Sample Selection Bias as a Specification Error

Author(s): James J. Heckman

Source: *Econometrica*, Jan., 1979, Vol. 47, No. 1 (Jan., 1979), pp. 153-161

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/1912352

Econometrica, Vol. 47, No. 1 (January, 1979)

SAMPLE SELECTION BIAS AS A SPECIFICATION ERROR

BY JAMES J. HECKMAN[1]

This paper discusses the bias that results from using nonrandomly selected samples to estimate behavioral relationships as an ordinary specification error or "omitted variables" bias. A simple consistent two stage estimator is considered that enables analysts to utilize simple regression methods to estimate behavioral functions by least squares methods. The asymptotic distribution of the estimator is derived.

THIS PAPER DISCUSSES the bias that results from using nonrandomly selected samples to estimate behavioral relationships as an ordinary specification bias that arises because of a missing data problem. In contrast to the usual analysis of "omitted variables" or specification error in econometrics, in the analysis of sample selection bias it is sometimes possible to estimate the variables which when omitted from a regression analysis give rise to the specification error. The estimated values of the omitted variables can be used as regressors so that it is possible to estimate the behavioral functions of interest by simple methods. This paper discusses sample selection bias as a specification error and presents a simple consistent estimation method that eliminates the specification error for the case of censored samples. The argument presented in this paper clarifies and extends the analysis in a previous paper **[6]** by explicitly developing the asymptotic distribution of the simple estimator for the general case rather than the special case of a null hypothesis of no selection bias implicitly discussed in that paper. Accordingly, for reasons of readability, this paper recapitulates some of the introductory material of the previous paper in, what is hoped, an improved and simplified form.

Sample selection bias may arise in practice for two reasons. First, there may be self selection by the individuals or data units being investigated. Second, sample selection decisions by analysts or data processors operate in much the same fashion as self selection.

There are many examples of self selection bias. One observes market wages for working women whose market wage exceeds their home wage at zero hours of work. Similarly, one observes wages for union members who found their nonunion alternative less desirable. The wages of migrants do not, in general, afford a reliable estimate of what nonmigrants would have earned had they migrated. The earnings of manpower trainees do not estimate the earnings that nontrainees would have earned had they opted to become trainees. In each of these examples, wage or earnings functions estimated on selected samples do not,

[1] This research was supported by a HEW grant to the Rand Corporation and a U.S. Department of Labor grant to the National Bureau of Economic Research. A previous version of this paper circulated under the title "Shadow Prices, Market Wages and Labor Supply Revisited: Some Computational Simplifications and Revised Estimates," June, 1975. An embarrassingly large number of colleagues have made valuable comments on this paper, and its various drafts. Particular thanks go to Takeshi Amemiya, Zvi Griliches, Reuben Gronau, Mark Killingsworth, Ed Leamer, Tom MaCurdy, Bill Rodgers, and Paul Schultz. I bear full responsibility for any remaining errors.

in general, estimate population (i.e., random sample) wage functions. Comparisons of the wages of migrants with the wages of nonmigrants (or trainee earnings with nontrainee earnings, etc.) result in a biased estimate of the effect of a *random* "treatment" of migration, manpower training, or unionism.

Data may also be nonrandomly selected because of decisions taken by data analysts. In studies of panel data, it is common to use "intact" observations. For example, stability of the family unit is often imposed as a requirement for entry into a sample for analysis. In studies of life cycle fertility and manpower training experiments, it is common practice to analyze observations followed for the full length of the sample, i.e., to drop attriters from the analysis. Such procedures have the same effect on structural estimates as self selection: fitted regression functions confound the behavioral parameters of interest with parameters of the function determining the probability of entrance into the sample.

1. A SIMPLE CHARACTERIZATION OF SELECTION BIAS

To simplify the exposition, consider a two equation model. Few new points arise in the multiple equation case, and the two equation case has considerable pedagogical merit.

Consider a random sample of I observations. Equations for individual i are

$$Y_{1i} = X_{1i}\beta_1 + U_{1i}, \tag{1a}$$

$$Y_{2i} = X_{2i}\beta_2 + U_{2i} \qquad (i = 1, \ldots, I), \tag{1b}$$

where X_{ji} is a $1 \times K_j$ vector of exogenous regressors, β_j is a $K_j \times 1$ vector of parameters, and

$$\begin{aligned} E(U_{ji}) = 0, \qquad E(U_{ji}U_{j'i''}) &= \sigma_{jj'}, \quad i = i'', \\ &= 0, \qquad i \neq i''. \end{aligned}$$

The final assumption is a consequence of a random sampling scheme. The joint density of U_{1i}, U_{2i} is $h(U_{1i}, U_{2i})$. The regressor matrix is of full rank so that if all data were available, the parameters of each equation could be estimated by least squares.

Suppose that one seeks to estimate equation (1a) but that data are missing on Y_1 for certain observations. The critical question is "why are the data missing?"

The *population* regression function for equation (1a) may be written as

$$E(Y_{1i} \mid X_{1i}) = X_{1i}\beta_1 \qquad (i = 1, \ldots, I).$$

The regression function for the subsample of available data is

$$E(Y_{1i} \mid X_{1i}, \text{ sample selection rule}) = X_{1i}\beta_1 + E(U_{1i} \mid \text{sample selection rule}),$$

$i = 1, \ldots, I$, where the convention is adopted that the first $I_1 < I$ observations have data available on Y_{1i}.

If the conditional expectation of U_{1i} is zero, the regression function for the selected subsample is the same as the population regression function. Least squares estimators may be used to estimate β_1 on the selected subsample. The only cost of having an incomplete sample is a loss in efficiency.

In the general case, the sample selection rule that determines the availability of data has more serious consequences. Suppose that data are available on Y_{1i} if $Y_{2i} \geqslant 0$ while if $Y_{2i} < 0$, there are no observations on Y_{1i}. The choice of zero as a threshold involves an inessential normalization.

In the general case

$$E(U_{1i} \mid X_{1i}, \text{sample selection rule}) = E(U_{1i} \mid X_{1i}, Y_{2i} \geqslant 0)$$
$$= E(U_{1i} \mid X_{1i}, U_{2i} \geqslant -X_{2i}\beta_2).$$

In the case of independence between U_{1i} and U_{2i}, so that the data on Y_{1i} are missing randomly, the conditional mean of U_{1i} is zero. In the general case, it is nonzero and the subsample regression function is

$$(2) \qquad E(Y_{1i} \mid X_{1i}, Y_{2i} \geqslant 0) = X_{1i}\beta_1 + E(U_{1i} \mid U_{2i} \geqslant -X_{2i}\beta_2).$$

The selected sample regression function depends on X_{1i} and X_{2i}. Regression estimators of the parameters of equation (1a) fit on the selected sample omit the final term of equation (2) as a regressor, so that the bias that results from using nonrandomly selected samples to estimate behavioral relationships is seen to arise from the ordinary problem of omitted variables.

Several points are worth noting. First, if the only variable in the regressor vector X_{2i} that determines sample selection is "1" so that the probability of sample inclusion is the same for all observations, the conditional mean of U_{1i} is a constant, and the only bias in β_1 that results from using selected samples to estimate the population structural equation arises in the estimate of the intercept. One can also show that the least squares estimator of the population variance σ_{11} is downward biased. Second, a symptom of selection bias is that variables that do not belong in the true structural equation (variables in X_{2i} not in X_{1i}) may appear to be statistically significant determinants of Y_{1i} when regressions are fit on selected samples. Third, the model just outlined contains a variety of previous models as special cases. For example, if $h(U_{1i}, U_{2i})$ is assumed to be a singular normal density ($U_{1i} \equiv U_{2i}$) and $X_{2i} = X_{1i}$, $\beta_1 \equiv \beta_2$, the "Tobit" model emerges. For a more complete development of the relationship between the model developed here and previous models for limited dependent variables, censored samples and truncated samples, see Heckman [**6**]. Fourth, multivariate extensions of the preceding analysis, while mathematically straightforward, are of considerable substantive interest. One example is offered. Consider migrants choosing among K possible regions of residence. If the self selection rule is to choose to migrate to that region with the highest income, both the self selection rule and the subsample regression functions can be simply characterized by a direct extension of the previous analysis.

2. A SIMPLE ESTIMATOR FOR NORMAL DISTURBANCES AND ITS PROPERTIES[2]

Assume that $h(U_{1i}, U_{2i})$ is a bivariate normal density. Using well known results (see [**10**, pp. 112–113]),

$$E(U_{1i} \mid U_{2i} \geqslant -X_{2i}\beta_2) = \frac{\sigma_{12}}{(\sigma_{22})^{\frac{1}{2}}}\lambda_i,$$

$$E(U_{2i} \mid U_{2i} \geqslant -X_{2i}\beta_2) = \frac{\sigma_{22}}{(\sigma_{22})^{\frac{1}{2}}}\lambda_i,$$

where

$$\lambda_i = \frac{\phi(Z_i)}{1-\Phi(Z_i)} = \frac{\phi(Z_i)}{\Phi(-Z_i)},$$

where ϕ and Φ are, respectively, the density and distribution function for a standard normal variable, and

$$Z_i = -\frac{X_{2i}\beta_2}{(\sigma_{22})^{\frac{1}{2}}}.$$

"λ_i" is the inverse of Mill's ratio. It is a monotone decreasing function of the probability that an observation is selected into the sample, $\Phi(-Z_i)\ (=1-\Phi(Z_i))$. In particular, $\lim_{\Phi(-Z_i)\to 1} \lambda_i = 0$, $\lim_{\Phi(-Z_i)\to 0} \lambda_i = \infty$, and $\partial\lambda_i/\partial\Phi(-Z_i) < 0$.

The full statistical model for normal population disturbances can now be developed. The conditional regression function for selected samples may be written as

$$E(Y_{1i} \mid X_{1i}, Y_{2i} \geqslant 0) = X_{1i}\beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{\frac{1}{2}}}\lambda_i,$$

$$E(Y_{2i} \mid X_{2i}, Y_{2i} \geqslant 0) = X_{2i}\beta_2 + \frac{\sigma_{22}}{(\sigma_{22})^{\frac{1}{2}}}\lambda_i,$$

(4a) $$Y_{1i} = E(Y_{1i} \mid X_{1i}, Y_{2i} \geqslant 0) + V_{1i},$$

(4b) $$Y_{2i} = E(Y_{2i} \mid X_{2i}, Y_{2i} \geqslant 0) + V_{2i},$$

where

(4c) $$E(V_{1i} \mid X_{1i}, \lambda_i, U_{2i} \geqslant -X_{2i}\beta_2) = 0,$$

(4d) $$E(V_{2i} \mid X_{2i}, \lambda_i, U_{2i} \geqslant -X_{2i}\beta_2) = 0,$$

(4e) $$E(V_{ji}V_{j'i'} \mid X_{1i}, X_{2i}, \lambda_1, U_{2i} \geqslant -X_{2i}\beta_2) = 0,$$

[2] A grouped data version of the estimation method discussed here was first proposed by Gronau [**4**] and Lewis [**11**]. However, they do not investigate the statistical properties of the method or develop the micro version of the estimator presented here.

for $i \neq i'$. Further,

(4f) $$E(V_{1i}^2 \mid X_{1i}, \lambda_i, U_{2i} \geqslant -X_{2i}\beta_2) = \sigma_{11}((1-\rho^2)+\rho^2(1+Z_i\lambda_i-\lambda_i^2)),$$

(4g) $$E(V_{1i}V_{2i} \mid X_{1i}, X_{2i}, \lambda_i, U_{2i} \geqslant -X_{2i}\beta_2) = \sigma_{12}(1+Z_i\lambda_i-\lambda_i^2),$$

(4h) $$E(V_{2i}^2 \mid X_{2i}, \lambda_i, U_{2i} \geqslant -X_{2i}\beta_2) = \sigma_{22}(1+Z_i\lambda_i-\lambda_i^2),$$

where

$$\rho^2 = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}$$

and

(5) $$0 \leqslant 1+\lambda_i Z_i - \lambda_i^2 \leqslant 1.$$

If one knew Z_i and hence λ_i, one could enter λ_i as a regressor in equation (4a) and estimate that equation by ordinary least squares. The least squares estimators of β_1 and $\sigma_{12}/(\sigma_{22})^{\frac{1}{2}}$ are unbiased but inefficient. The inefficiency is a consequence of the heteroscedasticity apparent from equation (4f) when X_{2i} (and hence Z_i) contains nontrivial regressors. As a consequence of inequality (5), the standard least squares estimator of the population variance σ_{11} is downward biased. As a consequence of equation (4g) and inequality (5), the usual estimator of the interequation covariance is downward biased. A standard GLS procedure can be used to develop appropriate standard errors for the estimated coefficients of the first equation (see Heckman [**6**]).

In practice, one does not know λ_i. But in the case of a censored sample, in which one does not have information on Y_{1i} if $Y_{2i} \leqslant 0$, but one does know X_{2i} for observations with $Y_{2i} \leqslant 0$, one can estimate λ_i by the following procedure:

(1) Estimate the parameters of the probability that $Y_{2i} \geqslant 0$ (i.e., $\beta_2/(\sigma_{22})^{\frac{1}{2}}$) using probit analysis for the full sample.[3]

(2) From this estimator of $\beta_2/(\sigma_{22})^{\frac{1}{2}}$ ($=\beta_2^*$) one can estimate Z_i and hence λ_i. All of these estimators are consistent.

(3) The estimated value of λ_i may be used as a regressor in equation (4a) fit on the selected subsample. Regression estimators of equation (4a) are *consistent* for β_1 and $\sigma_{12}/(\sigma_{22})^{\frac{1}{2}}$ (the coefficients of X_{1i} and λ_1, respectively).[4]

(4) One can consistently estimate σ_{11} by the following procedure. From step 3, one consistently estimates $C = \rho(\sigma_{11})^{\frac{1}{2}} = \sigma_{12}/(\sigma_{22})^{\frac{1}{2}}$. Denote the residual for the ith observation obtained from step 3 as $\hat{V}_{1i}$, and the estimator of C by $\hat{C}$. Then an estimator of σ_{11} is

$$\hat{\sigma}_{11} = \frac{\sum_{i=1}^{I_1} \hat{V}_{1i}^2}{I_1} - \frac{\hat{C}}{I_1}\sum_{i=1}^{I_1}(\hat{\lambda}_1\hat{Z}_i - \hat{\lambda}_i^2)$$

[3] In the case in which Y_{2i} is observed, one can estimate β_2, σ_{22}, and hence $\beta_2/(\sigma_{22})^{\frac{1}{2}}$ by ordinary least squares.

[4] It is assumed that vector X_{2i} contains nontrivial regressors or that β_1 contains no intercept or both.

where $\hat{\lambda}_i$ and $\hat{Z}_i$ are the estimated values of Z_i and λ_i obtained from step 2. This estimator of σ_{11} is consistent and positive since the term in the second summation must be negative (see inequality (5)).

The usual formulas for standard errors for least squares coefficients are *not* appropriate except in the important case of the null hypothesis of no selection bias ($C = \sigma_{12}/(\sigma_{22})^{\frac{1}{2}} = 0$). In that case, the usual regression standard errors are appropriate and an exact test of the null hypothesis $C = 0$ can be performed using the t distribution. If $C \neq 0$, the usual procedure for computing standard errors *understates* the true standard errors and overstates estimated significance levels.

The derivation of the correct limiting distribution for this estimator in the general case requires some argument.[5] Note that equation (4a) with an estimated value of λ_i used in place of the true value of λ_i may be written as

$$(4a') \qquad Y_{1i} = X_{1i}\beta_1 + C\hat{\lambda}_i + C(\lambda_i - \hat{\lambda}_i) + V_{1i}.$$

The error term in the equation consists of the final two terms in the equation.

Since λ_i is estimated by $\beta_2/(\sigma_{22})^{\frac{1}{2}}$ $(=\beta_2^*)$ which is estimated from the entire sample of I observations by a maximum likelihood probit analysis,[6] and since λ_i is a twice continuously differentiable function of β_2^*, $\sqrt{I}(\hat{\lambda}_i - \lambda_i)$ has a well defined limiting normal distribution

$$\sqrt{I}(\hat{\lambda}_i - \lambda_i) \sim N(0, \Sigma_i)$$

where Σ_i is the asymptotic variance-covariance matrix obtained from that of β_2^* by the following equation:

$$\Sigma_i = \left(\frac{\partial \lambda_i}{\partial Z_i}\right)^2 X_{2i} \Sigma X_{2i}',$$

where $\partial \lambda_i / \partial Z_i$ is the derivative of λ_i with respect to Z_i, and Σ is the asymptotic variance-covariance matrix of $\sqrt{I}(\hat{\beta}_2^* - \beta_2^*)$.

We seek the limiting distribution of

$$\sqrt{I_1}\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{C} - C \end{pmatrix} = I_1 \begin{pmatrix} \Sigma X_{1i}' X_{1i} & \Sigma X_{1i}' \hat{\lambda}_i \\ \Sigma X_{1i} \hat{\lambda}_i & \Sigma \hat{\lambda}_i^2 \end{pmatrix}^{-1} \frac{1}{\sqrt{I_1}} \begin{pmatrix} \Sigma X_{1i}'(C(\lambda_i - \hat{\lambda}_i) + V_{1i}) \\ \Sigma \hat{\lambda}_i (C(\lambda_i - \hat{\lambda}_i) + V_{1i}) \end{pmatrix}.$$

In the ensuing analysis, it is important to recall that the probit function is estimated on the entire sample of I observations whereas the regression analysis is performed solely on the subsample of I_1 ($<I$) observations where Y_{1i} is observed. Further, it is important to note that unlike the situation in the analysis of two stage least squares procedures, the portion of the residual that arises from the use of an estimated value of λ_i in place of the actual value of λ_i *is not* orthogonal to the X_1 data vector.

[5] This portion of the paper was stimulated by comments from T. Amemiya. Of course, he is not responsible for any errors in the argument.

[6] The ensuing analysis can be modified in a straightforward fashion if Y_{2i} is observed and β_2^* is estimated by least squares.

Under general conditions for the regressors discussed extensively in Amemiya [1] and Jennrich [9],

$$\operatorname*{plim}_{I_1\to\infty} I_1\begin{pmatrix}\Sigma X'_{1i}X_{1i} & \Sigma X'_{1i}\hat{\lambda}_i\\ \Sigma X_{1i}\hat{\lambda}_i & \Sigma\hat{\lambda}_i^2\end{pmatrix}^{-1} = \operatorname*{plim}_{I_1\to\infty} I_1\begin{pmatrix}\Sigma X'_{1i}X_{1i} & \Sigma X'_{1i}\lambda_i\\ \Sigma X_{1i}\lambda_i & \Sigma\lambda_i^2\end{pmatrix}^{-1} = B,$$

where B is a finite positive definite matrix.[7] Under these assumptions,

$$\sqrt{I_1}\begin{pmatrix}\hat{\beta}_1-\beta_1\\ \hat{C}-C\end{pmatrix} \sim N(0, B\psi B')$$

where

$$\psi = \operatorname*{plim}_{\substack{I_1\to\infty\\ I\to\infty}}\left[\sigma_{11}\begin{pmatrix}\dfrac{\Sigma X'_{1i}X_{1i}\eta_i}{I_1} & \dfrac{\Sigma X'_{1i}\lambda_i\eta_i}{I_1}\\ \dfrac{\Sigma\lambda_i X_{1i}\eta_i}{I_1} & \dfrac{\Sigma\lambda_i^2\eta_i}{I_1}\end{pmatrix} + C^2\left(\frac{I_1}{I}\right)\begin{pmatrix}\displaystyle\sum_{i=1,i'=1}^{I_1}\sum^{I_1}\frac{X'_{1i}X_{1i'}\theta_{ii'}}{I_1^2} & \displaystyle\sum_{i=1,i'=1}^{I_1}\sum^{I_1}\frac{X'_{1i}\pi_{ii'}}{I_1^2}\\ \displaystyle\sum_{i'=1,i=1}^{I_1}\sum^{I_1}\frac{X_{1i}\pi_{ii'}}{I_1^2} & \displaystyle\sum_{i=1,i'=1}^{I_1}\sum^{I_1}\frac{\Omega_{ii'}}{I_1^2}\end{pmatrix}\right]$$

$$\operatorname*{plim}_{\substack{I\to\infty\\ I_1\to\infty}}\frac{I_1}{I} = k, \quad 0 < k < 1,$$

where

$$C = \sigma_{12}/(\sigma_{22})^{\frac{1}{2}},$$

$$\eta_i = (1 + C^2(Z_i\lambda_i - \lambda_i^2)/\sigma_{11}),$$

$$\pi_{ii'} = \left(\frac{\partial\lambda_i}{\partial Z_i}\right)\left(\frac{\partial\lambda_{i'}}{\partial Z_{i'}}\right)\lambda_i X_{2i}\Sigma X'_{2i'},$$

$$\theta_{ii'} = \left(\frac{\partial\lambda_i}{\partial Z_i}\right)\left(\frac{\partial\lambda_{i'}}{\partial Z_{i'}}\right)X_{2i}\Sigma X'_{2i},$$

$$\Omega_{ii'} = (\lambda_i\lambda_{i'})\left(\frac{\partial\lambda_i}{\partial Z_i}\right)\left(\frac{\partial\lambda_{i'}}{\partial Z_{i'}}\right)X_{2i}\Sigma X'_{2i},$$

where $\partial\lambda_i/\partial Z_i$ is the derivative of λ_i with respect to Z_i,

$$\frac{\partial\lambda_i}{\partial Z_i} = \lambda_i^2 - Z_i\lambda_i.$$

Note that if $C = 0$, $B\psi B'$ collapses to the standard variance-covariance matrix for the least squares estimator. Note further that because the second matrix in ψ is positive definite, if $C \neq 0$, the correct asymptotic variance-covariance matrix $(B\psi B')$ produces standard errors of the regression coefficients that are larger than those given by the incorrect "standard" variance-covariance matrix $\sigma_{11}B$. Thus

[7] Note that this requires that X_{2i} contain nontrivial regressors or that there be no intercept in the equation, or both.

the usual procedure for estimating standard errors, which would be correct if λ_i were known, leads to an understatement of true standard errors and an overstatement of significance levels when λ_i is estimated and $C \neq 0$.

Under the Amemiya-Jennrich conditions previously cited, ψ is a bounded positive definite matrix. ψ and B can be simply estimated. Estimated values of λ_i, C, and σ_{11} can be used in place of actual values to obtain a consistent estimator of $B\psi B'$. Estimation of the variance-covariance matrix requires inversion of a $K_1+1\times K_1+1$ matrix and so is computationally simple. A copy of a program that estimates the probit function coefficients β_2^* and the regression coefficients $\hat{\beta}_1$ and $\hat{C}$, and produces the correct asymptotic standard errors for the general case is available on request from the author.[8]

It is possible to develop a GLS procedure (see Heckman [**7**]). This procedure is computationally more expensive and, since the GLS estimates are not asymptotically efficient, is not recommended.

The estimation method discussed in this paper has already been put to use. There is accumulating evidence [**3** and **6**] that the estimator provides good starting values for maximum likelihood estimation routines in the sense that it provides estimates quite close to the maximum likelihood estimates. Given its simplicity and flexibility, the procedure outlined in this paper is recommended for exploratory empirical work.

3. SUMMARY

In this paper the bias that results from using nonrandomly selected samples to estimate behavioral relationships is discussed within the specification error framework of Griliches [**2**] and Theil [**12**]. A computationally tractable technique is discussed that enables analysts to use simple regression techniques to estimate behavioral functions free of selection bias in the case of a censored sample. Asymptotic properties of the estimator are developed.

An alternative simple estimator that is also applicable to the case of truncated samples has been developed by Amemiya [**1**]. A comparison between his estimator and the one discussed here would be of great value, but is beyond the scope of this paper. A multivariate extension of the analysis of my 1976 paper has been performed in a valuable paper by Hanoch [**5**]. The simple estimator developed here can be used in a variety of statistical models for truncation, sample selection and limited dependent variables, as well as in simultaneous equation models with dummy endogenous variables (Heckman [**6**, **8**]).

University of Chicago

Manuscript received March, 1977; final revision received July, 1978.

[8] This offer expires two years after the publication of this paper. The program will be provided at cost.

REFERENCES

[1] AMEMIYA, T.: "Regression Analysis when the Dependent Variable is Truncated Normal," *Econometrica*, 41 (1973), 997–1017.
[2] GRILICHES, ZVI: "Specification Bias in Estimates of Production Functions," *Journal of Farm Economics*, 39 (1957), 8–20.
[3] GRILICHES, Z, B. HALL, AND J. HAUSMAN: "Missing Data and Self Selection in Large Panels," Harvard University, July, 1977.
[4] GRONAU, R.: "Wage Comparisons—A Selectivity Bias," *Journal of Political Economy*, 82 (1974), 1119–1144.
[5] HANOCH, G.: "A Multivariate Model of Labor Supply: Methodology for Estimation," Rand Corporation Paper R-1980, September, 1976.
[6] HECKMAN, J.: "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *The Annals of Economic and Social Measurement*, 5 (1976), 475–492.
[7] ———: "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions," NBER Working Paper #172, March, 1977 (revised).
[8] ———: "Dummy Endogenous Variables in a Simultaneous Equation System," April, 1977 (revised), *Econometrica*, 46 (1978), 931–961.
[9] JENNRICH, R.: "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40 (1969), 633–643.
[10] JOHNSON, N., AND S. KOTZ: *Distribution in Statistics*: *Continuous Multivariate Distributions.* New York: John Wiley & Sons, 1972.
[11] LEWIS, H.: "Comments on Selectivity Biases in Wage Comparisons," *Journal of Political Economy*, 82 (1974), 1145–1155.
[12] THEIL, H.: "Specification Errors and the Estimation of Economic Relationships," *Revue de l'Institut International de Statistique*, 25 (1957), 41–51.