371-76

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

THE IMPACT OF THE CENSORING PROBLEM ON ESTIMATING WOMEN'S OCCUPATIONAL ATTAINMENT EQUATIONS

Neil Fligstein
Wendy Wolf

UNIVERSITY OF WISCONSIN ~ MADISON

THE IMPACT OF THE CENSORING PROBLEM ON ESTIMATING WOMEN'S OCCUPATIONAL ATTAINMENT EQUATIONS

by

Neil Fligstein and Wendy Wolf*

*Authorship of this paper is shared equally by Wolf and Fligstein. We gratefully acknowledge the helpful comments and guidance of Martin Dooley, Joseph Hotz, Dave Crawford, William Sewell, Robert Hauser, Rachel Rosenfeld, and Willian Bielby. The authors and this research were supported by a grant to William Sewell and Robert Hauser entitled, "Social and Psychological Factors in Status Attainment" from the National Institute of Mental Health (M-06275), a grant from National Institute of General Medical Sciences (5T01GM01526-10) and a Center for Population Research Grant (HD 05876) to the Center for Demography and Ecology from the Center for Population Research of the National Institute of Child Health and Human Development.

ABSTRACT

Research on sex differences in occupational attainment suggests that working men and working women attain essentially the same mean level of occupational attainment and do so through quite similar processes. A possible explanation for these similarities is that the sample of working women contains an overrepresentation of successful women, since women who can afford not to work will stay out of the labor force unless they find a job commensurate with their education. This we define as a censoring problem. By extending a technique developed by Heckman, we can estimate the structural parameters for all women, regardless of current employment status. This procedure allows us to assess the impact of the censoring problem on women's occupational attainment equations.

Recently, there have been several studies that have compared the occupational attainments of men and women (Treiman and Terrell, 1975; Featherman and Hauser, 1976; McClendon, 1976; Alexander and Eckland, 1975). The results of this research suggest that working women and men attain essentially the same mean level of occupational status and do so through quite similar processes, implying sexual equality in occupational rewards. Because the samples of women for these analyses are restricted to working women, it has been suggested that these findings may be explained in part by the fact that there has been selection into the sample on the basis of the dependent variable (Featherman and Hauser, 1976; Wolf, 1975; 1976; McClendon, 1976). That is, women who can not find a job commensurate with their education and who can afford not to work will opt to remain out of the labor force. If this were the case, the sample of employed women could include an overrepresentation of those who have found jobs that are commensurate with their training and background.

If one is interested in obtaining population parameters describing the process of occupational attainment for all women, restricting the sample to employed women could result in a bias in the structural parameters. If potential occupational status affects a woman's decision to work, the sample of employed women is a nonrandom sample of the population of all women. This can be viewed as a censoring problem (Heckman, 1975). In this paper, we (1) review and reject some alternatives that could correct for this problem and (2) present a technique for obtaining the structural parameters for the whole population by accounting for the censoring problem.

One way to deal with the censoring problem is to include women who are not employed into an equation predicting occupational status. There are three alternatives for doing this: (1) assign them their husband's score; (2) assign them a score for the role of housewife (Bose, 1973); or (3) assign them a zero on an occupational status scale. Assigning women who are not employed a housewife status score or their husbands' status score is inappropriate because it confuses the concept of status obtained through the woman's own labor market activities with status obtained by other means. Allowing over half the women in the sample to have status scores which do not relate to their own labor market activity is not only arbitrary but the interpretation of any regressions based on such assignment is, at best, dubious. Applying a score of zero to women not employed at the relevant times poses difficulties for at least two reasons. Occupational status scales (in this case Duncan's (1961) Socioeconomic Index) are rank orderings of occupations and therefore do not have meaningful zero points. Such an expedient assignment would be arbitrary, if not meaningless. Second, if women who are employed were assigned a score which ranges from zero to ninety-six and those not employed were scored zero, other variables in the regression equation would be highly related to the dependent variable merely because they have a strong effect on whether the woman is working; thus, labor force participation and occupational attainment would be confounded. In summation, it seems difficult to include women who are not gainfully employed into equations predicting occupational status, particularly if one's interest is women's occupational attainments through their own activities in the labor market.

In attempting to take account of all women regardless of their current employment status, one could assert that there is a structural equation that describes the process of occupational attainment for all women. If a woman were to enter the labor force tomorrow, there is a set of structural parameters that describes the returns she would receive for her education and the effects of her other characteristics. The problem in estimating this structural equation is that we can not observe the occupational statuses of women who are not employed. Estimating the structural equation solely for employed women biases the structural parameters because of the censoring problem. By reformulating a technique originally suggested by Heckman (1974; 1975), we are able to estimate the structural parameters by correcting for the bias introduced through the restriction of the sample to employed women. We demonstrate that the error in an occupational status equation estimated solely for working women has a non-zero expectation and is correlated with the exogenous variables in that equation, thus biasing the structural parameters. This correlation is due to the fact that the error in the occupational status equation is related to the decision to work. Using Heckman's technique, we are able to obtain the structural parameters for all women and test whether the censoring problem is empirically important.

Heckman's Model and Our Modification

Heckman (1974) produces a set of equations that relate a woman's decision to work, how many hours she works, her wage rate, and her asking wage rate. To do this he sets up a structural model of the following form:

$$\ell(W_i^*) = \beta_0 + \beta_1 h_i + \beta_2 (W_m)_i + \beta_3 P_i + \beta_4 A_i + \beta_5 Z_i + \varepsilon_i \qquad (1)$$

$$\ell(W_i) = b_0 + b_1 S_i + b_2 E_i + u_i \qquad (2)$$

where $\ell(W_i^*)$ is an appropriately transformed "shadow price" of the woman's time in the home; $\ell(W_i)$ is an appropriately transformed wage; h_i is hours worked; W_m is the wage of the husband; P_i is a vector of good prices; A_i is the asset income of the household; Z_i is a vector of constraints which arise from previous economic decisions and chance; S_i is years of schooling; E_i is work experience; β's and b's are parameters and ε_i and u_i are disturbances that are normally distributed with zero expectation and non-zero variances and covariance.

$$\begin{pmatrix} \varepsilon \\ \\ u \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\varepsilon} & \sigma_{\varepsilon u} \\ \\ \sigma_{\varepsilon u} & \sigma_{u} \end{pmatrix} \right)$$

The problem in estimating the structural equations is that shadow prices can not be observed and wages can only be observed for working women.

Heckman's technique takes account of the censoring problem and is thus able to estimate the structural parameters in Equations (1) and (2). If a woman's shadow wage exceeds her offered wage at zero hours of work, she does not work. If her offered wage exceeds her shadow price at zero hours of work, then she will work, i.e.:

$$\ell(W_i) > \ell(W_i^*) \text{ at } h = 0.$$

$$b_o + b_1 S_i + b_2 E_i + u_i > \beta_0 + \beta_2 (W_m)_i + \beta_3 P_i + \beta_4 A_i + \beta_5 Z_i + \varepsilon_i, \text{ then}$$

$$b_0 - \beta_0 + b_1 S_i + b_2 E_i - \beta_2 (W_m)_i - \beta_3 P_i - \beta_4 A_i - \beta_5 Z_i > \varepsilon_i - u_i \qquad (3)$$

Economic theory predicts that above zero hours of work, a woman adjusts her hours (most possibly her annual hours) so that $W_i = W_i^*$. If the inequality in Equation 3 holds (i.e. the woman works), two reduced form equations can be estimated: one determining observed hours worked and one determining observed wages:

$$h_i = \frac{1}{\beta_1}(b_o - \beta_o + b_1S_i + b_2E_i - \beta_2(W_m)_i - \beta_3P_i - \beta_4A_i - \beta_5Z_i) + \frac{u_i - e_i}{\beta_1} \quad (4)$$

$$\ell(W_i) = b_0 + b_1S_i + b_2E_i + u_i \quad (5)$$

The basic insight in Heckman's work (1975) is that for the subsample of working women the inequality in Equation (3) ($W_i > W_i^*$ at $h = 0$) implies that the conditional means for $\frac{u_i - e_i}{\beta_1}$ and u_i are non-zero and are systematically related to the exogenous variables in their respective equations. To demonstrate this, he derives the expectations of the error terms in Equations (4) and (5), conditional upon the woman working:

$$E\left(\frac{u_i - \varepsilon_i}{\beta_1} \,\middle|\, \frac{u_i - e_i}{\sigma^*} > \phi_i\right) = \frac{\lambda_i\sigma^*}{\beta_1} \quad (6)$$

$$E\left(u_i \,\middle|\, \frac{u_i - e_i}{\sigma^*} > \phi_i\right) = \left(\frac{\sigma_u - \sigma_{\varepsilon u}}{\sigma^*}\right)\lambda_i \quad (7)$$

where $\sigma^* = (\sigma_u - 2\sigma_{\varepsilon u} + \sigma_\varepsilon)^{1/2}$

$$\phi_i = \frac{-(b_0 - \beta_0 + b_1S_i + b_2E_i - \beta_2(W_m)_i - \beta_3P_i - \beta_4A_i - \beta_5Z_i)}{\sigma^*}$$

$$\lambda_i = \frac{\frac{1}{\sqrt{2\pi}}\, e^{-\phi_i^{1/2}}}{\int_{\phi_i}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-t^{1/2}}dt}$$

(λ_i, the inverse of the Mill's ratio, which is the ratio of the ordinate of a standard normal to the right tail.)

Thus the expectation of the error terms in the equations for observed hours and wages are non-zero and are correlated with the exogenous variables in each respective equation. The correlation is due to the fact that

λ_i, a component of the error term in Equations (4) and (5), is a function of ϕ_i, which is a linear combination of some of the exogenous variables in Equations (4) and (5). By including λ_i as an additional regressor and thus correcting for the censoring problem, the structural parameters in Equation (5) can be estimated. Heckman argues: "This representation demonstrates that empirical studies which neglect the censoring problem and apply ordinary least squares to subsamples of working women simply omit λ_i as an explanatory variable" (Heckman, 1975:5). We use this technique to estimate the structural equation describing the occupational attainment process for all women.

Heckman's model is reformulated in a manner that (1) implicitly contains some, but not all, of his assumptions, (2) utilizes his technique, and (3) is more tailored to our concern with occupational attainment. We are interested in two equations: one that estimates whether a woman is employed and the structural equation predicting her occupational attainment. To expedite the following argument, the presentation of the exogenous variables in each equation is delayed.

$$p(EMP_i) = F(X_i'\underline{\gamma}) \tag{8}$$

$$SEI_i = Y_i'\underline{\delta} + \varepsilon_{2i} \tag{9}$$

where $p(EMP_i)$ is the probability that the i^{th} woman works, $X_i'\underline{\gamma}$ is a set of explanatory variables and parameters from a probit analysis, SEI_i is occupational status (Duncan, 1961), $Y_i'\underline{\delta}$ is a set of explanatory variables and parameters in the structural equation, and ε_{2i} is an error term in that equation.

Considering Equation 8:

$$\hat{\phi}_i = I(X)_i = X_i' \underline{\hat{\gamma}} \tag{10}$$

where $I(X)_i$ is the predicted value from a probit analysis for the i^{th} woman. We assert that there exists (see Crawford, 1975a; 1975b):

$$\varepsilon_{1i} = I_i^* \sim N(0,1) \tag{11}$$

where ε_{1i} is the threshold level of work for the i^{th} woman. It is a function of unmeasured variables such as tastes for work, ability, and labor market factors.

It follows (see Crawford, 1975a; 1975b) that:

$$\text{EMP} = 1 \text{ (i.e. the woman is employed) if } I(X)_i \geq I_i^* \text{ or } \phi_i \geq \varepsilon_{1i} \tag{12}$$

$$\text{EMP} = 0 \text{ (i.e. the woman is not employed) if } I(X)_i < I_i^* \text{ or } \phi_i < \varepsilon_{1i} \tag{13}$$

$$\Pr(\text{EMP}_i = 1|X) = F(\phi_i) \tag{14}$$

where $F(\cdot)$ is the cumulative standard normal density function evaluated at ϕ_i.

Now given that ε_{1i} is the stochastic element in the decision to work and ε_{2i} is the disturbance in Equation (9), it is reasonable to assert:

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right).$$

Standardizing ε_{2i}, this becomes $\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i}/\sqrt{\sigma_{22}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12}/\sqrt{\sigma_{22}} \\ \sigma_{12}/\sqrt{\sigma_{22}} & 1 \end{pmatrix} \right).$

We are interested in deriving the conditional expectation of ε_{2i}, given that the woman works ($\phi_i \geq \varepsilon_{1i}$).

From Johnson and Kotz (1970: 82-83), it is known that if $X \sim N(\mu, \sigma^2)$ then:

$$E(X|\phi_i > X) = \mu - \sigma^2\lambda_i . \qquad (15)$$

Thus,

$$E(\varepsilon_{1i}|\phi_i > \varepsilon_{1i}) = -\lambda_i = \frac{f(\phi_i)}{F(\phi_i)} \ \underline{1/} \qquad (16)$$

since $\mu = 0$ and $\sigma = 1$. λ_i is the aforementioned inverse of the Mill's ratio.

The conditional expectation of ε_{2i} given that a woman is employed can be derived as follows (Johnson and Kotz, 1972: 113):

$$E(\varepsilon_{2i}/\sqrt{\sigma_{22}} \mid \phi_i > \varepsilon_{1i}) = \rho\, E(\varepsilon_{1i} \mid \phi_i > \varepsilon_{1i}) = -\rho\lambda_i \text{ where } \rho = \frac{\sigma_{12}}{\sqrt{\sigma_{22}}} . \qquad (17)$$

Unstandardizing, we get:

$$E(\varepsilon_{2i} \mid \phi_i > \varepsilon_{1i}) = -\rho\sqrt{\sigma_{22}}\,\lambda_i = -\sigma_{12}\lambda_i . \qquad (18)$$

When the error term in the SEI equation is treated as conditional on employment it has a non-zero expectation and is correlated with the exogenous variables in that equation. Since some of the exogenous variables are the same in Equations (8) and (9), the correlation is due to the fact that λ_i is a function of ϕ_i, which is a linear combination of the exogenous variables in the equation predicting employment. The quantity λ_i is the inverse of the Mill's ratio--the ratio of the ordinate of a standard normal to the right tail (Heckman, 1975: 3). Its denominator is the probability that a woman works. As $\phi_i \to +\infty$, $\lambda_i \to 0$. In populations where the probability of working is near 1 and therefore λ_i is near 0, the bias is minimal since the conditional means of the errors are near zero (Heckman, 1975: 4).

Equation (19) is the expectation function predicting occupational status conditional on the woman working:

$$E(SEI \mid \phi_i > \varepsilon_{1i}) = Y_i'\underline{\delta} + -\lambda_i\sigma_{12} \tag{19}$$

$$SEI_i = E(SEI \mid \phi_i > \varepsilon_{1i}) + V_{2i} \text{ where } V_{2i} = \sigma_{12}\lambda_i + \varepsilon_{2i} \text{ and } E(V_{2_i}) = 0 \tag{20}$$

Equation (20) is the equation we estimate solely for working women that allows us to obtain the structural parameters for all women. Since $V_{2i} = \sigma_{12}\lambda_i + \varepsilon_{2i}$, Equation (20) is merely another way of rewriting Equation (9). Substituting from Equations (19) and (20):

$$SEI_i = Y_i'\underline{\delta} - \sigma_{12}\lambda_i + \sigma_{12}\lambda_i + \varepsilon_{2i} = Y_i'\underline{\delta} + \varepsilon_{2i} \tag{21}$$

By including λ_i as a regressor, we have derived an equation to be estimated for employed women that controls for the potential bias due to the censoring problem. The structural parameters for Equation (9) are obtained from Equation (20), which includes λ_i. The coefficients in Equation (20) are the structural parameters; one does not treat the coefficient of λ_i as one of the structural parameters. The parameter estimated for λ_i in Equation (20) is an estimator of σ_{12}: the covariance between the errors in the equation predicting employment and the errors in the equation predicting occupational status.

Equation (20) is estimated in a fashion suggested by Heckman (1975). Equation (8) is estimated with a probit analysis, thus obtaining ϕ_i's for all persons in the sample. Then λ_i's are obtained for all employed women in the sample by using ϕ_i and generating the inverse of the Mill's ratio for each individual. Finally, using ordinary least squares, Equation (20) is estimated, which includes λ_i as an additional regressor.

Equations to be Estimated

At this point the equations that are estimated are presented. Equation (22) is the model that is estimated through a probit analysis of the total sample. Equation (23) is the equation with occupational status in 1967 as the dependent variable, corrected for the censoring problem. It is estimated for the sample of employed women using ordinary least squares.

$$p(EMP_i) = F(\gamma_0 + \gamma_1 AGE_i + \gamma_2 KIDLT6_i + \gamma_3 KID613_i + \gamma_4 KID1417_i + \gamma_5 ED_i + \gamma_6 AAM_i + \gamma_7 EVERTR_i + \gamma_8 OTFAMI_i + \gamma_9 EXPER_i + \gamma_{10} SES_i + \gamma_{11} FAM_i) \tag{22}$$

$$SEI_i = \delta_0 + \delta_1 AGE_i + \delta_2 KIDLT6_i + \delta_3 ED_i + \delta_4 EVERTR_i + \delta_5 OTFAMI_i + \delta_6 EXPER_i + \delta_7 SES_i + \delta_8 FAM_i + \sigma_{12}\lambda_i + V_{2i} \tag{23}$$

$p(EMP_i)$ is the probability that the i^{th} woman was employed at the time of the interview; AGE_i is the age of the woman in years; $KIDLT6_i$ is the number of children living in the household in 1967 who were under six years of age; $KID613_i$ is the number of children in the household in 1967 aged 6-13; $KID1417_i$ is the number of children living in the household aged 14-17; ED_i is the woman's number of years of formal schooling completed; AAM_i is the age of the respondent at first marriage; $EVERTR_i$ is a dummy variable which assumes a value of one if the woman has ever received training other than formal schooling; $OTFAMI_i$ is the total family income minus the wife's earnings if she was employed; $EXPER_i$ is the proportion of years between last attending school full-time and the time of interview that a woman was employed at least six months; SES_i is a linearly combined factor score of father's (head of household's) occupational status, father's (head of household's) education and mother's education; FAM_i is a factor score for farm origin and number of siblings; SEI_i is the Duncan (1961) socioeconomic index score of the occupation that the woman held at the time of interview; γ_0 and δ_0 are intercepts; $\gamma_{1}\ldots\gamma_{11}$ are the parameter estimates from the probit analysis; $\delta_{1}\ldots\delta_8$ are estimates of the structural parameters from the ordinary least squares, σ_{12} is the covariance of the errors across equations, and λ_i is the inverse of the Mill's ratio.

We briefly discuss the equation with employment as the dependent variable since it is not of major concern here and is only important in that it provides estimates of ϕ_i from which λ_i' s are obtained. EXPER, ED, and EVERTR are expected to have substantial positive effects while KIDLT6, OTFAMI and AAM are expected to negatively affect employment (Sweet, 1973; Mott, 1972; Bowen and Finegan, 1969; Cain, 1966; Waite, 1976). It should be noted that neither the woman's potential wage rate nor her potential occupational status is included explicitly in this equation despite the fact that both of these variables would be expected to positively affect labor force participation. Instead, potential wage rate and potential occupational status are implicitly included by entering into the equation the determinants of these variables. For example, schooling, experience, and training are included as proxies for potential wage.

The major concern is the parameter estimates from the equation with SEI as the dependent variable, Included in this equation are (1) variables that have been shown to have effects on occupational attainment, i.e. education, family of origin characteristics and labor force participation (Wang, 1973; Wolf, 1975; Featherman and Hauser, 1976; Treiman and Terrell, 1975); (2) variables that had been expected by several researchers to affect a woman's occupational attainment, but whose effects have not been borne out empirically, i.e. KIDLT6 and AGE (Wolf, 1975; Sheehy, 1975; McClendon, 1976); and (3) other family income. This second group of variables which represent career contingencies and other factors related to the family of procreation is included because it is possible that the bias due to the censoring problem could be affecting the parameter estimates of these variables. Other family income is included because it implicitly allows us to inspect the selectivity bias hypothesis. It is possible that women whose families have high "other family incomes" are less likely to be employed. However, if they do choose to take a job, they can be selective in the types of jobs they take.

Data for this study are from the 1967 National Longitudinal Survey of Mature Women, aged 30-44 (Parnes, et al., 1970), chosen because it is the only national data set with satisfactory labor force experience measures for women. The subpopulation used in this study is all white, currently married females who were 30-44 years old in 1967. Of the 3112 women in the subpopulation, 1679 had data on all variables and, therefore, could be used in the probit analysis. Seven hundred sixty five women who were employed and met the other criteria were included in the SEI regression. The missing data were a problem. Those responding to all items tended to have slightly higher levels of education, occupational status and labor force participation. While the mean levels differ, their effects on the correlations and parameter estimates are minimal.[2/]

Empirical Results

Table 1 presents the results from the probit analysis. The number of children in the household under six years of age, other family income and age at first marriage have negative net effects on the probability of employment while extent of labor market experience, educational attainment and the number of children living in the household who are 14 to 17 have positive net effects.

TABLE 1 -- Results of Probit Analysis (Equation 22) Where the Dependent Variable is Employment at the Time of Interview (N = 1679)

Variable	Maximum Likelihood Estimate of Coefficient	Standard Error	Ratio of MLE/STD Error
Constant	-.457	.425	-1.076
AGE	.005	.010	.551
KIDLT6	-.479	.053	-9.105
KID613	.009	.031	.293
KID1417	.256	.050	5.126
ED	.088	.018	4.816
AAM	-.078	.012	-6.578
EVERTR	-.075	.076	- .986
OTFAMI	-.000027	.000008	-3.321
EXPER	2.463	.141	17.412
SES	-.016	.043	- .385
FAM	-.004	.037	- .023

Where KIDLT6 = number of children in household under 6 years old; KID613 = number of children in household ages 6 to 13; KID1417 = number of children in household ages 14 to 17; EVERTR = dummy variable signifying whether the woman has experienced training other than formal schooling; EXPER = proportion of years since last school attendance in which the woman worked at least six months; AGE = age in years; ED = number of years of formal schooling completed; AAM = age at first marriage; FAM = factor score for farm origin and number of siblings; SES = factor score for socioeconomic status of family of origin; OTFAMI = other family income in 1966.

These results are as expected and do not warrant further discussion. Table 2 presents the means and standard deviations of the variables in Equations (22) and (23) as well as the correlations between the variables in Equation (23) and λ_i. It should be noted that if a variable is positively related to employment, it is negatively related to λ_i. This is a manifestation of the fact that $\phi_i \to +\infty$, $\lambda_i \to 0$.

Table 3 presents the results from two sets of ordinary least squares regressions. The first three columns are the results of the estimation of an equation that does not take into account the conditional distribution of the errors. This is the same type of equation that is usually estimated by researchers interested in female occupational attainment. The next three columns present the estimates of Equation (23) or (20) without presenting the coefficient for λ_i. These coefficients represent the structural parameters of the process of occupational attainment of married women. This regression, by adding λ_i as an additional regressor, eliminates the potential bias in the original equation due to the fact that its error is conditional on the woman being employed.

The inclusion of λ_i alters some of the coefficients and standard errors. This is because it is correlated with the exogenous variables in the occupational attainment equation. In the equation without λ_i, EXPER has a positive, barely statistically significant effect.[3/] If a woman works at least six months or more in all years since leaving school full-time, as opposed to not working at all, she would gain 7.63 SEI points according to the misspecified equation. In the equation with λ_i, the unstandardized effect of EXPER, given that the woman has worked in all years, is 8.618 SEI points and its standard error has increased from 2.07 to 3.45. The effect of EXPER is no longer statistically significant.[4/]

TABLE 2 -- Means and Standard Deviations for the Variables in Equations 22 and 23; Correlations Between λ and the Variables in the Occupational Attainment Equations (N = 765)

	Mean	S.D.	Correlations With λ
SEI	40.103	20.485	-.143
KIDLT6	.324	.624	.249
KID613	1.212	1.176	
KID1417	.642	.783	
EVERTR	.352	.478	-.049
EXPER	.562	.293	-.759
AGE	37.498	4.260	-.030
ED	11.698	2.418	-.130
AAM	19.814	3.385	
FAM	.120	.968	-.014
SES	-.052	.968	.041
OTFAMI	8086.200	4394.554	.067
λ	.609	.339	1.000

Where SEI = occupational attainment in 1967; KIDLT6 = number of children in household under 6 years old; KID613 = number of children in household ages 6-13; KID1417 = number of children in household ages 14-17; EVERTR = dummy variables signifying whether the woman has experienced training other than formal schooling; EXPER = proportion of years since last school attendance in which the woman worked at least six months; AGE = age in years; ED = number of years of formal schooling completed; AAM = age at first marriage; FAM = factor score for farm origin and number of sibs; SES = factor score for socioeconomic status of family of origin; λ = inverse of the Mill's Ratio; OTFAMI = other family income in 1966.

TABLE 3 -- Parameter Estimates of Occupational Status Equations for the Subsample of Working Women (N = 765)

	Parameter Estimates Without λ			Structural Parameters Obtained from Equation 23		
	Regression Coefficient	Standard Error	Standardized Coefficient	Regression Coefficient	Standard Error	Standardized Coefficient
KIDLT6	- 1.503	1.013	-.046	- 1.680	1.129	-.051
EVERTR	3.636*	1.287	.085*	3.591*	1.293	.084*
EXPER	7.630*	2.070	.109*	8.618	3.452	.123
AGE	.082	.149	.017	.081	.149	.017
ED	4.078*	.295	.481*	4.098*	.300	.484*
FAM	-.132	.628	-.006	-.141	.629	-.007
SES	1.207	.729	.057	1.185	.732	.056
OTFAMI	.00070*	.00014	.150*	.00069*	.00014	.148*
Intercept	-21.304			-22.609		
R^2	.370			.370		

*Significant at .01 level. (See Footnote 3.)

Where KIDLT6 = number of children in household less than 6 years old, EVERTR = dummy variable signifying whether the woman has experienced training, other than formal schooling, EXPER = proportion of years since last school attendance in which the woman worked at least six months, AGE = age in years, ED = number of years of formal schooling completed, AAM = age at first marriage, FAM = factor score for farm origin and number of siblings, SES = factor score for socioeconomic status of family of origin, λ = inverse of Mill's ratio, OTFAMI = other family income.

Although the size of the parameter estimate for work experience in the corrected equation is large, so is its standard error. Although past work experience has the most powerful net effect on whether a woman is currently employed, the effect of work experience on occupational attainment is minimal, in that it is highly variable across individuals. This result is not in conflict with earlier research which suggests that labor force interruptions have minimal effects on women's occupational attainments (Wolf, 1975; Rosenfeld, 1976), despite their documented effects on women's earnings.

Except for the difference in the effects of experience, the coefficients of the other variables are remarkably similar in both equations, suggesting that the censoring problem has only minimal effects on the structural parameters.

After correctly specifying the occupational attainment equation, there are three variables that have statistically significant and substantively important effects on a woman's occupational attainment: ED, EVERTR, and OTFAMI. The education effect is such that a one-year increase in educational attainment results in a 4.10 point increase in current occupational status. If a woman has had non-formal schooling, she experiences an increase of 3.59 points in current occupational status.

The OTFAMI effect is suggestive. It is statistically significant; a $10,000 increase in other family income results in a 6.9 point increase in current occupational attainment. Although this is not a large effect since $10,000 is twice the standard deviation, this is the second most powerful effect in the corrected occupational attainment equation. Other family income's positive effect in the equation without λ_i coupled with its negative effect in the probit analysis suggested to us that women whose families have high other family income are less likely to be employed, but if they are employed they are likely to have higher status jobs, net of all of the variables included in the model. This might be due to the fact that women who could afford not to return to employment, would wait to return to work until they found a job commensurate with their education. The fact that the other family income effect is stable after controlling for the censoring problem indicates that this explanation is unacceptable. Two alternative explanations seem possible. First, women whose families have high incomes probably have more and better contacts in the job market and thus are better able to find high status jobs. Second, the other family income effect could be due to assortative mating; that is, people of like statuses tend to intermarry. These data do not allow us to discriminate between the two.

Conclusion

This paper investigates one potential source of bias in estimating equations for women's occupational attainments. This bias is due to the exclusion of nonworking women from the occupational attainment equation. We present a technique which allows us to estimate the structural parameters for all currently married women, regardless of their employment status. The fact that the structural parameters obtained by including λ_i as a regressor are, in general, remarkably similar to the ordinary least squares estimates for working women suggests that the bias due to the censoring problem is minimal. However, the structural parameters are superior to the ordinary least squares estimates without λ_i because the structural parameters better describe the process for the total population of currently married women.

Because the censoring problem appears to be minimal, selection into the sample of working women on the basis of the dependent variable is not a reasonable explanation for the apparent similarities between men and working women in the process of occupational attainment. Given that the censoring problem was the major speculation for the apparent sex similarities in occupational attainment and it was not found to be important empirically, we are left with three possible arguments concerning sexual inequality in occupational rewards. First, it could be argued that the model is misspecified and if certain variables (for example, the status of first job) were included, the process of occupational attainment for the sexes would differ. Second, one could accept the finding of sexual equalities in occupational rewards, using the evidence from sex differences in occupational status attainment. Last, one could argue that certain dimensions of sexual inequality in occupational rewards are not being tapped by the concept, occupational status. One possible example is authority relations in the work setting. This dimension of jobs has been shown to have an effect on income net of occupation status for both sexes (Wright, 1977). The implication of this finding is that authority relations could be an important aspect of sexual inequality in occupational rewards. Before making generalizations about sexual inequality in the occupational structure from studies of sex differences in occupational status attainment, it seems reasonable to inspect other dimensions of jobs that might be important in the study of sexual inequality in occupational rewards. We are reluctant to accept the finding of sexual equality in occupational rewards before exploring the first and third possibilities.

Footnotes

1/ Whereas Heckman is concerned with truncation from below (i.e. $E(X \mid X > \phi)$, which equals $\mu + \sigma^2\lambda$) we are concerned with truncation from above (i.e. $E(X \mid \phi > X)$ which equals $\mu - \sigma^2\lambda$) (Crawford, 1975). Lambda, in Heckman, is equal to $\frac{f(-\phi)}{1-F(-\phi)}$; our lambda equals $\frac{f(\phi)}{F(\phi)}$. These are equivalent.

2/ We constructed ϕ's and λ's in two fashions: (1) assigned a missing value whenever a value was missing on any of the exogenous variables in the probit equation; and (2) substituted means for missing values on the exogenous variables. The parameter estimates in Equation (23) did not vary depending on (1) whether we deleted cases listwise or pairwise and/or (2) used the λ_i's constructed in the different fashions.

3/ We use the .01 level as a criterion for statistical significance because of the effect of the nonrandom sampling design. By using the .01 level, we have, in effect, a .05 level of significance (Rosenfeld, 1976).

4/ Heckman (1975) argues that the errors in the structural equations, after correcting for censoring, are heteroscedastic. This tends to increase the estimates of the standard error and our criteria for significance are, thus, slightly conservative (Theil, 1971).

References

Alexander, Karl L. and Bruce K. Eckland

1975 "School experience and status attainment." Pp. 171-209 in S. E. Dragastin and G. E. Elder, Jr. (eds), Adolescence in the Life Cycle. Washington, D.C.: Hemisphere Publishing Company.

Bose, Christine

1973 "Jobs and gender: sex and occupational prestige." Ph.D. dissertation, Johns Hopkins University.

Bowen, William G. and T. Aldrich Finegan

1969 The Economics of Labor Force Participation. Princeton: Princeton University Press.

Cain, Glen

1966 Married Women in the Labor Force. Chicago: University of Chicago Press.

Crawford, David L.

1975a "The wage work response of elderly heads." Mimeographed paper, University of Wisconsin-Madison.

1975b "Estimating earnings functions from truncated samples." Institute for Research on Poverty Discussion Paper No. 287-75, University of Wisconsin-Madison.

Duncan, Otis Dudley

1961 "A socioeconomic index for all occupations." Chapter 6 in Albert J. Reiss, Jr., _et al_., Occupations and Social Status. Glencoe: Free Press.

Featherman, David L. and Robert M. Hauser

1976 "Sexual inequalities and socioeconomic achievement in the U.S., 1962-1973." American Sociological Review 41 (June): 462-484.

Heckman, James J.

1974 "Shadow prices, market wages, and labor supply." Econometrica 42: 679-694.

1975 "Shadow prices, market wages, and labor supply revisited: some computational and conceptual simplifications and revised estimates." Mimeographed paper, University of Chicago.

Hotz, V. Joseph

1975 "The labor supply response of married women in the New Jersey negative income tax experiment." Mimeographed paper, University of Wisconsin-Madison.

Johnson, Norman L. and Samuel Kotz

1970 Continuous Univariate Distributions, Volume 1. New York: Houghton-Mifflin.

1972 Continuous Multivariate Distributions. New York: John Wiley.

McClendon, McKee J,

1976 "The occupational status attainment processes of males and females." American Sociological Review 41 (February): 52-65.

Mott, Frank L.

1972 "Labor force participation and fertility for women with small children in Rhode Island: an analysis of their interactions and antecedents." Ph.D. dissertation, Brown University.

Nerlove, Marc and S. James Press

1973 Univariate and Multivariate Log-Linear and Logistic Models. California: Rand.

Parnes, Herbert S., John Shea, Ruth Spitz and Frederick Zeller

1970 Dual Careers, Volume 1. Washington, D.C.: U.S. Department of Labor, Manpower Research Monograph No. 21

Rosenfeld, Rachel

1976 "Women's employment patterns and occupational achievements." Ph.D. dissertation, University of Wisconsin-Madison.

Sheehy, Linda L.

1975 "A model of educational and occupational attainment for female labor force participants." Master's thesis, University of Wisconsin-Madison.

Sweet, James A.

1973 Women in the Labor Force. New York: Seminar.

Theil, Henri

1971 Principles of Econometrics. New York: John Wiley.

Treiman, Donald J. and Kermit Terrell

1975 "Sex and the process of status attainment: a comparison of working women and men." American Sociological Review 40 (April): 174-201.

Waite, Linda J.

1976 "Working wives: 1940-1960." American Sociological Review 41 (February): 65-80.

Wang, Linda

1973 "The female occupational attainment process and occupational mobility." Master's thesis, University of Wisconsin-Madison.

Wolf, Wendy C.

1975 "Occupational attainment of married women: do career contingencies matter?" Ph.D. dissertation, Johns Hopkins University.

1976 "Occupational attainments of married women: do career contingencies matter?" Working Paper No. 76-3, Center for Demography and Ecology, University of Wisconsin-Madison.

Wright, Erik O. and Luca Perrone

1977 "Marxist class categories and income inequality." American Sociological Review (February, 1977), forthcoming.