (likelihood Ratio)	184.446		179.419	184.446
χ^2 degree freedom				

NOTE: a = t is asymptotic t-statistic

b = DK: Don't know, dummy variable.

c = NR: No response, dummy variable.