University of Wisconsin-Madison

IRP Discussion Papers

Anders Björklund
Robert Moffitt

ESTIMATION OF WAGE GAINS AND
WELFARE GAINS FROM
SELF-SELECTION MODELS

DP #735-83

Institute for Research on Poverty
Discussion Paper 735-83

THE ESTIMATION OF WAGE GAINS AND WELFARE GAINS FROM SELF-SELECTION MODELS

Anders Björklund
The Industrial Institute for Economic and Social Research and
The Institute for Social Research in Stockholm

Robert Moffitt
Department of Economics
Rutgers University

August 1983

The authors would like to thank the participants of workshops at the University of Wisconsin and Mathematica Policy Research for comments, as well as Charles Brown, Randy Brown, Christopher Flinn, and Arthur Goldberger. The research was partially supported by the Institute for Research on Poverty at the University of Wisconsin.

Abstract

In this paper we consider the basic self-selection model for the effects of education, training, unions, and other activities on wages. We show that past models have ignored "heterogeneity of rewards" to the activity--i.e., differences across individuals in the rate of return to the activity--as a source of selection bias. We model such heterogeneity, show how its presence can be tested, and draw out its implications for the wage and welfare gains to the activity. An empirical application provides strong support for such heterogeneity.

THE ESTIMATION OF WAGE GAINS AND WELFARE GAINS
FROM SELF-SELECTION MODELS

Economists are often interested in estimating the effect of various types of choices on wages. In labor economics, applications frequently have been made in four areas: (1) education, (2) unions, (3) manpower training, and (4) migration. Researchers on these subjects have become increasingly concerned with the potential self-selection that may arise, mainly because the decisions are made by the individuals themselves. In general, self-selection has been regarded as a disturbing problem for the issue under examination, for ordinary least squares (OLS) or otherwise unadjusted estimates of the parameters of interest are biased if self-selection is present. The usual remedies have been to control for self-selection either by applying the techniques developed by Maddala and Lee (1976), Heckman (1978, 1979), and Lee (1979) (see also Barnow, Cain and Goldberger, 1980), or by trying to avoid the problem by using panel data. Examples of the first approach are Willis and Rosen (1979) and Kenny et al. (1979) for education, Lee (1978) for unions, Nakasteen and Zummei (1980) for migration, and Mallar, Kerachsky, and Thornton (1980) for a jobs program. Examples of the second approach are Kiefer (1979), Bassi (forthcoming), and Nickell (1982) for manpower training.[1]

In this paper we demonstrate the importance of the selection mechanism per se in these types of problems. Our primary goal is to demonstrate the implications of interpreting the self-selection model as a basic model of <u>consumer demand</u>. In the context of the consumer-demand model we show that selection bias occurs because population heterogeneity causes different individuals to make different choices--the total population of individuals divides itself into participants and nonparticipants (college attendees and nonattendees, trainees and nontrainees, union members and nonunion members, etc.). Our most important point is that population differences arise either because of what we term "heterogeneity of rewards"--that is, heterogeneity in the rate of return to the activity--or because of "heterogeneity of costs" to undertake the activity (both monetary and nonmonetary) or both.[2] We then show that this distinction has important implications both for the econometric specification of the model and for the interpretation of the results--in particular, for the appropriate estimation of (1) the wage gain to the activity and (2) the welfare gain of the activity. The presence of costs creates a wedge between the welfare gain and wage gain. Our initial point that the selection mechanism is important per se follows from the fact that the estimation of welfare gains requires structural estimates of the selection model. Wage gains, even if obtainable by OLS of the wage equation, are not enough.

The idea of heterogeneity of rewards, which is perhaps our most important idea, corresponds in regression terms simply to a random coefficients model. We allow the return to education or manpower training or some other activity to vary across individuals. We then explicitly model the endogeneity of education (or the decision to participate in a training program) by allowing it to be a <u>function</u> of the random coefficient, or the reward. Thus, rather than specifying an ad hoc equation for education choice, we relate education choice directly to the parameters in the earnings function. This generates a set of cross-equation constraints between the earnings equation and the education-choice equation which we impose in the estimation.

The presence of heterogeneity in the return also has strong implications for public policy, for it implies that those already participating are, in general, those with the highest return. Expanding the participant population--such as by providing educational subsidies or higher stipends to trainees--draws into the activity those who get less out of it. One of the strengths of our model is that it makes this point explicit. Indeed, with our model we can estimate both the mean rewards for those currently participating as well as the reward for those who would participate if the costs of participating were lowered.

In the next section we present our model. Then we provide an empirical illustration, using the case of manpower training. Our empirical results indicate rather dramatically that heterogeneity of rewards are present. We end with suggestions for future research.

I. HETEROGENEITY IN SELF-SELECTION MODELS

As a point of departure we let the individual maximize the utility function $U(Y_i - \phi_i T)$, where T is a dummy variable for participation in the activity, ϕ_i denotes the costs of participating in the activity for individual i, and Y denotes the wage. We assume that ϕ_i captures both monetary costs and a monetized utility component. Let α_i be the earnings gain from participation. Thus, the individual participates in the activity if

$$U(Y_i + \alpha_i - \phi_i) > U(Y_i) \text{ or } \alpha_i > \phi_i$$

where Y_i is now interpreted as earnings in the absence of participation.

This simple analog with demand theory suggests two sources of heterogeneity that might divide the population into participants and nonparticipants: (1) heterogeneity in rewards (α_i) and (2) heterogeneity in costs or preferences (ϕ_i). In terms of elementary demand theory it is obvious that people make different choices if they either face different "prices" in the budget constraint, α_i, or have different utility functions or different costs, in our case ϕ_i.

Consider now one specification that has been used in the research in the four areas mentioned at the beginning of this paper:[3]

$$Y_i = X_i\beta + \alpha T_i + \varepsilon_i \tag{1}$$

$$\begin{aligned} T_i &= 1 \text{ if } T_i^* > 0 \\ T_i &= 0 \text{ if } T_i^* \leq 0 \end{aligned} \tag{2}$$

$$T_i^* = W_i\eta + v_i \tag{3}$$

where Y_i is the wage of individual i; X_i is a vector of wage determinants and β is its associated coefficient vector; T_i is a dummy variable for education (e.g., high school or college), union status, manpower-training participation, or residence; T_i^* is a latent variable determining the dichotomous variable T_i; W_i is a vector of variables affecting T_i and η is its associated coefficient vector; and ε_i and v_i are error terms. We shall formulate our models in terms of a dummy variable, T_i, throughout, but it will be clear that the analysis would carry through equally if T_i were continuous.

The model in (1) - (3) implies a specific form of self-selection. It is obvious that the wage gain, α, is constant and equal for all. In terms of demand theory all individuals face the same price of non-participation. Hence some dispersion or heterogeneity in preferences or other costs must be present. Therefore we can only interpret $W_i\eta$ and v_i in the choice equation as observed and unobserved costs, respectively. This follows because the specification of the choice equation in terms of our framework should instead be:[4]

$$T_i^* = \alpha - W_i\eta - v_i,$$

where T_i^* is the net reward. The assumption of homogeneous rewards is in our view rather restrictive. In all the applications we have mentioned it can be argued that every individual is unique in terms of his skills and labor-market situation. Therefore it is reasonable to allow the wage gains to differ between individuals. A straightforward specification allowing for both observed and unobserved heterogeneity would be:

$$\alpha_i = Z_i\delta + u_i \tag{4}$$

where Z_i is a vector of observed variables, δ is its coefficient vector, and u_i is an error term. Reformulating the model with (4) gives us the following:

$$Y_i = X_i\beta + \alpha_i T_i + \varepsilon_i \tag{5}$$

$$\left.\begin{aligned} T_i &= 1 \text{ if } T_i^* > 0 \\ T_i &= 0 \text{ if } T_i^* \leq 0 \end{aligned}\right. \tag{6}$$

$$T_i^* = \alpha_i - \phi_i \tag{7}$$

$$\alpha_i = Z_i\delta + u_i \tag{8}$$

$$\phi_i = W_i\eta + v_i \tag{9}$$

$$E(\varepsilon_i) = E(u_i) = E(v_i) = 0$$

$$E(v_i^2) = \sigma_v^2 \qquad E(u_i^2) = \sigma_u^2 \qquad E(\varepsilon_i^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_i u_i) = \sigma_{\varepsilon u} \qquad E(\varepsilon_i v_i) = \sigma_{\varepsilon v} \qquad E(u_i v_i) = \sigma_{uv}$$

In reduced form the model comes down to:[5]

$$Y_i = X_i\beta + Z_i\delta + \varepsilon_i + u_i, \text{ if } T_i = 1 \tag{10}$$

$$Y_i = X_i\beta \qquad + \varepsilon_i \qquad , \text{ if } T_i = 0 \tag{11}$$

$$\begin{aligned} T_i &= 1 \qquad , \text{ if } T_i^* > 0 \\ T_i &= 0 \qquad , \text{ if } T_i^* \leq 0 \end{aligned} \tag{12}$$

$$T_i^* = Z_i\delta - W_i\eta + u_i - v_i \tag{13}$$

Here we have let both rewards and costs be a function of a vector of observed variables (Z and W respectively) and its own individual error term. The relative importance of δ and σ_u^2, on the one hand, versus η and σ_v^2, on the other hand, provides the empirical basis for judging whether heterogeneity of rewards or costs is more important. The test for heterogeneity of rewards is the test of the full model (10)-(13) against a model with the restrictions $\sigma_u = \sigma_{\varepsilon u} = \sigma_{uv} = 0$ and $\delta = 0$ except for a constant. The test for heterogeneity of costs is the test of the full model against one with the restrictions $\sigma_v = \sigma_{\varepsilon v} = \sigma_{uv} = 0$ and $\eta = 0$ except for a constant.[6]

This formulation of the problem alters in a rather interesting fashion the interpretation of the estimated parameters from that usually given. First, note that there is no longer a single "effect" of the program since rewards are heterogeneous. Of course, we can speak of a mean reward, or the reward for an individual with a given characteristics vector Z_i. This could be calculated from the estimated parameter vector δ. The "average" reward is, we assume, that to which the usual constant-parameter estimate in the original equation (1) must correspond. But note that it could easily be negative, zero, or positive but small. Since an individual chooses T=1 only if the unobserved component u_i is sufficiently high, there is no reason for the mean reward to be positive. This obviously has major implications for the interpretation of the wage coefficients in previous studies. A more relevant measure of the rate of return in the mean size of the reward conditional upon choosing T=1:

$$\begin{aligned} E(\alpha_i \mid T_i^* > 0, Z_i\delta, W_i\eta) &= Z_i\delta + E(u_i \mid u_i - v_i > -Z_i\delta + W_i\eta) \\ &= Z_i\delta + (\sigma_{u,u-v}/\sigma_{u-v})[f(s)/(1-F(s))] \end{aligned} \tag{14}$$

where $s = (-Z_i\delta + W_i\eta)\sigma_{u-v}$, and f and F are the standard normal density and distribution functions, respectively (we have assumed normality for the errors). This expression is, we argue, the appropriate measure of the expected wage gain from the activity T.

Likewise, note that the T_i^* quantity in equation (6) is simply the dollar amount of the net reward (net of costs, that is). At mean values or any other values of Z_i and W_i in the population, it may be negative even if the reward α_i is positive. But we can use it to determine the dollar gain from the activity T if an individual undergoes it, for T_i^* is just the amount by which income must be reduced in order to leave the individual indifferent to participation (i.e., at the same utility level in either case). This value corresponds in consumer-demand theory to the compensating variation, or "willingness to pay" for the activity. This consumer-surplus measure can be calculated as:

$$\begin{aligned} E(T^*_i \mid T^*_i > 0, Z_i\delta, W_i\eta) &= Z_i\delta - W_i\eta + E(u_i - v_i \mid u_i - v_i > -Z_i\delta + W_i\eta) \\ &= Z_i\delta - W_i\eta + \sigma_{u-v}[f(s)/(1 - F(s))] \end{aligned} \tag{15}$$

where again $s = (-Z_i\delta + W_i\eta)/\sigma_{u-v}$. This welfare gain will equal the wage gain only in the special case $\sigma_v = \sigma_{uv}$, $\eta = 0$ (i.e., no costs).

Finally, some comments on the role of selection bias in our full model are warranted. The term selection bias generally refers to the bias of the OLS estimates of the parameters β and δ in the wage equation that arises if the error terms of these equations are correlated with the error term of the choice equation. In our full model it appears that if there is unobserved heterogeneity of rewards, the error term u_i will appear in both equations. Hence selection bias will be present. It is also important to emphasize that this source of selection bias cannot be avoided by using panel data and working with first differences--that is, by formulating the dependent variable as the difference in earnings from one time period to another. Such a model can be exactly represented by (10)-(13), with Y_i replaced by ΔY_i and by reinterpreting ε_i as the difference in the level errors in two periods.[7]

On the other hand, there is also selection bias in the β and δ coefficients if the unobserved costs, v_i, are correlated with the error term in the earnings equation, ε_i. This is the more usual case of selection bias. It will be eliminated by employing a first-difference technique if v_i is correlated only with some permanent component in the level error. However, note that even if there is no such correlation, the complete self-selection model (10)-(13) must still be estimated if we want to compute our measure of consumer surplus. Hence we conclude that the self-selection model is important per se irrespective of any bias of OLS estimates of the wage equation.

Identification and Estimation

The identification conditions in the full model (10)-(13) are virtually identical to those in the Lee (1979) model and therefore need little discussion. From our two earnings equations (10) and (11), it is clear that the coefficient vectors β and δ are identified, as are their error variances, $(\sigma_\varepsilon^2 + 2\sigma_{\varepsilon u} + \sigma_u^2)$ and σ_ε^2. In equation (13) the vector of parameters η is identified only if there is at least one variable in Z_i that is not in W_i (a similar condition appears in the Lee model). The variance in the same equation, $(\sigma_u^2 - 2\sigma_{uv} + \sigma_v^2)$ is also identified,[8] as are the covariances between the error in equation (13) and those in equations (10) and (11). From these composite variances it can be shown that some normalization is necessary for complete model identification.[9] We have chosen $\sigma_{uv} = 0$.[10] Subject to this normalization we can identify σ_ε, σ_u, σ_v, $\rho_{\varepsilon u}$, and $\rho_{\varepsilon v}$.

The estimation of the model is also no more or less difficult than the estimation of the usual self-selection model. The model can be estimated with either full-information or limited-information techniques. In the first case the log likelihood function is maximized with respect to the unknown parameters (β, δ, η, σ_ε, σ_u, σ_v, $\sigma_{\varepsilon u}$, $\sigma_{\varepsilon v}$):

$$L = \sum_{T=1} \log P_1 + \sum_{T=0} \log P_0$$

where[11]

$$P_1 = \text{Prob}(\varepsilon_i + u_i = Y_i - X_i\beta - Z_i\delta,\ u_i - v_i > W_i\eta - Z_i\delta)$$

$$P_0 = \text{Prob}\ (\varepsilon_i = Y_i - X_i\beta,\ u_i - v_i \leqslant W_i\eta - Z_i\delta).$$

The evaluation of P_1 and P_0 in terms of normal probabilities and densities is shown in Appendix A.

In the limited-information case a three-step method can be used (Lee, 1979). In the first step a reduced-form version of equation (13) is estimated with probit:

$$T_i = 1 \text{ if } T_i^* > 0$$

$$T_i = 0 \text{ if } T_i^* \leqslant 0$$

$$T_i^* = V_i\gamma + u_i - v_i$$

where V_i is the union of the vectors Z_i and W_i. In the second step an augmented earnings equation is estimated with OLS:[12]

$$Y_i = X_i\beta + T_iZ_i\delta + \psi_1\lambda_1T_i + \psi_2\lambda_2(1-T_i) + \xi_i$$

where

$$\lambda_1 = f(-V_i(\hat{\gamma/\sigma_{u-v}}))/[1-F(-V_i(\hat{\gamma/\sigma_{u-v}})]$$

$$\lambda_2 = -f(-V_i(\hat{\gamma/\sigma_{u-v}}))/F(-V_i(\hat{\gamma/\sigma_{u-v}}))$$

$$\psi_1 = \frac{\sigma_{\varepsilon+u,u-v}}{\sigma_{u-v}}$$

$$\psi_2 = \frac{\sigma_{\varepsilon,u-v}}{\sigma_{u-v}}$$

The results give consistent estimates of β and δ. In the third step a modified equation is estimated with probit:

$$T_i = 1 \text{ if } T_i^* > 0$$

$$T_i = 0 \text{ if } T_i^* \leqslant 0$$

$$T_i^* = c(Z_i\hat{\delta}) + cW_i\eta + u_i - v_i$$

where $c = 1/\sigma_{u-v}$. Since the coefficient on $(Z_i\hat{\delta})$ is one over the standard deviation, the parameter vector η can be obtained by dividing it into the probit coefficient vector $(\hat{c\eta})$.

This provides estimates of all the coefficients. The composite variance parameters are also obtainable: σ_{u-v} from the aforementioned "c" coefficient, $\sigma_{\varepsilon+u,u-v}$ and $\sigma_{\varepsilon,u-v}$ thence from the estimates of ψ_1 and ψ_2, and σ^2_ε and $\sigma^2_{\varepsilon+u}$ by a procedure explained in Lee (1979). (Note that the last two estimates are not needed for evaluation of the wage gain and welfare gain.) The underlying variance parameters are then obtainable from these composites (see n. 9).

The full-information technique is to be preferred for many reasons, most of all because it is more efficient than the limited-information technique. Also, the cross-equation restrictions for common parameters are directly imposed in the full-information case. In addition, the OLS standard errors in the limited-information case are inconsistent; correct estimates can only be obtained at some effort. This last point is particularly important for us, for we wish to test the significance of heterogeneity of rewards and costs. A correct significance test cannot be performed from the limited-information estimates.

II. AN EMPIRICAL APPLICATION

A. Data and Empirical Specification

To illustrate the model, we have applied it to the government manpower training program in Sweden. The program provides classroom and other forms of training in a large variety of fields. The purpose of the training is to raise the future earnings of the participants and at the same time provide the expanding sectors of the economy with trained labor. Although unemployment or risk of becoming unemployed is the common eligibility criterion, some of the courses are open to anyone. The data are from the Swedish Level of Living Survey (see Vuksanović, 1979), a longitudinal data base from a representative sample of the Swedish population. The sample size in the survey is about 6,500 persons, interviewed in 1974 and 1981. The data base provides information about personal characteristics and traditional human capital variables such as schooling, experience, and the wage level for those who were employed at the time of the interview. Recently the data base has been supplemented with register data from the National Labor Market Board containing information on those individuals who undertook manpower training provided by the Board. Information is now available on the persons in the sample who started manpower training from 1976 onwards; 470 persons in the total sample started manpower training from 1976 until May 1982.

Our basic model can be formulated both in terms of wage levels and first differences. In order to maintain comparability with recent American studies of manpower training (especially Kiefer, 1979; and Bassi, forthcoming) we have chosen the latter formulation, even though it has the disadvantage of reducing the sample substantially. The sample characteristics are presented in Appendix B. Our outcome variable is the difference of the log of wages between 1981 and 1974.[13]

For our X, Z, and W variables, we have only the standard choices. Among the X variables we have only included truly predetermined variables--experience, schooling, age, and sex. Experience and schooling are measured prior to training to avoid endogeneity problems with training.[14]

The exact same variables are included in Z, for we have no strong arguments for excluding any of them.[15] Our a priori assumption is that the skills provided by the courses are more useful for those with little general schooling and little experience. The ability to learn might also vary with age. Earlier studies have also shown that the gains are higher for women even though the reason is not clear (see Bassi, forthcoming). We will also experiment with health status and a dummy variable for immigrants among the Z variables.

The costs are more difficult to specify since items like preferences for schooling, foregone income, and size of the training stipend are not included in our data. However, it can be argued that age should be included because it determines the length of the horizon. Also, women may have a shorter planning horizon than men because their labor-force activity is more intermittent. We will therefore test age and sex as cost determinants.

B. Results

The maximum-likelihood estimates of the full model are shown in column (1) of Table 1. The first five rows show the coefficients on the variables affecting the reward to participation in the program. Older individuals and those with more education have significantly lower rewards. Neither experience nor sex has a significant effect. Other variables tried in the equation (health status and immigrant status) were also insignificant. The next rows show two cost parameters. Both are quite insignificant, presaging a finding that reoccurs throughout--that costs are weakly significant, if at all. However, note that the point estimate of the costs is negative rather than positive at all relevant ages. Although insignificant, what we have termed "costs" could in fact include some benefits--namely, any benefit not captured by current earnings like more pleasant jobs. Other benefits could include stipends from participation in the program, nonmonetary rewards to participation, and other such items.

The following coefficients show that several standard earnings determinants are significant in our model as well. Increased age and experience have negative effects on the growth rate of earnings, as has been found in most past studies of earnings profiles. Women have higher wage growth, possibly because of various policy measures to decrease discrimination against women. Those with more schooling have lower earnings growth, possibly the result of decreasing returns to education.

Table 1

Estimates of Full Model

	Full-Information Maximum Likelihood (FIML) (1)	(2)	Earnings Equation (OLS)	Three-Step Estimates[a]
Rewards ($Z\delta$):				
Constant	.023 (.441)	.004 (.441)	.724* (.255)	-3.324 (2.840)
Age	-.025* (.013)	-.025* (.013)	.022* (.008)	-.237 (.126)
Experience	-.016 (.012)	-.016 (.012)	.014 (.009)	-.171 (.109)
Female	-.056 (.127)	-.032 (.156)	.256* (.082)	-.879 (.667)
Schooling	-.081* (.030)	-.079* (.030)	-.013 (.017)	-.783 (.457)
Costs ($W\eta$):				
Constant	-.322 (.272)	-.328 (.285)	—	-5.577
Age	.003 (.010)	.002 (.010)	—	.018
Female	—	.033 (.128)	—	—
General Earnings Growth ($X\beta$):				
Constant	1.074* (0.119)	1.074* (0.124)	1.083	4.203
Age	-.004* (.002)	-.004* (.002)	-.004* (.001)	.0025 (.002)

-table continues-

Table 1, continued

	Full-Information Maximum Likelihood (FIML) (1)	(2)	Earnings Equation (OLS)	Three-Step Estimates[a]
Experience	-.005*	-.005*	-.005*	-.001
	(.002)	(.002)	(.001)	(.002)
Female	.038*	.037*	.034*	.065*
	(.016)	(.016)	(.016)	(.018)
Schooling	-.007*	-.007*	-.008*	.010*
	(.004)	(.004)	(.002)	(.005)
Covariance Matrix				
σ_ε	.315*	.315*	—	3.83
	(.005)	(.004)		
σ_u	.986*	.969*	—	10.99
	(.207)	(.218)		
σ_v	.228	.197	—	—[b]
	(.692)	(1.794)		
$\rho_{\varepsilon u}$[c]	-.238	-.220	—	-1.98
	(.524)	(.547)		
$\rho_{\varepsilon v}$[d]	-.673*	-.673	—	—[b]
	(.239)	(.473)		
Composite Variances				
$\sigma_{(\varepsilon+u)}$	.961*	.951*	—	2.00
	(.113)	(.132)		
$\sigma_{(u-v)}$	1.012	.989	—	9.71
	(0.185)	(.187)		
$\rho_{(\varepsilon+u)(u-v)}$	.973	.972	—	5.36
	(.013)	(.019)		

-table continues-

Table 1, continued

	Full-Information Maximum Likelihood (FIML) (1)	(2)	Earnings Equation (OLS)	Three-Step Estimates[a]
$\rho_{\varepsilon(u-v)}$	-.081 (.865)	-.081 (.860)	--	-.46
Log likelihood	-908.28	-908.18	--	--

Asymptotic standard errors in parentheses.

*Significant at 10 percent level.

[a]Standard errors of η vector and covariances not obtainable in this method.

[b]Estimated $\sigma_v^2 = -26.72$.

[c]$\rho_{\varepsilon u} = \sigma_{\varepsilon u}/(\sigma_\varepsilon \sigma_u)$

[d]$\rho_{\varepsilon v} = \sigma_{\varepsilon v}/(\sigma_\varepsilon \sigma_v)$

The estimates of the covariance matrix are in some sense the most important, since the main hypotheses to be tested in the paper are related to them. As the results show, unobserved heterogeneity of rewards (σ_u) is highly significantly determined, while unobserved heterogeneity of costs (σ_v) is insignificant at conventional levels. Thus our preliminary judgment is that heterogeneity of rewards is more important than heterogeneity of costs. The covariance terms show that unobserved earnings growth (ε) is negatively correlated with both heterogeneity of rewards and costs, although insignificantly in the former case. The composite variances show that the error term in the selection equation is weakly negatively correlated with the error term in the nonparticipant earnings equation but strongly positively correlated with the error term in the participant earnings equation.

The estimates in the second column show the result of testing the identification of the cost parameter vector. As will be recalled, these parameters are identified only if some variables in Z are not in W. On the hypothesis that women may have shorter time horizons than men and may have different costs, the female dummy was tested in the cost vector. As the results indicate, its coefficient was very insignificant and a likelihood ratio test cannot reject a zero coefficient. Apparently the results are not sensitive to simple changes in the identification of those costs.

The third and fourth columns show the results of using simpler techniques. Ignoring selection bias altogether and estimating the earnings equation with OLS yields coefficients on the Z variables that are far from their FIML values. The last column of the table shows that estimates from the three-step technique are also quite far from the FIML values. Indeed, not only are the coefficient magnitudes often implausible, but the estimated variances are sometimes negative and the estimated correlation coefficients are sometimes greater than one in absolute value.[16] We have not conducted any systematic examination of the reasons for these results, but they may be related to the rather low trainee participation rates in the sample (about 5 percent). The three-step technique may be particularly unreliable when such a small tail of the distribution is being fitted.[17]

Table 2 shows the results of testing several of the restrictions regarding heterogeneity in which we are interested. The first column replicates the results from column (1) of Table 1. The second column tests the restriction that all cost parameters are zero ($\sigma_v = \rho_{\varepsilon v} = \eta = 0$). A likelihood ratio test indicates that the restriction is rejected at the 90 percent level but cannot be rejected at the 95 percent level. Thus the four cost parameters are, as a whole, barely significant. Note that in this case the wage gain equals the welfare gain, for participation is determined solely by the reward. Thus we can also conclude that the difference between the welfare gain and the wage gain in the full model is only barely significant.

In the next column we test the restriction that there is no unobserved heterogeneity of rewards ($\sigma_u = \rho_{\varepsilon u} = 0$). A likelihood ratio test overwhelmingly rejects this restriction ($\chi^2 = 56$). Note too that this restriction has a large effect on the δ parameters. This means that merely interacting T with other variables will not give correct estimates. Next we further restrict the model by having no observed heterogeneity of rewards--that is, we restrict the model to have only a constant wage effect of participation. The OLS estimates of this model,

Table 2

Estimates of Restricted Models

	Full Specification (FIML)	No Costs[a]	No Unobserved Heterogeneity of Rewards[b]	Complete Homogeneity of Rewards[c]		
				FIML	OLS	Three-Step[d]
Rewards ($Z\delta$)						
Constant	.023	.176	.626*	.096	.045	1.688
	(.441)	(.328)	(.324)	(.286)	(.036)	(.353)
Age	−.025*	−.025*	−.008	—	—	—
	(.013)	(.010)	(.008)			
Experience	−.016	−.012	−.003	—	—	—
	(.012)	(.010)	(.004)			
Female	−.056	−.054	.011	—	—	—
	(.127)	(.117)	(.036)			
Schooling	−.081*	−.069*	−.025	—	—	—
	(.030)	(.024)	(.022)			
Costs ($W\eta$)						
Constant	−.322	—	.528	.246	—	.160
	(.272)		(.358)	(.287)		
Age	.003	—	.002	.001		
	(.010)		(.008)	(.002)	—	.011
General Earnings Growth ($X\beta$)						
Constant	1.074*	1.016*	1.078*	1.090*	1.107	.857
	(.119)	(0.047)	(0.133)	(0.064)		
Age	−.004*	−.003*	−.004*	−.004*	−.004*	−.0001
	(.002)	(.001)	(.002)	(.002)	(.001)	(.002)
Experience	−.005*	−.004*	−.005*	−.005*	−.005*	−.005*
	(.002)	(.002)	(.002)	(.002)	(.001)	(.001)
Female	.038*	.041*	.045*	.044*	.044*	.044*
	(.016)	(.016)	(.016)	(.016)	(.016)	(.016)
Schooling	−.007*	−.006*	−.007*	−.008*	−.008*	−.008*
	(.004)	(.003)	(.004)	(.002)	(.002)	(.002)

(table continues)

Table 2 (cont.)

	Full Specification (FIML)	No Costs[a]	No Unobserved Heterogeneity of Rewards[b]	Complete Homogeneity of Rewards[c] FIML	OLS	Three-Step[d]
Covariance Matrix						
σ_ε	.315* (.005)	.321* (.003)	.323* (.003)	.321* (.003)	--	3.180
σ_u	.986* (.207)	.946* (.101)	--	--	--	--
σ_v	.228 (.692)	--	.338 (.299)	--[e]	--	--
$\rho_{\varepsilon u}$	-.268 (.524)	-.477* (.065)	--	--	--	--
$\rho_{\varepsilon v}$	-.673* (.239)	--	.037 (.872)	.083 (.404)	--	-.238
Composite Variances						
$\sigma(\varepsilon+u)$	.961* (.113)	.841* (.028)	.323* (.003)	.321* (.003)	--	3.180
$\sigma(u-v)$	1.012* (0.185)	.945* (.101)	.338 (.299)	--[e]	--	--
$\rho(\varepsilon+u)(u-v)$	.973* (.013)	.942* (.010)	-.037 (.872)	-.083 (.404)	--	.238
$\rho_\varepsilon(u-v)$	-.081 (.865)	-.477* (.065)	-.037 (.872)	-.083 (.404)	--	.238
Log Likelihood	-908.28	-912.91	-936.03	-947.43	--	--

Asymptotic standard errors in parentheses
*Significance at 10-percent level
[a] $\sigma_v = \rho_{\varepsilon v} = \eta = 0$
[b] $\sigma_u = \rho_{\varepsilon u} = 0$
[c] $\sigma_u = \rho_{\varepsilon u} = 0$, δ = constant. All η coefficients relative to σ_v = .338.
[d] Standard errors on covariances and on η vector not obtainable in this method.
[e] Not estimated; σ_v normalized at .338.

commonly used in past studies, give low (.045) and insignificant wage gains. The FIML estimates, which allow for self-selection via the costs, give higher (.096) but still insignificant wage gains. Note too that the estimate of the correlation across equations in this model is low and insignificant (.08). An analyst who has estimated this model alone might conclude that there is no selection bias (perhaps because a first-difference technique has been employed), but in fact that correlation is an average of positive and negative composite correlations in the full model (.973 and -.081, weighted toward the latter because 95.8 percent of the sample are nonparticipants). Finally, it is interesting to compare the FIML estimates with the results from the widely used three-step method. The differences are again surprisingly large. Our conclusion is that the efficiency gains of the FIML method are important in models like ours.

C. Implications for Wage and Welfare Gains

Using the parameter estimates from our full FIML specification, we can calculate the expected wage and welfare gains from participating as given in equations (14)-(15) above. The results are shown in Table 3. The first entry indicates that a participant with the Z and W characteristics of the mean individual in the total population would have a wage gain of approximately 6.5 percent. The mean wage gain of those who do not participate is negative, as should be expected--this is part of the reason that such individuals do not participate. The mean welfare gain from participation is equal to about a 40 percent increase in the wage. The welfare gain is larger than the wage gain because, in our application, mean costs are estimated to be negative. There is nothing necessary in this result, and we expect that positive costs would occur in other applications and that therefore the welfare gain would be smaller than the wage gain.

The standard error on the wage gain is fairly large, equal to .402. This may seem at odds with our above results in the significance of heterogeneity of rewards, but the two findings are quite compatible, as illustrated in Figure 1. The mean reward in the population is -2.11, but the fraction participating (about 5 percent) occupies only the small upper tail of the reward distribution (the shaded region).[18] The conditional mean in that tail is not far from zero, partly because the conditional variance in the tail is naturally large, and partly because in our application we have estimated negative mean costs, as already mentioned--hence many individuals participate even though they have negative wage gains (although a few have very large wage gains).[19] Nevertheless, the likelihood ratio tests reported above indicate that the wage-gain distribution as a whole is a good explainer of participation and the wage gain from participation. A model which collapsed the distribution on the mean would be significantly worse.

The diagram in Figure 1 also shows how our model can be used to predict the effect on earnings of changing the participant population. For example, lowering costs--such as by paying stipends to a training program or providing scholarships for education--would shift $\bar{W}\eta$ to the left (as shown by the arrow) and enlarge the number of participants. Those brought into the program obviously would have smaller wage gains than those already in--hence the mean wage gain must fall. Mathematically, the effect on the mean wage gain of changing costs is:

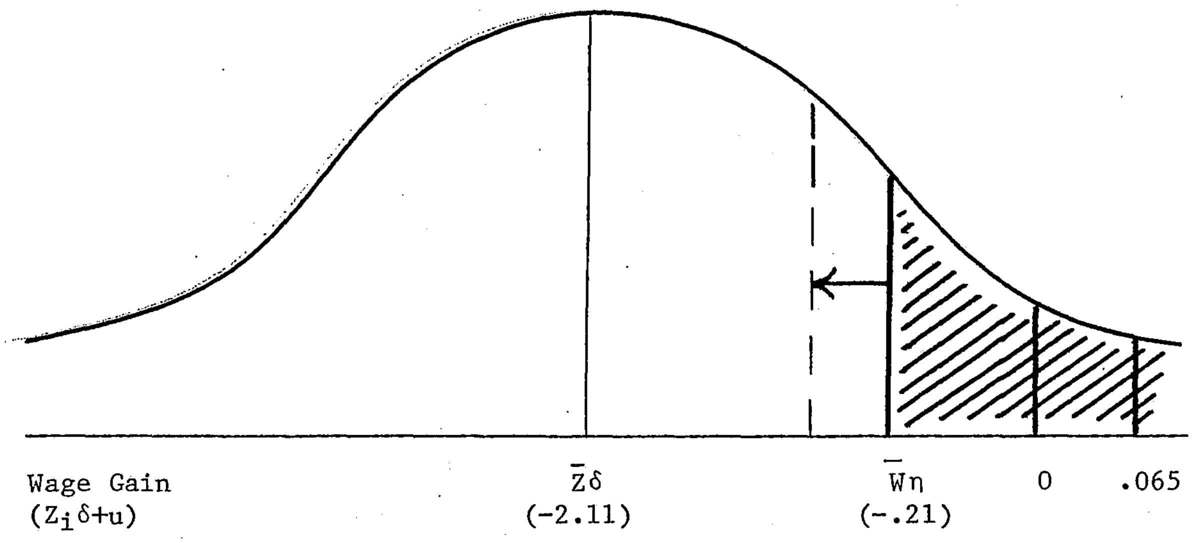

Figure 1

$$\frac{\partial E(Z_i\delta + u_i \mid T_i = 1, Z_i\delta, W_i\eta)}{\partial(W_i\eta)} = \frac{\sigma_u^2 f(s_i)}{\sigma_{u-v}^2[1 - F(s_i)]} \left\{\frac{f(s_i)}{1 - F(s_i)} - s_i\right\} > 0$$

where the notation can be seen in the notes to Table 3. This expression must be positive because the term in curly brackets is positive (it is the expectation of (T_i^*/σ_{u-v}) conditional upon its being positive).

A related question with a somewhat different answer is what the effect on mean wages in the total population would be if costs were lowered and participation expanded, i.e., whether economy-wide productivity would increase. Expected wage growth in the total population is:

$$\begin{aligned} E(\Delta W_i \mid X_i\beta, Z_i\delta, W_i\eta) &= X_i\beta + \text{Prob}(T_i{=}1)E[Z_i\delta + u_i \mid T_i = 1, Z_i\delta, W_i\eta] \\ &= X_i\beta + [1 - F(s_i)][Z_i\delta + (\sigma_u^2/\sigma_{u-v})f(s_i)/(1 - F(s_i))]. \end{aligned}$$

Hence

$$\partial E(\Delta W_i \mid X_i\beta, Z_i\delta, W_i\eta)/\partial(W_i\eta) = -f(s_i)[(Z_i\delta)\sigma_v^2 + (W_i\eta)\sigma_u^2]/\sigma_{u-v}^3.$$

This effect is a weighted average of rewards and costs, and hence is ambiguous in sign. In particular, the sign can differ between groups with different Z and W characteristics. In our sample, since the mean wage gain and mean costs are both negative, the expression is positive--hence lowering costs and increasing participation would lower mean wage growth. This is again because negative costs imply that many participants who are on the margin have negative wage gains. If costs were instead positive, subsidizing them would bring in participants with positive wage gains and hence could improve mean economy-wide wages. The strength of our model is that these effects can be calculated explicitly.

The rest of the results in Table 3 are also of some interest. Evaluating wage and welfare gains at the mean characteristics of the participant population gives higher values of both (almost 100 percent higher in the case of wage gains). As should be expected, those who participate have the Z characteristics for which rewards are higher. Likewise, as also shown in the table, nonparticipants have characteristics for which the wage gains are more negative. A "high-reward" population--the young, the inexperienced, and those with less schooling--have wage gains of almost 20 percent, almost double those of the mean participant. Finally, the table results for the three-step technique confirm the lack of robustness indicated in Tables 1 and 2. The wage and welfare gains implied by the parameters are extreme and implausible.

III. SUMMARY AND CONCLUSIONS

In this paper we have extended the basic self-selection model--appropriate for estimating the effect on earnings of education, training, unions, or migration--incorporate heterogeneity of rewards in returns to the activity. We show that such heterogeneity can be neatly specified in a model that has a close relationship to basic consumer demand theory, and that the welfare gain to the activity can be estimated in the model. We also demonstrate that the notion of heterogeneity of rewards has strong implications for public policy, for it implies that bringing more people into the activity lowers the mean rate of return. One of the strengths of our model is that it makes these points explicit and provides the means to calculate directly the effect of changing the cost

Table 3

Expected Wage and Welfare Gains[a]

	FIML	Three-Step Technique
Total Population $\bar{Z}$, $\bar{W}$:		
Exp. Wage Gain$_{T=1}$	.065 (.402)	5.306 --[b]
Exp. Wage Gain$_{T=0}$	-2.177 (.919)	-24.791 (10.143)
Exp. Welfare Gain$_{T=1}$	.343 (.353)	3.643 (3.297)
Exp. Welfare Gain$_{T=0}$	-1.969 (.934)	-14.809 (9.126)
Participant Population $\bar{Z}$,$\bar{W}$:		
Exp. Wage Gain$_{T=1}$	.103 (.427)	4.821 --[b]
Exp. Welfare Gain$_{T=1}$	.434 (.384)	4.023 (3.583)
Non-Participant Population $\bar{Z}$,$\bar{W}$:		
Exp. Wage Gain$_{T=0}$	-2.189 (0.920)	-24.902 (10.162)
Exp. Welfare Gain$_{T=0}$	-1.981 (0.941)	-19.929 (9.139)
High-Reward Population $\bar{Z}$,$\bar{W}$:[c]		
Exp. Wage Gain$_{T=1}$	.187 (.484)	4.110 --[b]
Exp. Welfare Gain$_{T=1}$	.534 (.453)	4.922 (4.215)

-table continues-

Notes to Table 3

[a]Standard errors in parentheses.

$$\text{Exp. Wage Gain}_{T=1} = E(\bar{Z}_i\delta + u \mid T_i^* > 0,\ \bar{Z}_i\delta,\ \bar{W}_i\eta)$$

$$= \bar{Z}_i\delta + (\sigma_u^2/\sigma_{u-v})\lambda_1,\ \lambda_1 = f(s)/[1 - F(s)]$$

$$s = (\bar{W}_i\eta - \bar{Z}_i\delta)/\sigma_{u-v}$$

$$\text{Variance} = \sigma_u^2[1 + (\sigma_u^2/\sigma_{u-v})(s\lambda_1 - \lambda_1^2)]$$

$$\text{Exp. Wage Gain}_{T=0} = E(\bar{Z}_i\delta + u \mid T_i^* < 0,\ \bar{Z}_i\delta,\ \bar{W}_i\eta)$$

$$= \bar{Z}_i\delta - (\sigma_u^2/\sigma_{u-v})\lambda_2,\ \lambda_2 = f(s)/F(s)$$

$$\text{Variance} = \sigma_u^2[1 - (\sigma_u^2/\sigma_{u-v}^2)(s\lambda_2 + \lambda_2^2)]$$

$$\text{Exp. Welfare Gain}_{T=1} = E(T_i^* \mid T_i^* > 0,\ Z_i\delta,\ \bar{W}_i\eta)$$

$$= \bar{Z}_i\delta - \bar{W}_i\eta + \sigma_{u-v}\lambda_1$$

$$\text{Variance} = \sigma_{u-v}^2(1 + s\lambda_1 - \lambda_1^2)$$

$$\text{Exp. Welfare Gain}_{T=0} = E(T_i^* \mid T_i^* < 0,\ \bar{Z}_i\delta,\ \bar{W}_i\eta)$$

$$= Z_i\delta - W_i\eta - \sigma_{u-v}\lambda_2$$

$$\text{Variance} = \sigma_{u-v}^2(1 - s\lambda_2 - \lambda_2^2)$$

[b]Negative variance.

[c]Age = 20, Experience = 2, Schooling = 9, Female = .44.

structure on participation probabilities and mean rewards of participants. Our empirical application to a Swedish manpower training program provides strong evidence of the existence of heterogeneity of rewards.

There are several areas of additional research on this topic. First, it would be interesting to incorporate uncertainty into the model, for participation decisions are presumably based upon some guess about the future returns--the actual return is not known. Another extension would be the incorporation of involuntary nonparticipation into the model, such as would occur if an individual desires to be a member of a union and cannot get a union job, or if an individual desires to enroll in an educational or training program but cannot.[20] Finally, it would of course be interesting to see this model applied to wage equations for education, unions, migration, and other training programs.

NOTES

[1]Kiefer and Bassi employ a first-difference technique to eliminate selection bias. Their stochastic model is a special case of a larger class of panel data models that assume fixed effects. See Chamberlain (1982) for a discussion of such models. In this paper we will only be concerned with the first difference technique. Note also that the manpower-training application of Ashenfelter (1978) uses panel data. However, selection bias is avoided in that model only if selection is based soley upon lagged earnings, an observable variable. The more serious problem arises when selection is affected by unobservables, the case we are concerned with here. See Barnow, Cain, and Goldberger (1980) for a discussion of this distinction.

[2]This distinction was made in an earlier paper by Moffitt (1981).

[3]For example, this is the model discussed by Barnow, Cain, and Goldberger (1980) and used by Kiefer (1979) and Bassi (forthcoming).

[4]To put it differently: if there is no heterogeneity of costs or preferences, and if the rate of return is constant (and positive), why do not all individuals participate? The common statement that bias arises because participation is correlated with "ability" and because "ability" is in the linear error term ε_i cannot be correct, for ε_i cancels out in the comparison of earnings with and without participation.

[5]Note that the equation system (10)-(13) is observationally equivalent to the general Lee (1979) model in which each regime is allowed to have its own error term with a separate variance, and where free correlation between the choice-equation error term and the two regime error terms is allowed. Compared to that general model, which has sometimes

been estimated in the applied literature, our formulation only provides an alternative interpretation of the various correlations (albeit an economically important one). The importance of the interpretation of the error terms can be seen by comparing our interpretation to the union model of Lee (1978). Such a comparison has been made recently by Björklund (1983), who shows that in the context of our model Lee's parameter estimates have very different implications for the magnitude of union-wage effects than he supposed.

[6]Both hypotheses are nested in the full model. See notes 9 and 10.

[7]To see this, note that the earnings obtainable by participating and by not participating in the activity respectively can be denoted

$$Y_{ti} = Y_{si} + X_i\beta + Z_i\delta + u_i + \varepsilon_i \text{ if } T_i = 1$$

$$Y_{ti} = Y_{si} + X_i\beta + \varepsilon_i \qquad \text{if } T_i = 0$$

where t and s represent time periods after and before the activity.

[8]Note that the variance of the T* equation is identified, unlike that of a probit equation. The reason is important. The variance is identified because the wage gain $Z_i\delta$ appears in the T* equation with a coefficient of one. This is our restriction from theory--that the participation decision must be a direct function of the dollar wage gain. T* is thus measurable in dollar terms and its scale can be fixed. This also relates to the identification condition on the W and Z vectors. In the Lee model, the same condition appears as a requirement that the coefficients on Y in the selection equation be identified--the variance cannot be identified and is normalized to one. In our model, the theoretical restriction we impose on those coefficients allows us to identify

the variance instead. This is what allows us to identify the welfare gain in dollar terms, a crucial contribution of the paper.

[9]The five identifiable composite variances are:

$$\sigma^2_{\varepsilon+u} = \sigma^2_{\varepsilon} + 2\sigma_{\varepsilon u} + \sigma^2_{u}$$

$$\sigma^2_{\varepsilon} = \sigma^2_{\varepsilon}$$

$$\sigma^2_{u-v} = \sigma^2_{u} - 2\sigma_{uv} + \sigma^2_{v}$$

$$\sigma_{\varepsilon+u,u-v} = \sigma_{\varepsilon u} - \sigma_{\varepsilon v} + \sigma^2_{u} - \sigma_{uv}$$

$$\sigma_{\varepsilon,u-v} = \sigma_{\varepsilon u} - \sigma_{\varepsilon v}$$

The need for normalization can be seen by noting that there are six unknown underlying parameters in the five equations.

[10]Lee normalized the covariance across the two earnings equations to be zero. This cannot be the case in our model. Nor would $\sigma_{\varepsilon u} = 0$ be plausible on a priori grounds. Hence we set $\sigma_{uv} = 0$. Note too that our measures of wage and welfare gains as well as tests for heterogeneity of rewards and heterogeneity of costs are invariant to this normalization because these measures and tests only involve the composite variances. Regardless of the normalization, the test of $\sigma_u = \sigma_{uv} = \sigma_{\varepsilon u} = 0$ is a test for whether (1) $\sigma^2_{\varepsilon+u} = \sigma^2_{\varepsilon}$ and (2) $\sigma_{\varepsilon+u,u-v} = \sigma_{\varepsilon,u-v}$. The test of $\sigma_v = \sigma_{uv} = \sigma_{\varepsilon v} = 0$ is a test for whether (1) $\sigma^2_{u-v} = \sigma_{\varepsilon+u,u-v} - \sigma_{\varepsilon,u+v}$ and (2) $\sigma^2_{\varepsilon+u} - \sigma^2_{\varepsilon} = \sigma_{\varepsilon+u,u-v} + \sigma_{\varepsilon,u-v}$.

[11]These probability estimates are to be interpreted loosely, for the probability mass at a single point is zero.

[12]The two equations could also be estimated separately.

[13]The concept of fixed effect in this model is thus in relative terms.

[14]In wage change equations it is common to include the change in experience and schooling on the right-hand side. In our case we find that inappropriate because both variables are endogenous; the choice beween participation and nonparticipation implies a choice between different changes in experience and schooling.

[15]Note that there is no identification requirement on X and Z.

[16]The lambda variables in these equations and those in Table 2 are all significant at the 10 percent level.

[17]Another source of imprecision in the model may lie in our not having any variables in the selection equation that can be reasonably excluded from the earnings equation.

[18]We assume $v = 0$ for illustration.

[19]Such would occur in any case, of course, since v ranges to minus infinity. But clearly a negative mean cost results in more participants with negative wage gains than would be the case if costs were positive.

[20]Such a specification would lead to a disequilibrium model of participation (Moffitt, 1981). The bivariate probit model with partial observability would be applicable (Poirier, 1980).

References

Ashenfelter, O. 1978. "Estimating the Effect of Training Programs on Earnings." Review of Economics and Statistics (February): 47-57.

Barnow, B., G. Cain and A. Goldberger. 1980. "Issues in the Analysis of Selection Bias." In Evaluation Studies Review Annual, edited by E. Stromsdorfer and G. Farkas, Vol. 5. Beverly Hills: Sage.

Bassi, L. "The Effect of CETA on the Post-Program Earnings of Participants." Journal of Human Resources, forthcoming.

Bjorklund, A. 1983. "A Note on the Interpretation of Lee's Self-selection Model," mimeo.

Chamberlain, G. 1982. "Panel Data." Working Paper 8209, Social Systems Research Institute, University of Wisconsin.

Heckman, J. 1978. "Dummy Endogenous Variables in a Simultaneous Equations System." Econometrica, 46 (July): 931-959.

_____. 1979. "Sample Selection Bias as a Specification Error." Econometrica, 47 (January): 153-161.

Kenny, L., L. Lee, G. Maddala. and R. Trost. 1979. "Returns to College Education: An Investigation of Self-Selection Bias Based on Project Talent Data." International Economic Review, 20 (October): 775-789.

Kiefer, N. 1979. "Population Heterogeneity and Inference from Panel Data on the Effects of Vocational Education." Journal of Political Economy (October): S213-S226.

Lee, L. 1978. "Unionism and Wage Rates: A Simultaneous Equations Model With Qualitative and Limited Dependent Variables." International Economic Review, 19: 415-433.

Lee, L. 1979. "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables." Econometrica, 47 (July): 966-977.

Maddala, G. and L. Lee. 1976. "Recursive Models With Qualitative Endogenous Variables." Annals of Economic and Social Measurement, 5 (Fall): 525-544.

Mallar, C., S. Kerachsky and C. Thornton. 1980. "The Short-Term Economic Impact of the Job Corps Program." In E. Stromsdorfer and G. Farkas, op. cit.

Nakasteen, R. and M. Zummei. 1980. "Migration and Income: The Question of Self-Selection." Southern Economic Journal, 46: 840-851.

Moffitt, R. 1981. "Varieties of Selection Bias in Program Evaluations." Rutgers University, mimeo.

Nickell, S. 1982. "The Determinants of Occupational Success in Britain." Review of Economic Studies, XLIX: 43-53.

Poirier, D. 1980. "Partial Observability in Bivariate Probit Models." Journal of Econometrics, 12 (February): 209-217.

Vuksanovic, M. 1979. "Codebook for the Level of Living Survey 1974" (Kodbok for 1974 års levnadsnivåundersokning, in Swedish), Institute for Social Research, Stockholm.

Willis, R. and S. Rosen. 1979. "Education and Self-Selection." Journal of Political Economy (October): S7-S36.

Appendix A

Evaluation of the Likelihood Function

The log likelihood function is:

$$L = \sum_{T=1} \log(P_1) + \sum_{T=0} \log(P_0)$$

where

$$P_1 = \text{Prob}\ (\varepsilon_i + u_i = Y_i - X_i\beta - Z_i\delta,\ u_i - v_i > W_i\eta - Z_i\delta)$$

$$P_0 = \text{Prob}\ (\varepsilon_i = Y_i - X_i\beta,\ u_i - v_i < W_i\eta - Z_i\delta)$$

Letting f be the unit normal density function and F the cumulative normal distribution function, the two probabilities can be factored into a conditional univariate c.d.f. and a marginal univariate p.d.f.:

$$P_1 = [1 - F(r_1)]\ f(z_1)/\sigma_e$$

$$P_0 = F(r_2)\ f(z_3)/\sigma_\varepsilon$$

where

$$e = \varepsilon + u$$

$$f = u - v$$

$$\rho_{\varepsilon f} = \sigma_{ef}/(\sigma_e\sigma_f)$$

$$\rho_{\varepsilon f} = \sigma_{\varepsilon f}/(\sigma_\varepsilon\sigma_f)$$

$$z_1 = (Y_i - X_i\beta - Z_i\delta)/\sigma_e$$

$$z_2 = (W_i\eta - Z_i\delta)/\sigma_f$$

$$z_3 = (W_i\eta - Z_i\beta)/\sigma_\varepsilon$$

Appendix A (cont.)

$$r_1 = (z_2 - \rho_{ef} z_1)/(1 - \rho_{ef}^2)^{1/2}$$

$$r_2 = (z_2 - \rho_{\varepsilon f} z_3)/(1 - \rho_{\varepsilon f}^2)^{1/2}$$

Appendix B

Sample Characteristics[a]

	Participants	NonParticipants
Number in the sample	87	2014
Log wage after training (1981)	3.502	3.561
Change in log wages, 1974 to 1981	.895	.790
Age (1981)	35.45	42.07
Years of schooling before training (1974)	9.55	10.18
Years of work experience before training (1974)	9.18	14.95
Fraction of women	.44	.44

[a]Sample includes only those with wges in both 1974 and 1980, a subset of the full sample.