The MIT Press

The Estimation of Wage Gains and Welfare Gains in Self-Selection Models

Author(s): Anders Björklund and Robert Moffitt

Source: *The Review of Economics and Statistics*, Feb., 1987, Vol. 69, No. 1 (Feb., 1987), pp. 42-49

Published by: The MIT Press

Stable URL: https://www.jstor.org/stable/1937899

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *The Review of Economics and Statistics*

JSTOR

This content downloaded from 193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

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

Anders Björklund and Robert Moffitt*

Abstract—We modify the basic self-selection model for the effects of education, training, unions, and other activities on wages, by including "heterogeneity of rewards" to the activity—i.e., differences across individuals in the rate of return to the activity. We show that such heterogeneity creates a new form of selection bias. We provide tests for its presence and we draw out its implications for the wage and welfare gains to the activity. An empirical application provides strong support for such heterogeneity in one particular training program.

ONE of the largest literatures in empirical labor economics concerns the estimation of the determinants of the effects of various individual choices on wages. For example, the effects on wages of education, unions, migration, and manpower training have been heavily studied. Nowadays the potential self-selection bias that arises from ordinary least squares (OLS) estimation of such relationships is widely recognized, a bias that arises if those participating in the activity are systematically different from those who are not participating. Since the development of selectivity-bias methods by Maddala and Lee (1976), Heckman (1978, 1979), Lee (1979), and Barnow et al. (1980), a fairly standardized methodology has been available for the elimination of this bias. Examples of applications of these techniques are Willis and Rosen (1979) and Kenny et al. (1979) for education, Lee (1978) for unions, Nakosteen and Zimmer (1980) for migration, and Mallar et al. (1980) for manpower training.

In this paper we show the implications of modifying the standard self-selection model to allow for what we call "heterogeneity of rewards." The standard model assumes implicitly that there is a constant coefficient on the participation dummy variable (for union status, educational level, and so on) in the wage equation; we assume instead that the coefficient has an unobserved random component. This specification reflects more accurately the language commonly used to describe selection bias, an example of which is the hypothesis that "those individuals participating in an activity are those who get more out of it." We assert that the standard model is not consistent with this language.

The implications of heterogeneity in rewards—or heterogeneity in the rate of return—are many and interesting. First, we are able to estimate not only the *average* wage gain to the activity of those who are currently participating, but also the *marginal* wage gain of the individuals who are on the margin. The two may be quite different, and it is the latter that is important if one is attempting to expand the activity and draw more people into it. Second, we show that the selection equation should be specified formally as the difference between benefits and costs to participation. We draw an analogy with consumer demand theory to show that, as a result, we can distinguish the *wage* gain from the *welfare* gain to the activity. The two are different if costs are nonzero. To estimate the welfare gain, one must estimate a structural form of the selection equation; estimating an ad-hoc selection equation together with the wage equation is not enough.

In the next section we present our model. Following that we present an application to a data set of manpower trainees. Our empirical results indicate rather dramatically that heterogeneity of rewards is present.

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_i 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. Then the

Received for publication April 29, 1985. Revision accepted for publication May 1, 1986.

* The Industrial Institute for Economic and Social Research and the Institute for Social Research, University of Stockholm; and Brown University, respectively.

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

Copyright ©1987

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

individual participates in the activity if

$$U(Y_i + \alpha_i - \phi_i) > U(Y_i)$$

or

$$\alpha_i > \phi_i$$

where Y_i is now interpreted as earnings in the absence of participation. This simple analogue with demand theory suggests two sources of heterogeneity that might divide the population into participants and nonparticipants: heterogeneity in rewards (α_i) and 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.

The standard model of self-selection that has been used in the research in this area is the following:

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

$$T_i = 1 \quad \text{if } T_i{*} > 0; \quad T_i = 0 \quad \text{otherwise} \tag{2}$$

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

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.

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 nonparticipation. Hence, some dispersion or heterogeneity in preferences or 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.[1] This follows because the specification of the choice equation in terms of our framework should instead be $T_i{*} = \alpha - W_i\eta - v_i$, where $T_i{*}$ is the net reward. Our model allows instead both observed and unobserved heterogeneity in rewards $\alpha_i = Z_i\delta + u_i$, where Z_i is a vector of observed variables, δ is its coefficient vector, and u_i is an error term. This will generate a random coefficients model, as recently also pointed out by Heckman and Robb (1985).

Reformulating the full model in this way gives us the following:

$$Y_i = X_i\beta + \alpha_i T_i + \epsilon_i \tag{4}$$

$$T_i = 1 \text{ if } T_i{*} > 0; \ T_i = 0 \text{ otherwise} \tag{5}$$

$$T_i{*} = \alpha_i - \phi_i \tag{6}$$

$$\alpha_i = Z_i\delta + u_i \tag{7}$$

$$\phi_i = W_i\eta + v_i \tag{8}$$

$$\begin{gathered} E(\epsilon_i) = E(u_i) = E(v_i) = 0 \\ E(v_i^2) = \sigma_v^2, \quad E(u_i^2) = \sigma_u^2, \quad E(\epsilon_i^2) = \sigma_\epsilon^2 \\ E(\epsilon_i u_i) = \sigma_{\epsilon u}, \quad E(\epsilon_i v_i) = \sigma_{\epsilon v}, \quad E(u_i v_i) = \sigma_{uv}. \end{gathered} \tag{9}$$

In reduced form the model comes down to[2]

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

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

$$T_i = 1 \quad \text{if } T_i^* > 0; \qquad T_i = 0 \quad \text{otherwise} \tag{12}$$

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

Here we have let both rewards and costs be functions of vectors of observed variables (Z and W, respectively) and individual error terms. 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 $\sigma_u = \sigma_{\epsilon u} = \sigma_{uv} = 0$ 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 $\sigma_v = \sigma_{\epsilon v} = \sigma_{uv} = 0$ and $\eta = 0$ except for a constant.[3]

Our model allows us to make a number of interesting distinctions. We can distinguish be-

[1] See Rosen (1977) for a good discussion of this issue in the context of human capital theory.

[2] 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 has been demonstrated by Björklund (1983), who has reinterpreted the union model of Lee (1978) in our terms.

[3] Both hypotheses are nested in the full model. See notes 5 and 6.

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

tween the average wage gain and the marginal wage gain; between the effect of increased participation on participants alone and the population as a whole; and between wage gains and welfare gains. To illustrate these distinctions, consider the following expressions calculable from our model:

$$\begin{aligned} E(\alpha_i|T_i = 1, Z_i\delta, W_i\eta) &= Z_i\delta + E(u_i|u_i - v_i > -Z_i\delta + W_i\eta) \\ &= Z_i\delta + (\sigma_{u,u-v}/\sigma_{u-v})\lambda_i \end{aligned} \tag{14}$$

$$\begin{aligned} \partial E(\alpha_i|T_i = 1, Z_i\delta, W_i\eta)/\partial(W_i\eta) &= [(\sigma_{u,u-v})/\sigma^2_{u-v}]\lambda_i(\lambda_i - s_i) > 0 \end{aligned} \tag{15}$$

$$\begin{aligned} E(Y_i|X_i\beta, Z_i\delta, W_i\eta) &= X_i\beta + \text{Prob}(T_i = 1)E[Z_i\delta + u_i| \\ &\quad T_i = 1, Z_i\delta, W_i\eta] \\ &= X_i\beta + [1 - F(s_i)] \\ &\quad \times [Z_i\delta + (\sigma_{u,u-v}/\sigma_{u-v})\lambda_i] \end{aligned} \tag{16}$$

$$\begin{aligned} \partial E(Y_i|X_i\beta, Z_i\delta, W_i\eta)/\partial(W_i\eta) &= -f(s_i)[(Z_i\delta)d + (W_i\eta)\sigma_{u,u-v}]/\sigma^3_{u-v} \end{aligned} \tag{17}$$

$$\begin{aligned} E(T_i^*|T_i = 1, Z_i\delta, W_i\eta) &= Z_i\delta - W_i\eta \\ &\quad + E(u_i - v_i|u_i - v_i > -Z_i\delta + W_i\eta) \\ &= Z_i\delta - W_i\eta + \sigma_{u-v}\lambda_i \end{aligned} \tag{18}$$

where $\lambda_i = f(s_i)/[1 - F(s_i)]$, $s_i = (-Z_i\delta + W_i\eta)/\sigma_{u-v}$, $d = \sigma^2_{u-v} - \sigma_{u,u-v}$ and where f and F are the standard normal density and distribution functions, respectively.

Equation (14) shows the expected wage gain of those who participate, that is, the average wage gain. This is, roughly speaking, what an OLS coefficient provides. Note that $Z_i\delta$, the wage gain in the entire population, could be zero or negative. Indeed, it is likely to be negative in the entire population if only a small fraction of the population participates in the activity. Equation (15) shows the effect of a change in costs on the wage gain, such as the effect on the wage gain of changing tuition (education), union dues (unions), migration costs (migration), or training stipends (training). This expression must be positive, assuming $\sigma_{u,u-v}$ is positive, for increasing costs must eliminate those with low wage gains on the margin; hence, the mean wage gain amongst participants must rise.

Equation (16) shows the expression for the average earnings of the entire population, which is just a weighted average of participants and nonparticipants. Taking the derivative with respect to costs gives equation (17), which shows the effect of changing costs (and hence either expanding or contracting the size of the program) on the mean wage in the population. If costs are lowered, the negative of this expression gives the marginal gain from increased participation on the economy-wide wage, and it can be positive or negative. It will certainly be lower than the average wage gain in (14) because those on the margin who are brought into the program will have the lowest potential wage gains.

Equation (18) shows the expected value of T_i^* for participants, which we term the "welfare gain." It equals the wage gain only if costs are zero. It must be positive for participants. This welfare gain will be lower than the wage gain if costs are positive, but if costs are negative (e.g., if there are stipends or subsidies to the activity), the welfare gain may be greater than the wage gain. In fact, the wage gain may even be negative in the latter case.

Inspection of the reduced form in equations (10)–(13) shows that selection bias could occur in our model solely from the presence of heterogeneity of rewards, for the error term u_i appears both in the selection equation and in the wage equation. Additional selection bias will occur if the error v_i is correlated with the earnings error, ϵ_i; this is the usual form of selection bias in the standard model. Although such correlation may be present in our model, it does not have as "structural" an explanation as heterogeneity of rewards.

Note too that these selectivity biases cannot be eliminated by first-differencing even if v_i is only correlated with a permanent component of ϵ_i which cancels out in the differencing. This follows because u_i will appear even in an earnings-change equation, for it affects earnings growth from the $T_i = 0$ state to the $T_i = 1$ state.

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

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

is clear that the coefficient vectors β and δ are identified, as are their error variances, $(\sigma_\epsilon^2 + 2\sigma_{\epsilon u} + \sigma_u^2)$ 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,[4] 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 to identify all of the underlying variances and covariances.[5] We have chosen $\sigma_{uv} = 0$; none of our coefficient estimates or heterogeneity tests is affected by this normalization.[6] We estimate the model with full-information maximum likelihood rather than with a limited information method because the former is more efficient, because the cross-equation restrictions for common parameters are directly imposed, and because correct standard errors are directly obtained from the inverse of the information matrix. The last advantage is particularly important for us, for we wish to perform hypothesis tests on the variances and covariances in the model to test for the presence of heterogeneity of rewards.[7] The log likelihood function is given in the appendix.

II. An Empirical Application

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. Unfortunately, space constraints do not allow us to discuss the institutional details of this program; the reader interested in such details is encouraged to write to the first author.

The data are from the Swedish Level of Living Survey (see Vuksanovic, 1979), a longitudinal data base from a representative sample of the Swedish population, supplemented by register data from the National Labor Market Board containing information on individuals who undertook manpower training. In order to maintain comparability with recent American studies of manpower training (Kiefer (1979), Bassi (1983)), we formulate our dependent variable in first-differences —we define it as the difference in the log wage from 1974 to 1981. There are 2,101 observations in the sample. The characteristics of the sample are described in Björklund and Moffitt (1983).

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.[8] The

[4] Note that the variance of the T^* equation is identified, unlike that of a standard probit equation, 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. It should be noted that an equivalent model would result if the wage gain were allowed to have a non-unitary coefficient, for in that case the variance of the error term could not be identified. In the probit model the coefficients can only be identified up to a factor of proportionality, and in our case we normalize by setting the coefficient on the wage gain to one rather than setting the variance of the error term to one.

[5] The five identifiable composite variances are

$$\sigma_{\epsilon+u}^2 = \sigma_\epsilon^2 + 2\sigma_{\epsilon u} + \sigma_u^2, \quad \sigma_\epsilon^2 = \sigma_\epsilon^2,$$
$$\sigma_{u-v}^2 = \sigma_u^2 - 2\sigma_{uv} + \sigma_v^2,$$
$$\sigma_{\epsilon+u,\,u-v} = \sigma_{\epsilon u} - \sigma_{\epsilon v} + \sigma_u^2 - \sigma_{uv}, \text{ and}$$
$$\sigma_{\epsilon,\,u-v} = \sigma_{\epsilon u} - \sigma_{\epsilon v}.$$

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

[6] Lee normalized the covariance across the two earnings equations to be zero. We cannot do this in our model because we have no corresponding structural parameter. We could set $\sigma_{\epsilon u} = 0$ or $\sigma_{\epsilon v} = 0$, but $\sigma_{uv} = 0$ seems more plausible. However, the important point is that our measures of wage and welfare gains as well as our tests for heterogeneity of rewards and of costs are invariant to this normalization, because these measures and tests only involve the composite variances (see note 5). For example, our measures in (14)–(18) involve only σ_{u-v}—which is a composite variance—and $\sigma_{u,\,u-v}$, which equals the composite-variance difference $(\sigma_{\epsilon+u,\,u-v} - \sigma_{\epsilon,\,u-v})$ according to note 5. Also, the heterogeneity tests are just tests on composite variances. The test of $\sigma_u = \sigma_{uv} = \sigma_{\epsilon u} = 0$ is a test for whether (1) $\sigma_{\epsilon+u}^2 = \sigma_\epsilon^2$ and (2) $\sigma_{\epsilon+u,\,u-v} = \sigma_{\epsilon,u-v}$; and the test of $\sigma_v = \sigma_{uv} = \sigma_{\epsilon v} = 0$ is a test for whether (1) $\sigma_{u-v}^2 = \sigma_{\epsilon+u,\,u-v} - \sigma_{\epsilon,\,u+v}$ and (2) $\sigma_{\epsilon+u}^2 - \sigma_\epsilon^2 = \sigma_{\epsilon+u,\,u-v} + \sigma_{\epsilon,\,u-v}$.

[7] See our background paper (Björklund and Moffitt, 1983) for estimates using the limited-information method.

[8] We do not include the change in experience and schooling on the right-hand side because both variables are endogenous;

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

exact same variables are included in Z, for we have no strong arguments for excluding any of them. 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, 1983). 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 training, 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.

Results

The maximum-likelihood estimates of the full model are shown in column (1) of table 1. Age and schooling have significantly negative effects on rewards but experience and sex do not. Costs have no significant determinants; even the constant is insignificant. Earnings growth is negatively affected by age, experience, and schooling (recall that the dependent variable is the growth rate of earnings) and positively affected by being female. The variance estimates show that unobserved heterogeneity of rewards (σ_u) is highly significantly determined but 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.

Column (2) shows the effect of adding sex to the cost vector. Its coefficient is insignificant by a t-test and by a likelihood ratio test. Thus the results are not sensitive to simple changes in the identification of costs.

Columns (3)–(5) show the results of testing several of the restrictions regarding heterogeneity. Column (3) tests the restriction that all cost parameters are zero ($\sigma_v = \rho_{\epsilon v} = \eta = 0$). A likelihood ratio test indicates that the restriction is rejected at the 90% level but cannot be rejected at the 95% level.[9] Thus the four cost parameters are, as a whole, barely significant. We should note that this finding may be a partial result of poor cost proxies. In any case, this implies that the wage gain equals the welfare gain; 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 ($\sigma_u = \rho_{\epsilon u} = 0$). 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 implies 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 aside from a constant—this is the standard selectivity-bias model. This restriction is also rejected by a likelihood ratio test. The effect of the program on earnings in this case is 0.096 and is insignificant. The cross-equation error correlation is 0.08 and is also insignificant. Thus an analyst who had estimated the standard model might conclude that there is no selection bias, whereas in fact the low composite correlation is an average of positive and negative correlations in the full model (0.973 and −0.081, weighted toward the latter because 95% of the sample are nonparticipants).

The last two columns show OLS estimates of the wage equation with and without interactions with the X variables. The last column, showing homogeneous rewards, shows an insignificant effect of 0.045, more than a 50% bias from the 0.096 estimate in the standard model. Note too that the interaction coefficients in column (6) are often

the choice between participation and nonparticipation implies a choice between different changes in experience and schooling.

[9] We must note that we have not adjusted for testing parameter values on the boundary of the parameter space. See Cox and Hinckley (1974, p. 332).

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

TABLE 1.—ESTIMATES OF THE MODEL

	Full Model (1)	Full Model (2)	No Costs[a] (3)	No Unobserved Reward Heterogeneity[b] (4)	Complete Reward Homogeneity[c] (5)	OLS (6)	OLS (7)
Rewards ($Z\delta$)							
Constant	0.023	0.004	0.176	0.626[d]	0.096	0.724[d]	0.045
	(0.441)	(0.441)	(0.328)	(0.324)	(0.286)	(0.255)	(0.036)
Age	−0.025[d]	−0.025[d]	−0.025[d]	−0.008	—	−0.022[d]	—
	(0.013)	(0.013)	(0.010)	(0.008)		(0.008)	
Experience	−0.016	−0.016	−0.012	−0.003	—	0.014	—
	(0.012)	(0.012)	(0.010)	(0.004)		(0.009)	
Female	−0.056	−0.032	−0.054	0.011	—	0.256[d]	—
	(0.127)	(0.156)	(0.117)	(0.036)		(0.092)	
Schooling	−0.081[d]	−0.079[d]	−0.069[d]	−0.025	—	−0.013	—
	(0.030)	(0.030)	(0.024)	(0.022)		(0.017)	
Costs ($W\eta$)							
Constant	−0.322	−0.328	—	0.528	0.246	—	—
	(0.272)	(0.285)		(0.358)	(0.287)		
Age	0.003	0.002	—	0.002	0.001	—	—
	(0.010)	(0.010)		(0.008)	(0.002)		
Female	—	0.033	—	—	—	—	—
		(0.128)					
General Earnings Growth ($X\beta$)							
Constant	1.074[d]	1.074[d]	1.016[d]	1.078[d]	1.090[d]	1.083	1.107
	(0.119)	(0.124)	(0.047)	(0.133)	(0.064)		
Age	−0.004[d]	−0.004[d]	−0.003[d]	−0.004[d]	−0.004[d]	−0.004[d]	−0.004[d]
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Experience	−0.005[d]	−0.005[d]	−0.004[d]	−0.005[d]	−0.005[d]	−0.005[d]	−0.005[d]
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Female	0.038[d]	0.037[d]	0.041[d]	0.045[d]	0.044[d]	0.034[d]	0.044[d]
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Schooling	−0.007[d]	−0.007[d]	−0.006[d]	−0.007[d]	−0.008[d]	−0.008[d]	−0.008[d]
	(0.004)	(0.004)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Covariance Matrix							
σ_ϵ	0.315[d]	0.315[d]	0.321[d]	0.323[d]	0.321[d]	—	—
	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)		
σ_u	0.986[d]	0.969[d]	0.946[d]	—	—	—	—
	(0.207)	(0.218)	(0.101)				
σ_v	0.228	0.197	—	0.338	—[e]	—	—
	(0.692)	(1.794)		(0.299)			
$\rho_{\epsilon u}$[f]	−0.268	−0.220	−0.477[d]	—	—	—	—
	(0.524)	(0.547)	(0.065)				
$\rho_{\epsilon v}$[g]	−0.673[d]	−0.673	—	0.037	0.083	—	—
	(0.239)	(0.473)		(0.872)	(0.404)		
Composite Variances							
$\sigma_{(\epsilon+u)}$	0.961[d]	0.951[d]	0.841[d]	0.323[d]	0.321[d]	—	—
	(0.113)	(0.132)	(0.028)	(0.003)	(0.003)		
$\sigma_{(u-v)}$	1.012[d]	0.989	0.945[d]	0.338	—[e]	—	—
	(0.185)	(0.187)	(0.101)	(0.299)			
$\rho_{(\epsilon+u)(u-v)}$	0.973[d]	0.972	0.942[d]	−0.037	−0.083	—	—
	(0.013)	(0.019)	(0.010)	(0.872)	(0.404)		
$\rho_{\epsilon(u-v)}$	−0.081	−0.081	−0.477[d]	−0.037	−0.083	—	—
	(0.865)	(0.860)	(0.065)	(0.872)	(0.404)		
Log Likelihood	−908.28	−908.18	−912.91	−936.03	−947.43	—	—

Note: Asymptotic standard errors are in parentheses.
[a] $\sigma_v = \rho_{\epsilon v} = \eta = 0$.
[b] $\sigma_u = \rho_{\epsilon u} = 0$.
[c] $\sigma_u = \rho_{\epsilon u} = 0$, δ = constant. All η coefficients relative to $\sigma_v = 0.338$.
[d] Significant at the 10% level.
[e] Not estimated; σ_v normalized at 0.338.
[f] $\rho_{\epsilon u} = \sigma_{\epsilon u}/(\sigma_\epsilon \sigma_u)$.
[g] $\rho_{\epsilon v} = \sigma_{\epsilon v}/(\sigma_\epsilon \sigma_v)$.

This content downloaded from 193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

quite different than those in column (4), showing again the magnitude of the bias in the OLS estimates.

Implications for Wage and Welfare Gains

Table 2 shows our calculations of expected wage and welfare gains given in (14) and (18) above, using the parameter estimates from our full model. The table shows that the expected wage gain among the participants in a population with mean observed characteristics is 6.5% and that the expected wage gain among nonparticipants is negative (which is why they do not participate). The mean welfare gain is 34.3% of the wage, considerably larger than the wage gain. Mean costs are thus estimated to be negative, which may be a result of the stipends paid to the trainees, the placement advantages of the program, or some nonpecuniary gain. The rest of the table shows gains evaluated at different means of the observed variables. The results indicate that, as should be expected, gains are larger for individuals with participant characteristics (10.3% and 43.4%), smaller for individuals with nonparticipant means, and larger for those with high-reward characteristics.

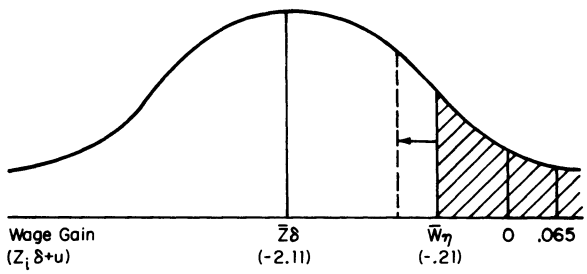
What would be the effect of expanding this program on mean wages (i.e., on productivity growth) in Sweden? Evaluating equation (17) at the means indicates that mean wages would *fall* if the program were expanded by lowering costs (e.g., raising stipends). This is because, at the margin, those who would be brought into the program have negative wage gains according to our estimates. This is illustrated in figure 1, which

TABLE 2.—EXPECTED WAGE AND WELFARE GAINS

Total Population $\bar{Z}, \bar{W}$:	
Expected Wage Gain$_{T=1}$	0.065
Expected Wage Gain$_{T=0}$	−2.177
Expected Welfare Gain$_{T=1}$	0.343
Expected Welfare Gain$_{T=0}$	−1.969
Participant Population $\bar{Z}, \bar{W}$:	
Expected Wage Gain$_{T=1}$	0.103
Expected Welfare Gain$_{T=1}$	0.434
Non-Participant Population $\bar{Z}, \bar{W}$:	
Expected Wage Gain$_{T=0}$	−2.189
Expected Welfare Gain$_{T=0}$	−1.981
High-Reward Population $\bar{Z}, \bar{W}$:[a]	
Expected Wage Gain$_{T=1}$	0.187
Expected Welfare Gain$_{T=1}$	0.534

[a] Age = 20, Experience = 2, Schooling = 9, Female = 0.44.

FIGURE 1.—DISTRIBUTION OF WAGE GAINS IN THE POPULATION



shows the distribution of rewards in the population. The mean reward in the total population is −2.11; those with rewards above the mean cost, −0.21, occupy the upper tail of the distribution (the shaded region). The conditional mean reward in that tail is 0.065, as table 2 indicates.[10] Since mean costs are negative, the marginal individual has a negative wage gain—hence expanding the program by shifting $W\eta$ to the left would not only lower the average gain but would lower mean wages in Sweden. On the basis of this evidence—it is, of course, only one study—one could argue for a reduction in the size of the program.

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—to 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 also be estimated in the model. We also demonstrate that the notion of heterogeneity of rewards has 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 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.

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

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms

APPENDIX

Evaluation of the Likelihood Function

The log likelihood function is

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

where

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

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

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.:

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

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

where

$$e = \epsilon + u, f = u - v, \rho_{ef} = \sigma_{ef}/(\sigma_e\sigma_f), \rho_{\epsilon f} = \sigma_{\epsilon f}/(\sigma_\epsilon\sigma_f),$$

$$z_1 = (Y_i - X_i\beta - Z_i\,\delta)/\sigma_e, z_2 = (W_i\eta - Z_i\,\delta)/\sigma_f,$$

$$z_3 = (Y_i - X_i\beta)/\sigma_\epsilon,$$

$$r_1 = (z_2 - \rho_{ef}z_1)/\left(1 - \rho_{ef}^2\right)^{1/2}, \quad \text{and}$$

$$r_2 = (z_2 - \rho_{\epsilon f}z_3)/\left(1 - \rho_{\epsilon f}^2\right)^{1/2}.$$

REFERENCES

Barnow, Burt, Glen Cain, and Arthur Goldberger, "Issues in the Analysis of Selection Bias," in Ernst Stromsdorfer and George Farkas (eds.), *Evaluation Studies Review Annual*, Vol. 5 (Beverly Hills: Sage, 1980).

Bassi, Laurie, "The Effect of CETA on the Post-Program Earnings of Participants," *Journal of Human Resources* 18 (Fall 1983), 539–556.

Björklund, Anders, "A Note on the Interpretation of Lee's Self-Selection Model," mimeo, The Industrial Institute for Economic and Social Research, Stockholm, 1983.

Björklund, Anders, and Robert Moffitt, "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," Discussion Paper 735-83, Institute for Research on Poverty, University of Wisconsin, 1983.

Cox, D. R., and C. V. Hinkley, *Theoretical Statistics* (London: Chapman and Hall, 1974).

Heckman, James, "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica* 46 (July 1978), 931–959.

______, "Sample Selection Bias as a Specification Error," *Econometrica* 47 (Jan. 1979), 153–161.

Heckman, James, and Richard Robb, "Alternative Methods for Evaluating the Impact of Interventions," in James Heckman and Burton Singer (eds.), *Longitudinal Analysis of Labor Market Data* (Cambridge: Cambridge University Press, 1985).

Kenny, Laurence, Lung-fei Lee, G. S. Maddala, and Robert Trost, "Returns to College Education: An Investigation of Self-Selection Bias Based on Project Talent Data," *International Economic Review* 20 (Oct. 1979), 775–789.

Kiefer, Nicholas, "Population Heterogeneity and Inference from Panel Data on the Effects of Vocational Education," *Journal of Political Economy* 87 (Oct. 1979), S213–S226.

Lee, Lung-fei, "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review* 19 (June 1978), 415–433.

______, "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica* 47 (July 1979), 966–977.

Maddala, G. S., and Lung-fei Lee, "Recursive Models with Qualitative Endogenous Variables," *Annals of Economic and Social Measurement* 5 (Fall 1976), 525–544.

Mallar, Charles, Stuart Kerachsky, and Craig Thornton, "The Short-term Economic Impact of the Job Corps Program," in Ernst Stromsdorfer and George Farkas (eds.), *Evaluation Studies Review Annual*, Vol. 5 (Beverly Hills: Sage, 1980).

Nakosteen, Robert, and Michael Zimmer, "Migration and Income: The Question of Self-Selection," *Southern Economic Journal* 46 (Jan. 1980), 840–851.

Rosen, Sherwin, "Human Capital: A Survey of Empirical Research," in Ronald Ehrenberg (ed.), *Research in Labor Economics* Vol. 1 (Greenwich, CT: JAI Press, 1977).

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

Willis, Robert, and Sherwin Rosen, "Education and Self-Selection," *Journal of Political Economy* 87 (Oct. 1979), S7–S36.

This content downloaded from
193.50.45.171 on Fri, 29 Nov 2024 19:18:27 UTC
All use subject to https://about.jstor.org/terms