

NBER Working Paper Series

SAMPLE SELECTION BIAS AS A  
SPECIFICATION ERROR

(with an application to the estimation  
of labor supply functions)

James J. Heckman

Working Paper No. 172

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR  
AND SOCIAL INSTITUTIONS

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

March 1977  
(revised)

Preliminary; not for quotation.

NBER working papers are distributed informally and in limited number for comments only. They should not be quoted without written permission of the author.

This report has not undergone the review accorded official NBER publications; in particular, it has not yet been submitted for approval by the Board of Directors.

The research reported in this paper was supported by an 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. I have greatly benefited from conversations with Paul Schultz and James Smith, and from detailed written comments of Mark Killingsworth on the first draft of this paper. Takeshi Amemiya, John Cogan, Reuben Gronau, Ed Leamer, H. G. Lewis, and Bill Rodgers all made valuable comments on the second draft. Gary Becker, Tom MaCurdy and Arnold Zellner made valuable comments on this draft. Ralph Shnelvar performed the computations. I assume all responsibility for any remaining errors.

In this paper, the bias that results from using nonrandomly selected samples to estimate behavioral relationships is shown to arise because of a missing data problem. In contrast with the standard omitted variable problem in econometrics, in which certain explanatory variables of a regression model are missing, the problem of sample selection bias arises because data are missing on the dependent variable of an analysis. Regressions estimated on the data available from the nonrandom sample will not, in general, enable the analyst to estimate parameters of direct interest to economists. Instead, such regression coefficients confound meaningful structural parameters with the parameters of the function determining the probability that an observation makes its way into the nonrandom sample.

Sample selection bias may arise for two distinct reasons. First, there may be self selection by the individuals being investigated. One observes market wages for certain women because their productivity in the market exceeds their productivity in the home.<sup>1</sup> Similarly, one observes wages for union members who found their nonunion alternative less desirable. Finally, the wages of migrants (or manpower trainees) do not, in general, afford an estimate of what nonmigrants (nontrainees) would have earned had they migrated (participated in training). In each of these cases, wage functions fit on the available data do not estimate the wage function that characterizes a randomly selected member of the general population subject to the "treatment" of work, unionism, migration, and manpower training programs, respectively.

---

<sup>1</sup>Note that this does not imply that the more market productive women are the ones observed working.

Moreover, in each case, if it were possible to obtain the missing wage data for either the treatment or nontreatment population, it would be possible to utilize simple regression techniques to estimate the parameters of population functions. Simple comparisons between pre and post treatment wages would yield unbiased estimates of the economic benefits of the treatment.

Sample selection bias may also arise as a direct consequence of actions taken by the analyst. In studies of panel data, it is common to require that "intact" observations be employed. For example, in analyses of the time series of the labor supply of married women, stability of the family unit is often required for an observation sequence to be analyzed. The effect of such criteria operates in precisely the same fashion as self selection: fitted functions confound behavioral functions with sample selection functions.

It is fair to say that most competent analysts have been aware of the possibility of both sources of selection bias. It is also fair to say that the accepted econometric practice has been to ignore the problem in making parameter estimates but to verbally qualify the estimates in light of possible selection biases.

Recent work in econometrics has attempted to improve on previous work by making specific assumptions about the source of selection bias. In particular, this work assumes that both the missing data and the available data are drawn from a common probability distribution, typically assumed to be a normal law. Except for work by Amemiya (1973) and Gronau (1974), the authors of these studies rely on maximum likelihood estimators to produce parameter estimates free of selection bias.

In this paper, I present a simple characterization of the sample selection bias problem that is also applicable to the conceptually distinct

econometric problems that arise from truncated samples and from models with limited dependent variables.<sup>1</sup> The problem of sample selection bias is fit within the conventional specification error framework of Griliches and Theil. A simple estimator is discussed that enables analysts to utilize ordinary regression methods to estimate models free of selection bias.

The techniques discussed here are applied to reestimate and test a model of female labor supply developed by the author (1974). Besides providing an illustration of the methodology, this application is of interest in its own right for three reasons: (a) an important variable utilized in the author's previous analysis, the labor market experience of women, was incorrectly coded by the primary data source, (b) the simple estimators discussed here allow for much more extensive testing of the maintained hypotheses of the previous paper, (c) the method discussed here produces an initial consistent estimator for the likelihood equations of the previous paper. This last issue is important because the likelihood function proposed in the 1974 paper is not globally concave and hence the issue of selecting an initial starting value is an important one, since local optima will not yield consistent estimators.

Four conclusions emerge from the empirical analysis of female labor supply that is conducted on the 1967 National Longitudinal Survey for women age 30-44. First, estimated coefficients of labor supply and wage functions are quite sensitive to alternative treatment of the labor market experience of the wife. Recent work (Heckman, 1977) suggests that unmeasured factors that determine participation also determine past work behavior. Treating the wife's labor force experience as an endogenous variable in participation probabilities, using standard instrumental variable estimation techniques,

---

<sup>1</sup>This relationship is spelled out in greater detail in a companion paper (Heckman, 1976).

significantly alters the coefficients of estimated labor supply and wage functions. Second, in a model that treats the labor market experience of the wife as endogenous, there is evidence that selection bias is an important phenomenon in the estimates of labor supply functions, but there is little evidence of selection bias in estimates of the hourly wage function. Third, the empirical analysis casts some doubt on the validity of the simple model assumed in the 1974 paper. With a minor modification the basic structure of the model remains intact and concordant with data. Fourth, conventional measures of labor supply overstate the amount of measured work, create the statistical illusion of a standard work week and work year, and considerably understate the true sample variation in labor supply.

This paper is in three parts. In the first section, selection bias is presented within the specification error framework. In this section, general distributional assumptions are maintained. In section two, specific results are presented for the case of normal regression disturbances. Simple estimators are proposed and discussed. In the third section, empirical results are presented.

### I. Sample Selection Bias as a Specification Error

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  $T$  observations. The equations for individual  $i$  may be written as

$$(1a) \quad Y_{1i} = X_{1i}\beta_1 + U_{1i}$$

$$(1b) \quad Y_{2i} = X_{2i}\beta_2 + U_{2i}$$

exogenous

where  $X_{j1}$  is a  $1 \times K_j$  vector of regressors,  $\beta_j$  is a  $K_j \times 1$  vector of parameters,

$$(2) \quad E(U_{j1}) = 0, \quad E(U_{j1}U_{j'1}) = \sigma_{jj'}, \quad j=1,2,$$

$$E(U_{j1}U_{j'1'}) = 0, \quad i \neq i'.$$

The final assumption is an implication of a random sampling scheme. Denote the joint distribution of  $U_{11}, U_{21}$  by  $h(U_{11}, U_{21})$  which may be a singular distribution. The regressor matrix is assumed to be of full rank so that if all data were available, each equation could be estimated by least squares, and all parameters would be identified.

Suppose that one seeks estimates of equation (1a) but that data are missing on  $Y_{11}$  for certain observations. The crucial question is "why are data missing for certain observations?"

No matter what the answer to this question, one can write the population regression function for equation (1a) as

$$E(Y_{11}|X_{11}) = X_{11}\beta_1, \quad i = 1, \dots, T$$

while the regression function for the subsample of available data is

$$E(Y_{11}|X_{11}, \text{sample selection rule}) = X_{11}\beta_1 + E(U_{11}|\text{sample selection rule}),$$

$i = 1, \dots, T_1$  where, for convenience, the  $i$  subscripts are labeled so that the first  $T_1 < T$  observations have all data available. If the conditional expectation of  $U_{11}$  is zero, the selected sample regression function is the same as the population regression function. In this case, least squares may be applied to the subsample of the available data to estimate the population regression function. The only cost of having an incomplete sample is a loss in efficiency.

In the general case, the sample selection rule that determines the available data has more serious consequences. Consider the following selection rule: data are available on  $Y_{1i}$  if

$$(3) \quad Y_{2i} \geq 0$$

while if

$$Y_{2i} < 0$$

we do not obtain observations. Clearly the choice of zero as a threshold is an inessential normalization. Also, one could define a dummy variable  $d_i$  with the properties

$$(4) \quad d_i = 1 \text{ iff } Y_{2i} \geq 0$$

$$d_i = 0 \text{ iff } Y_{2i} < 0$$

so that one could analyze the joint distribution of  $Y_{1i}$ ,  $d_i$  dispensing with  $Y_{2i}$  altogether. The advantage in using the selection rule representation (3) is that it permits a unified summary of the existing literature. Utilizing this representation, one may write

$$\begin{aligned} E(U_{1i} | \text{sample selection rule}) &= E(U_{1i} | Y_{2i} \geq 0) \\ &= E(U_{1i} | U_{2i} > -X_{2i}\beta_2). \end{aligned}$$

In the case of independence between  $U_{1i}$  and  $U_{2i}$ , so that the selection rule is independent of the behavioral function being estimated, the conditional mean of  $U_{1i}$  is zero.

In general, the conditional mean of the  $U_{1i}$  disturbance does not vanish. Accordingly, the selected sample regression function may be written as

$$(5) \quad E(Y_{1i} | X_{1i}, Y_{2i} > 0) = X_{1i} \beta_1 + E(U_{1i} | U_{2i} > -X_{2i} \beta_2).$$

The selected sample regression function depends on  $X_{1i}$  and  $X_{2i}$ . Regression estimates of equation (1a) fit on the selected sample omit the final term of equation (5). Thus the problem of sample selection bias, initially viewed as a missing dependent variable problem may be reformulated as an ordinary omitted explanatory variable problem.

Several special cases of this model are of interest. First assume that the only variable in the regressor vector  $X_{2i}$  is the constant "1". In this case, the probability that an observation is included in the sample is the same for all observations and is not a function of any explanatory variables. The conditional mean of  $U_{1i}$  is a constant. Ordinary least squares estimators of equation (1a) yield unbiased estimators for slope coefficients but a biased estimator for the intercept, and the population variance  $\sigma_{11}$ . The same analysis applies to a more general model with  $X_{2i}$  regressors as long as the set of  $X_{1i}$  variables is uncorrelated with the conditional mean of  $U_{1i}$ . In particular, if  $X_{1i}$  and  $X_{2i}$  are independent random variables, this analysis continues to hold.

In the general case with nontrivial regressors included among the  $X_{2i}$  variables it is unreasonable to expect that the regressors of equation (1a) (i.e.,  $X_{1i}$ ) are uncorrelated with the conditional mean of  $U_{1i}$ . Accordingly, least squares estimators of the slope coefficients ( $\beta_1$ ) are biased. Without further assumptions about the distribution of  $U_{1i}$ , it is not possible to sign the bias. If the conditional mean of the  $U_{1i}$  disturbance is well approximated by the linear terms of a Taylor's series expansion, this approximation may be substituted in equation (5) and an ordinary specification error analysis may be performed.



From equation (5), it is evident that a symptom of selection bias is that variables that do not belong in the true structural equation (e.g., elements of  $X_{2i}$  not in  $X_{1i}$ ) may appear to be statistically significant determinants of  $Y_{1i}$  when regressions are fit on selected samples. For example, in Gronau's analysis of the selection bias that arises in using the wages of working women to estimate the potential wage of nonworking women, variables that affect the probability that a woman works, such as the presence of children, may appear to affect market wages when, in fact, no causal association exists. Thus regression evidence that women with children earn lower wages is not necessarily evidence that there is discrimination against such women or that women with lower market experience—as proxied by children—earn lower wages. Evidence that such extraneous variables "explain" wage rates may be interpreted as evidence in support of the selection hypothesis. However, even if no such extraneous variables appear in the selected sample regressions, estimates of the intercept and the population variance may be biased.

If one knew the conditional mean of  $U_{1i}$  or could estimate it, one could enter it as a regressor in equation (5) and use ordinary least squares to estimate the  $\beta_1$  parameters. In the next section, I discuss a method for estimating the conditional mean for the case of jointly normal disturbances. Before turning to this discussion, it is helpful to relate the simple model presented here to previous work in the literature.

The justly celebrated model of Tobin (1958) may be fit within this framework. (See also Amemiya, 1973.) In Tobin's model, data are missing on  $Y_{1i}$  if  $Y_{1i} < 0$ . Setting  $Y_{1i} \equiv Y_{2i}$ ,  $\beta_1 \equiv \beta_2$ ,  $X_{1i} \equiv X_{2i}$ , and  $U_{1i} \equiv U_{2i}$ , the "Tobit" model arises.<sup>1</sup> The bivariate density  $h(U_{1i}, U_{2i})$  becomes degenerate

<sup>1</sup>Tobin assumed a normal density of  $U_{1i}$ . The conceptual logic of his model does not rely on normality.

since  $U_{1i} = U_{2i}$ . Since  $X_{1i}$  and  $X_{2i}$  are identical, the conditional mean of  $U_{1i}$  is not orthogonal to  $X_{1i}$  and bias is guaranteed for the least squares estimators of equation (1a) applied to selected samples.

Tobin's model was a major stimulus to later work. Its simplicity and elegance mask two important ideas that have been confused. Most economists have interpreted his model as a prototype of a limited dependent variable model: the range of observed values of the random variable  $Y_{1i}$  cannot fall short of zero. Putting Tobin's model this way, it is less interesting. Most economists are willing to live with this type of truncation of the range of a variable and simple transformations can eliminate it (e.g., use of logs).

The important feature of Tobin's model is that a selection rule ( $Y_{1i} \geq 0$ ) generates the sample of observed data. Both Cragg (1971) and Nelson (1975) note that the selection rule generating observations on  $Y_{1i}$  need not be as closely related to the population regression function as Tobin assumed. Their models may be fit within the schema of equations (1) and (3).

For example, consider Nelson's model.  $Y_{1i}$  is observed if

$$Y_{1i} > Z_{2i}$$

where  $Z_{2i}$  is a random variable. In terms of the notation of equation (1), his model becomes  $Y_{2i} = Y_{1i} - Z_{2i}$ ,  $\beta_2 = 0$ . If  $Y_{2i} \geq 0$ ,  $Y_{1i}$  is observed while if  $Y_{2i} < 0$ ,  $Y_{1i}$  is observed to be zero.<sup>1</sup>

Elsewhere, I present a model that can be fit within the sample selection framework (Heckman, 1974). This model will be elaborated in Section III along with the closely related models of Gronau (1974) and Lewis (1974).

---

<sup>1</sup>Note, however, that  $Y_{1i}$  is not, strictly speaking, a limited dependent variable since nothing prevents  $Y_{1i}$  from becoming negative.

I note, in passing, that multivariate extensions of the preceding models, while mathematically straightforward, may be of considerable substantive interest. Two examples are offered. One concerns migrants choosing among  $K$  prospective regions. Each person can be viewed as possessing  $K$  distinct wage functions, one for each region. If the self selection rule is to choose that region with the highest income, both the selection rule and the subsample regression functions can be simply characterized by an obvious  $K + 1$  variable extension of the previous analysis. The second example concerns the measurement of union-nonunion wage differentials. Each person in a hypothetical population can be viewed as possessing both a union and a nonunion wage function. One self selection rule, based on the assumption of freedom of entry into unionism, is to select the unionism status with the highest wage. Estimators of wage functions based on pooled union and nonunion samples yield biased estimates of the economic return to unionism if selection into unionism status is nonrandom.

Before concluding this section, it is useful to clarify three concepts that are frequently confused in the literature. The first is the concept of a truncated variable. The second is the concept of a truncated sample. The third is the concept of a censored sample.

A sample is said to be censored when it is possible to use sample evidence to estimate the probability that a hypothetical observation will be observed. This is the situation assumed in the model of equations (1) and (3). A truncated sample differs from a censored sample because the probability of sample selection cannot be estimated from observed data. A random variable is said to be truncated when its range is limited. Clearly, random variables can be truncated in either censored or truncated samples. Also, quite clearly, the operational distinction between a censored and

truncated sample vanishes if there is a priori information about the probability of sample selection for a hypothetical observation. These categories often overlap. Thus in Tobin's model the sample is censored but the random variable is truncated.

## II. Simple Estimators for the Case of Normally Distributed $U_{1i}$ and $U_{2i}$

In this section, the model of equations (1), (3) and (4) is derived for the specific case of joint normality for  $U_{1i}$  and  $U_{2i}$ . The normality assumption is used in the models surveyed in Section I and is a natural starting point for any analysis. A simple estimator for this normal model is derived and discussed.

The joint distribution of  $U_{1i}, U_{2i}, h(U_{1i}, U_{2i})$ , is a bivariate normal density fully characterized by the assumptions stated in equation (2). It is permitted to be singular as in Tobin's model. Using well known results in the literature, (see Johnson and Kotz (1972)), pp. 112-113 or Gronau (1974), Lewis (1974)

$$E(U_{1i} | U_{2i} > -X_{2i}\beta_2) = \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda_i$$

$$E(U_{2i} | U_{2i} > -X_{2i}\beta_2) = \frac{\sigma_{22}}{(\sigma_{22})^{1/2}} \lambda_i$$

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

where  $\phi$  and  $\Phi$  are, respectively, the density and distribution function for the standard normal random variable and

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

" $\lambda_1$ " is the inverse of Mills' ratio, and is the ratio of the standard normal to the tail area of the distribution. There are several important features of  $\lambda_1$ . (1) Its denominator is the probability that a population observation with characteristics  $X_{2i}$  is selected into the observed sample. (2)  $\lambda(Z)$  is a monotone increasing function of  $Z$  and hence is a monotone decreasing function of the probability of sample selection  $\phi(Z)$ . In particular,  $\lim_{Z_1 \rightarrow -\infty} \lambda_1 = 0$ ,  $\lim_{Z_1 \rightarrow \infty} \lambda_1 \rightarrow \infty$ , and  $\frac{\partial \lambda_1}{\partial Z_1} > 0$ . Figure 1 displays the relationship between  $\lambda$  and  $\phi$ . In samples

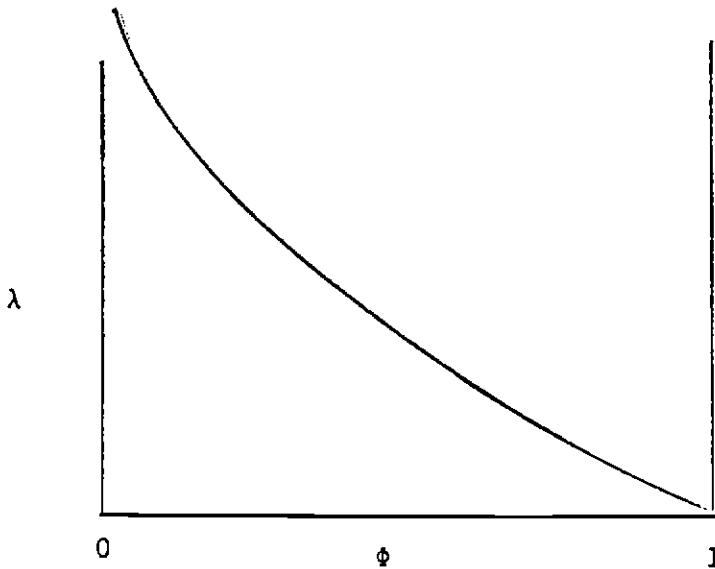


Fig. 1--Probability of sample inclusion.

in which the sample selection rule guarantees that all population observations have an equal chance of being sampled,  $\lambda(Z)$  is zero and the least squares estimator of equation (1a) has optimal properties.

Using these results, equation (5) becomes

$$(6a) \quad E(Y_{1i} | X_{1i}, Y_{2i} \geq 0) = X_{1i}\beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda(Z_1)$$

while the comparable expression for  $Y_{2i}$  becomes

$$(6b) \quad E(Y_{2i} | X_{2i}, Y_{2i} \geq 0) = X_{2i} \beta_2 + \frac{\sigma_{22}}{(\sigma_{22})^{1/2}} \lambda(Z_i)$$

If one could estimate  $Z_i$  and hence estimate  $\lambda_i$ , one could enter the latter variable as a regressor in equation (6a) and estimate  $\beta_1$  and  $\frac{\sigma_{12}}{(\sigma_{22})^{1/2}}$  by least squares. Similarly, if one could measure  $Y_{2i}$  when  $Y_{2i} \geq 0$ , as in Tobin's model, knowledge of  $Y_{2i}$  and  $\lambda_i$  would permit direct estimation of  $\beta_2$  and  $(\sigma_{22})^{1/2}$ . Representation (6a) reveals that if  $\sigma_{12} = 0$ , so that the disturbances that affect sample selection are independent of the disturbances affecting the behavioral functions of interest,  $\lambda_i$  may be omitted as a regressor. Thus, if either  $\lambda$  or  $\sigma_{12}$  is zero, or both, least squares estimators of  $\beta_1$  are unbiased.

The full statistical model of which equations (6a) and (6b) are expectations is now developed. One may write the model as

$$(7a) \quad Y_{1i} = X_{1i} \beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda_i + v_{1i}$$

$$(7b) \quad Y_{2i} = X_{2i} \beta_2 + \frac{\sigma_{22}}{(\sigma_{22})^{1/2}} \lambda_i + v_{2i}$$

where

$$E(v_{1i} | X_{1i}, \lambda_i, U_{2i} > -X_{2i} \beta_2) = 0$$

$$E(v_{2i} | X_{2i}, \lambda_i, U_{2i} > -X_{2i} \beta_2) = 0$$

and

$$E(v_{ji} v_{ki} | X_{1i}, X_{2i}, \lambda_i, U_{2i} > -X_{2i} \beta_2) = 0$$

for  $i \neq i'$ . It is straightforward to demonstrate that

$$(8a) \quad E(v_{1i}^2 | X_{1i}, \lambda_i, U_{2i} > -X_{2i} \beta_2) = \sigma_{11} ((1 - \rho^2) + \rho^2 (1 + Z_i \lambda_i - \lambda_i^2))$$

$$(8b) \quad E(v_{1i} v_{2i} | X_{1i}, X_{2i}, \lambda_i, U_{2i} > -X_{2i} \beta_2) = \sigma_{12} (1 + Z_i \lambda_i - \lambda_i^2)$$

$$(8c) \quad E(v_{2i}^2 | X_{2i}, \lambda_i, U_{2i} > -X_{2i} \beta_2) = \sigma_{22} (1 + Z_i \lambda_i - \lambda_i^2)$$

where

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

Moreover, one can easily establish that

$$(9) \quad 0 \leq (1 + \lambda_1 Z_1 - \lambda_1^2) \leq 1.$$

There are several important consequences of this inequality for the covariance structure of the disturbances of equations (7a, b). Suppose that one knows  $Z_1$  and  $\lambda_1$ , and enters  $\lambda_1$  as a regressor in equations (7). Standard least squares estimators of the population variance of  $\sigma_{11}$  and  $\sigma_{22}$  are downward biased estimators of the appropriate parameters. Also, the standard estimator of the interequation covariance is downward biased in absolute value. Note further that if  $Z_1$  contains regressors (apart from "1") the variances of the disturbances of equations 7 are heteroskedastic. Least squares/are not GLS estimators. The GLS estimators have an interesting interpretation. Unlikely observations (those with low probability of sample inclusion) receive greater weight than likely observations. This follows because the middle term in inequality (9) is a monotonically increasing function of the probability of sample inclusion,  $\phi(Z)$ .<sup>1</sup> Accordingly, less likely observations receive greater weight and observations with zero probability of sample inclusion receive the greatest weight.

The GLS estimators based on known  $\lambda_1$  possess unusual properties, not fully developed here. In contrast with the usual case for GLS estimators, parameters of the regression function enter the disturbance variance. This is seen most clearly in equations (7a) and (8a). Using the definition of  $\rho$ , presented below the equations (8a-8c), the coefficient on the  $\lambda_1$  variable in equation (7a) may be rewritten as  $\rho\sigma_{11}^{1/2}$  so that the dependence is explicit.

<sup>1</sup>An elementary application of L'Hospital's rule reveals that in the limit, as  $Z_1 \rightarrow -\infty$ ,  $\lambda_1 \rightarrow \infty$ , and  $\lim E(V_{21}^2) \rightarrow 0$ . Similarly,  $\lim E(V_{11}V_{21}) = 0$  and  $\lim_{Z_1 \rightarrow -\infty} E(V_{11}^2) = (1 - \rho^2)\sigma_{11}$ .

The consequences of this dependence are interesting although their full development is peripheral to this paper. Nonetheless, a brief outline may be of some interest.

With known  $\lambda_1$ , one may use least squares to produce unbiased estimates of the regression parameters of equation (7a):  $\beta_1$  and  $\rho\sigma_{11}^{1/2}$ . Using the least squares residuals from equation (7a), one may form a consistent estimator of  $\sigma_{11}$ .<sup>1</sup> Then, an approximate GLS estimator that converges in distribution to the true GLS estimator may be found by estimating equation (7a) by weighted least squares with the estimated weights obtained from equation (8a). An important feature of this problem is that one/<sup>round</sup>GLS estimators are not asymptotically efficient compared to the appropriate likelihood function estimates which are based on a truncated bivariate normal distribution with known points of truncations.<sup>2</sup>

The preceding analysis appears to be somewhat beside the point since as a practical matter one does not know  $Z_1$  and  $\lambda_1$  and hence one cannot directly estimate equation (7a). But in the case of a censored sample, it is possible to compute the probability that an observation has missing data so that it is possible to use probit analysis to estimate  $Z_1$  and hence  $\lambda_1$ .

In the case of a truncated sample this is not so. However, if prior information is available on the probability that  $Y_{1i}$  is observed it is possible

---

<sup>1</sup>Denote the residuals by  $\hat{V}_{1i}$ . Since  $\lambda_1$  and  $Z_1$  are known, and since  $\rho\sigma_{11}^{1/2}$  is estimated, one can estimate

$$\sigma_{11} = \frac{\sum \hat{V}_{1i}^2}{T} - \frac{(\rho^2 \sigma_{11})}{T} \sum (\lambda_1 Z_{1i} - \lambda_1^2).$$

This yields a consistent estimate of the variance that is guaranteed to be positive. Note, however, that nothing in the procedure guarantees a value of  $|\rho|$  inside the unit interval although in large samples it must lie in that interval.

<sup>2</sup>This is so because  $\rho$  and  $\sigma_{11}$  appear in the regression coefficients and in the variance so the information matrix is not block diagonal. An iterative estimator based on the initial consistent estimates previously discussed is asymptotically efficient



to estimate  $Z_i$  and  $\lambda_i$  so that prior information on the probability of sample inclusion eliminates the distinction between censored and truncated samples.

In the censored case, the probit likelihood function is

$$(10) \quad \mathcal{L} = \prod_{i=1}^T [\phi(Z_i)]^{d_i} [\phi(-Z_i)]^{1-d_i}$$

where  $d_i$  denotes the event "observation of  $Y_{1i}$ ." Under the standard conditions for identification in probit analysis (see Nerlove and Press, 1976), one may consistently estimate  $\beta_2/\sigma_{22}^{1/2}$ , and hence  $Z_i$  and  $\lambda_i$ . The estimated  $\lambda_i$  may be substituted for the actual  $\lambda_i$  in the preceding analysis.

In Appendix A, the asymptotic distribution is derived for the least squares estimator based on an estimated  $\lambda_i$  instead of the actual  $\lambda_i$ . The least squares estimators are consistent and asymptotically normally distributed. Moreover, in the important special case of the null hypothesis of no selection bias (e.g.,  $\sigma_{12} = 0$  in equation (7a)), the standard least squares estimator of the variance-covariance matrix of the regression coefficients is the appropriate estimator. However, if  $\sigma_{12} \neq 0$ , the standard estimator is inappropriate and the formula (A4) in Appendix A should be used instead.

As in the case of exact GLS estimators based on known values of  $\lambda_i$ , approximate GLS estimators are not asymptotically efficient nor do they converge in distribution to GLS estimators based on known  $\lambda_i$  except in the important special case of a null hypothesis of no selection bias. To achieve asymptotically efficient estimators, maximum likelihood estimators must be employed. The estimators suggested here provide initial consistent estimators for the likelihood equations so that a one step iteration (Rothenberg and Leenders, 1964) yields estimates that are asymptotically efficient. Thus the task of computing efficient estimates is simplified and the problem of locating a starting value for likelihood

function iterations is resolved. Elsewhere (Heckman, 1976) it is shown that for one problem the initial consistent estimators discussed here closely approximate the likelihood maximizing parameter estimates.

III. New Estimates of Female Labor Supply Functions  
and Wage Functions Free of Selection Bias:  
New Tests of an Old Model

A. The Model

In this section, the techniques of Section II are applied to estimate the labor supply and wage functions of married women. In the absence of fixed costs of entering and exiting the labor market, and under the assumption that workers are free to choose their hours of work, two functions fully characterize the labor supply decision.

The first function is the market wage function for the woman,  $Y_{1i}$ , defined by equation (1a). The second function is the reservation wage that records the value that a woman places on her time if she does not work ( $W_i^*$ ). If the market wage exceeds the reservation wage ( $Y_{1i} > W_i^*$ ), a woman works and her hours of work adjust so that in equilibrium the marginal value of her time equals her market wage rate.

Under certain simplifying assumptions elaborated <sup>more fully</sup> elsewhere (Heckman, 1974a), hours of work,  $h_i$ , are proportional to the gap between market wages and reservation wages. Denoting this proportionality factor by  $1/\gamma$ , and letting  $Y_{2i} = Y_{1i} - W_i^*$  be the gap, one is led to the following model:

$$(11a) \quad E(Y_{1i} | X_{1i}, Y_{2i} > 0) = X_{1i}\beta_1 + E(U_{1i} | U_{1i} > -X_{2i}\beta_2)$$

$$(11b) \quad E(h_i | X_{1i}, X_{2i}, Y_{2i} > 0) = \frac{1}{\gamma} E(Y_{2i} | X_{1i}, X_{2i}, Y_{2i} > 0) =$$

$$\frac{X_{2i}\beta_2}{\gamma} + \frac{1}{\gamma} E(U_{2i} | U_{2i} > -X_{2i}\beta_2)$$

$$(11c) \quad \gamma > 0$$

This model differs from the sample selection model of equations (1a) and (1b) in one important respect. Unlike the case in equation (1b) there is information about  $Y_{2i}$  up to a factor of proportionality ( $1/\gamma$ ) if a woman works ( $Y_{2i} \geq 0$ ). The decision function that characterizes labor force entry, which is the sample selection rule for this model, is closely related to the hours of work equation. The model of Lewis (1974) and Gronau (1974) is exactly the model of equations (1a) and (1b) and does not utilize the potential source of information that closely links the participation decision and the labor supply function.

From inspection of equations (11a) and (11b), it is clear that both wage and hours of work functions may be subject to selection bias. Least squares estimators of the wage and hours functions fit for working women confound the parameters of the sample selection function with the parameters of the behavioral functions of interest.

This is not to say that estimates of wage or labor supply functions fit on subsamples of working women are of no interest. A regression model that deletes the conditional expectation of the error terms approximates a function with a well defined interpretation. Consider equation (11b). The same set of variables ( $X_2$ ) appears in the regression function and in the conditional mean of  $U_2$ . If one deletes the conditional mean, to a first order approximation a regression equation estimates the vector

$$\frac{\hat{\beta}_2}{\gamma} = \frac{\beta_2}{\gamma} + \frac{1}{\gamma} \frac{\partial E(U_{2i} | U_{2i} > -X_{2i}\beta_2)}{\partial X_{2i}} .$$

Thus ordinary least squares coefficients estimate the effect of a variable moving along the behavioral function, the first term, and the effect of the variable in sorting people out in the taste distribution, the second term.

To clarify this decomposition, a concrete example may be helpful.

Let vector  $X_{2i}$  consist of one variable--say, ability to perform market tasks. Ability is expected to increase the supply of hours to the market for a working woman ( $\beta_2/\gamma > 0$ ). Moreover, ability is expected to increase the probability that a woman works. But this means that as one samples across working women with greater ability one is sampling women with progressively lower average tastes for work  $\left( \frac{\partial E(U_{2i} | U_{2i} > -X_{2i}\beta_2)}{\partial X_{2i}} < 0 \right)$ . Thus the regression coefficient on the ability variable is a downward biased estimate of  $\beta_2/\gamma$ .

Estimates of  $\beta_2/\gamma$  answer the question "what is the average effect of an additional unit of ability on the labor supply of women already working?" Economic theory provides a guide to the sign and magnitude of this coefficient. Estimates of  $\hat{\beta}_2/\gamma$  answer the question "what is the change in the average labor supply of women when one moves across ability groups?" These estimates give the basic ingredients required to estimate the aggregate labor supply curve. Given a distribution of ability in the population, one can add up the average labor supply at each ability class to compute aggregate labor supply.<sup>1</sup> Typically, economic theory does not directly yield predictions about this parameter which combines parameters describing movements along a given labor supply function with the parameters determining the entry of workers into the labor force.

The parameter  $1/\gamma$  plays a crucial role in this analysis, and may be interpreted as the uncompensated effect of a change in wage rates on labor supply. From equation (11b) it is not clear how this parameter may be estimated. Recall that  $Y_{2i}$  is defined as the difference between market wages and reservation wages ( $Y_{1i} - W_1^*$ ). To demonstrate how  $\gamma$  can be estimated one may introduce an explicit function for  $W_1^*$

$$W_1^* = N_1\psi + \epsilon$$

<sup>1</sup>More precisely, equation (11b) multiplied by the probability that a woman works, yields an estimate of the average hours supplied to the market by women with traits  $X_{1i}$ ,  $X_{2i}$ .

and note that

$$(12) \quad h_i = \frac{Y_{2i}}{\gamma} \equiv \frac{Y_{1i} - W_i^*}{\gamma} = \frac{X_{1i}\beta_1 - N_i\psi}{\gamma} + \frac{1}{\gamma} (U_{1i} - \epsilon_i).$$

Then it is clear that if one variable appears in  $X_{1i}$  that does not appear in  $N_i$ , such as the market human capital of the wife, given estimates of equation (11a) one can estimate both  $\gamma$  and  $\psi$ . Note, too, that one can follow conventions in simultaneous equation theory to avoid the multiplicity of estimates of  $\gamma$  that arise in the overidentified case if one inserts estimates of  $\beta_1$  (obtained from (11a)) to generate a predicted value of wage rates in (12), i.e., denoting such estimates by  $\hat{\beta}_1$ , estimating

$$\frac{Y_{2i}}{\gamma} = \frac{X_{1i}\hat{\beta}_1 - N_i\psi}{\gamma} + \frac{1}{\gamma} (U_{1i} - \epsilon_i) + \frac{1}{\gamma} X_{1i}(\beta_1 - \hat{\beta}_1)$$

where the final term vanishes in large samples.

The crucial feature of labor supply function (12) is that the supply of labor is assumed to be a function of the gap between market wages and reservation wages. This gap, in turn, is a measure of the probability that a woman works. Thus a strong assumption of this formulation is that a woman more likely to work is also more likely to supply more labor when she works.

Over the empirically relevant range, the labor supply curve may become backward bending ( $\gamma < 0$ ). Moreover, as noted by Cogan (1975) and Hanoch (1976), fixed costs of entry and exit may alter the simple relationship of equations (11b) and (12), and may even result in opposite signs for the effect of certain variables on labor supply and participation.

As an example, consider the effect of money costs of child care. Holding everything else the same, the greater the number of preschool children, the greater the cost of child care and hence the less likely is the event that

a woman works. However, given that a woman works, greater expenditure on child care results in a reduction of income and hence an expansion in hours worked if leisure time is a normal good. Time indivisibilities in the availability of child care, and commutation costs tend to reinforce the work increasing effect of child care costs.<sup>1</sup>

It is straightforward to extend the model of equations (11a) - (11c) to allow for these effects. A more general model is the three equation system

$$(13a) \quad Y_{1i} = X_{1i}\beta_1 + U_{1i}$$

$$(13b) \quad Y_{2i} = X_{2i}\beta_2 + U_{2i}$$

$$(13c) \quad h_i = Y_{3i} = X_{2i}\beta_3 + U_{3i},$$

$$E(U_{ji}) = 0, \quad E(U_{ji}U_{j'i'}) = \sigma_{jj'}, \quad i = i'$$

= 0 otherwise.

As before, a woman works, and her hours are positive if and only if  $Y_{2i} > 0$ . A noteworthy feature of equations (13b) and (13c) is that the same set of variables determine the participation decision and the quantity of hours supplied. Under the null hypothesis that equation (11b) is correct,  $\beta_2$  and  $\beta_3$  are equal up to a constant of proportionality ( $\beta_2 = \frac{\beta_3}{\gamma}$ ) and the joint distribution of the  $U_{ji}$  is a singular trivariate density. Assuming normality for

<sup>1</sup>Obviously, time costs decrease leisure consumed but need not decrease hours of work. Writing the leisure demand function in terms of wage rates  $W$  and full income  $(WT + A)$  where  $A$  is asset income and  $T$  is total time available,  $L = F(W, WT + A)$ ,  $\partial L/\partial T = F_2W$ ,  $\partial h/\partial T = (1 - F_2W)$ . Since  $F_2$  is positive the sign of the derivative is ambiguous. Ceteris paribus the higher the market wage rate, the more likely is it the case that time and money costs operate in the same direction.

the  $U_{j1}$  one can write

$$(14a) \quad E(Y_{11} | X_{21}, Y_{21} > 0) = X_{11} \beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda_1$$

$$(14b) \quad E(Y_{31} = h_1 | X_{21}, Y_{21} > 0) = X_{21} \beta_3 + \frac{\sigma_{23}}{(\sigma_{22})^{1/2}} \lambda_1$$

using the same definition of  $\lambda_1$  given before.<sup>1</sup>

### B. Main Empirical Results

The data utilized in the empirical analysis are a sample of 1735 women taken from the 1967 National Longitudinal Survey of Work Experience of Women Age 30-44 (The "Parnes" data) who are white, married with spouse present with husbands working in the previous year (1966). A woman is classified as working if she worked for pay in 1966 and satisfied the other sample selection criteria. Using this definition, 812 of the women work in 1966. The primary data source is described elsewhere in detail (Shea, et al., 1970). A more complete description of the means of the data used here and the sources of sample attribution is provided in Appendix A.

Given current professional ignorance about the appropriate dimension of labor supply, a variety of measures could be analyzed. Instead, a careful examination of the available data suggests that only one reliable measure is available: annual labor supply as defined by dividing annual earnings in 1966 by a questionnaire wage asked in early 1967. A superficial inspection of the data source suggests that a direct measure of labor supply is available

---

<sup>1</sup>Note that equation (13c) is cast in terms of hours of work. Technically speaking, hours should be treated as a limited dependent variable. In the empirical analysis presented below, I ignore this complication. Since hours of work distributions are concentrated far away from zero, and it is possible to use log hours rather than hours in a more general formulation.



by taking the product of weeks worked in 1966 and "usual hours worked." However these data are not usable since "weeks worked" includes vacation time and sick leave--two important margins of adjustment. Inspection of the histograms of both weeks worked and annual hours defined as the product of weeks worked and average weekly hours, suggests too much bunching of hours in standard reporting intervals. Appendix C presents the histograms for the measure used as well as the standard measures.<sup>1</sup> An important point to note in these histograms is that the distribution of annual hours of work, properly measured, shows much less of the bunching away from zero hours of work that is manifest in conventional measures of labor supply. It is precisely this artificial bunching that has stimulated recent work that introduces fixed costs into the analysis of labor supply behavior (Rosen, 1974). Accordingly, it is not surprising to find the result reported below, that with a proper measure of labor supply, there is much less evidence in favor of models with fixed costs of work.

---

<sup>1</sup>One disadvantage of the choice of labor supply measure that is offered in the text is that some women who work in 1966 did not supply a questionnaire wage. Only five percent of the sample is lost for this reason.

The specification of the economic relationships is conventional and requires little comment. Following Mincer (1974), the logarithm of wage rates is assumed to depend on schooling and market experience. Experience is defined as the number of years since leaving school that a woman has worked six months or longer. Following much previous research, female labor supply is postulated to depend on wage rates of the head and wife, the presence of children, family assets and wife's education.

Recent work by the author (Heckman (1977)) presents evidence that the labor market experience of the wife cannot be treated as an exogenous variable in the participation decision. (Evidence on this is offered below.) This variable records the wife's previous work history and is highly correlated with unmeasured determinants of current labor force participation. The empirical analysis discussed below explicitly deals with the endogeneity issue and considerable evidence is found for endogeneity of experience in labor supply and participation equations but little evidence for endogeneity in wage functions. Estimates of labor supply functions that purge "experience" of its endogenous component produce more plausible labor supply estimates.

The structure of the discussion of the empirical results is as follows. First, estimates of equation (13b) are discussed. Then estimates of the labor supply and wage functions are presented. Finally, some tests of the simple model of equations (11a) and (11b) are performed in a separate section.

Table 1 records the estimates of the  $(\sigma_{22}^{1/2})$  normalized coefficients of equation (13b) which generates the probability that a woman works. The first column presents estimates based on the assumption that experience is exogenous. The second column presents estimates based on predicted experience. The instrumental variables used to predict experience are reported below the table.

TABLE 1

PROBIT ESTIMATES OF THE PARAMETERS DETERMINING THE  
 PROBABILITY THAT A WOMAN WORKS (EQUATION (13b))  
 (Asymptotic normal statistics in parentheses)

	$\beta_2 / (\sigma_{22})^{1/2}$	
	(1) Estimates That Treat "Experience" As Exogenous	(2) Estimates That Treat "Experience" As Endogenous
Intercept	-.817 (4.7)	-.412 (1.56)
Nbr. of children less than 6	-.504 (10)	-.493 (9.11)
Assets	.436x10 <sup>-7</sup> (.25)	.619x10 <sup>-6</sup> (.29)
Husband's hourly wage rate (\$/hr.)	-.177 (8.0)	-.167 (7.81)
Wife's labor market experience (yrs.)	.098 (15.0)	.046 (1.81)
Wife's education	.080 (15.3)	.074 (5.3)
Log likelihood	-920.9	-1073.1
Observations	1735	1735

The probability that woman  $i$  works is

$$\frac{X'_{21}\beta_2}{(\sigma_{22})^{1/2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

The instrumental variables used to predict experience are linear and squared terms for children less than six, 1967 assets, husband's age, husband's education, husband's hourly wage, wife's education, and interactions of all linear terms.

As expected from a reading of the literature, the presence of small children, and a higher husband's wage rate lower the probability that a randomly selected woman works. Women with greater education are more likely to work. For both sets of estimates, greater work experience raises the probability of participation although both the size of the effect, and its statistical significance, are diminished when predicted experience is used in place of the actual variable in the estimation of the probit coefficients. A straightforward application of the Wu (1973) test rejects the null hypothesis that "experience" is uncorrelated with the error term in (13b).<sup>1</sup>

Following the methodology outlined in Section II, the probit coefficient estimates may be used to consistently estimate  $\beta_2/\sigma_{22}^{1/2}$ ,  $Z_1$  and hence  $\lambda(Z_1)$ . Hourly wage regressions with and without these estimated regressors are presented in Table 2 which also presents some evidence on the endogeneity of experience in the wage function. Column 1 presents the estimates of the traditional wage function. The estimates of the traditional equation corrected for censoring, but assuming experience to be endogenous, are presented in column 2. There is some indication of sample censoring but it is not overwhelming. The test statistic on " $\lambda$ " in column 2 is only marginal, and the wage coefficient estimates are essentially unchanged from column 1. Columns 3 and 4 record the results of an analysis that predicts experience and tests whether or not regression specifications based on predicted experience differ significantly from regression specifications with actual experience. Inspection of the Wu statistic on the bottom line of columns 3 and 4 suggest that the endogeneity

---

<sup>1</sup>The Wu test as used here consists of entering both "experience" and the residual of "experience" from predicted "experience" in the probit function. If the coefficient on the residual is significantly different from zero, one rejects the null hypothesis of uncorrelatedness of experience with the error term. The test statistic for this model gave a "t" of 2.1.

TABLE 2  
 RESULTS FOR HOURLY WAGE RATES (EQUATION (13a))  
 (Asymptotic normal statistics in parentheses)

Variables	(1) Traditional Regression	(2) Traditional Regression Corrected for Censoring	(3)* Traditional Regression Corrected for Endogeneity	(4)* Traditional Regression Corrected for Endogeneity and Censoring
Wife's labor market experience (yrs.)	.0167 (7.85)	.0207 (5.99)	.003 (.33)	.0151 (1.2)
Wife's education (yrs.)	.0763 (13.6)	.0779 (13.7)	.074 (12.7)	.076 (12.8)
$\lambda$	. . . .	.0878 (1.48)	. . . .	.1002 (1.4)
Intercept	-.401	-.515	-.226	-.419
$R^2$	.230	.235	.168	.1709
Wu statistic**	. . . .	. . . .	.34 (t)	1.17 (F)

\* The instrumental variables for experience are listed in Table 1.

\*\* The Wu statistic is the "t" score on the residual of predicted experience from actual experience in column 3 and is the "F" score on this residual and the residual between predicted  $\lambda$  and actual  $\lambda$  (based on measured experience). Predicted  $\lambda$  is obtained from a regression of  $\lambda$  based on measured experience on polynomials in the instrumental variables.

of experience is not an important issue in the estimation of the coefficients of the wage function. In my judgment, the best wage function is the traditional one recorded in column 1, but there is little to choose from between the estimates presented in column 1 and those presented in column 2.

The story with respect to the estimates of the labor supply functions is different. There is strong evidence for both sample censoring and endogeneity/ of experience. The estimates of the traditional regression specification are displayed in column 1 of Table 3. These estimates are in agreement with those in previous studies and require little comment. The regression estimates recorded in column 2 are unreasonable. There is little evidence of sample censoring, but the coefficients of the equation are not fit with much precision. Column 3 differs from column 1 in that experience is treated as an endogenous variable. The result of the Wu test applied to this equation strongly rejects the null hypothesis that experience is an exogenous or predetermined variable.

Column 4 displays the estimates of the labor supply function accounting for sample censoring and endogeneity of experience. The null hypothesis that experience is predetermined and the null hypothesis of no censoring are both rejected. Accordingly, the estimates in column 4 are offered as the best in this table.

A comparison of columns 3 and 4 reveals important differences. Except for insignificant coefficients, all of the slope coefficients in the labor supply equation presented in column 4 are larger in absolute value than the coefficients in column 3. The elasticity of labor supply in the column 3 estimates is high but not too much outside the range of estimates presented by Schultz (1975, p. 31). The elasticity of 4.5 derived from the specification in column 4 seems unduly large and requires some comment.

TABLE 3

ANNUAL HOURS WORKED DEFINED BY DIVIDING EARNINGS BY WAGE RATES (EQUATION (13c))  
(Asymptotic normal statistics in parentheses)

Variables	(1) Traditional Regression	(2) Traditional Regression Corrected for Censoring	(3) Traditional Regression w/Experience Treated as Endogenous	(4) Traditional Regression Corrected for Censoring w/ Exper. Treated as Endogenous	(5)** Marginal Effect of Variables from (4) Evaluated at Sample Mean
Nbr. of children less than 6	-155.8 (3.73)	-94.1 (.93)	-141.04 (2.61)	-925.1 (2.75)	-207.2
Assets (\$)	$3.1 \times 10^{-3}$ (.18)	$3.2 \times 10^{-3}$ (.191)	$3.2 \times 10^{-3}$ (.17)	$3.1 \times 10^{-3}$ (.17)	$+3.1 \times 10^{-3}$
husband's hourly wage rate (\$/hr.)	-33.1 (1.9)	-12.4 (.35)	-16.9 (.9)	-275.4 (2.5)	-32.9
Wife's experience (yrs.)	48.6 (12.1)	39.1 (2.66)	57.2 (2.55)	128.9 (3.4)	60.0
Wife's education (yrs.)	21.1 (1.89)	11.7 (.66)	10.9 (.89)	119.5 (2.5)	11.70
$\lambda$	. . . . .	-201.8 (.67)	. . . . .	2401 (2.36)	. . . . .
Intercept	664.2	912.1	765.1	-1755	. . . . .
$R^2$	.17	.176	.115	.121	. . . . .
Wu statistic*	. . . . .	. . . . .	3.1 (t)	9.8 (F)	. . . . .
Implied labor supply elasticity w/respect to hourly wage rates***	1.68	1.35	1.99	4.47	2.08

\* The Wu statistic is explained in Table 2.

\*\* These estimates are obtained from the coefficients in column 4 added to the effect of the variable on the conditional mean of the disturbance,  $(\partial \lambda / \partial X_{21})$ , multiplied by the coefficient on  $\lambda$  (2401).  
\*\*\* These estimates are obtained by dividing the experience coefficient in the wage function into the experience coefficient in the labor supply function and dividing by average labor supply.

The important point to note is that traditional estimates of the coefficients of labor supply functions of working women confound two effects: movement along a given labor supply function for working women and movement across taste distributions. Thus, for example, presence of an additional child under six has a dramatically negative effect on hours of work for a working woman (-925 hours reduction in supply). But working women with an additional child have a greater average taste for market work since only the most work prone women remain at work after the imposition of a child.

The two separate effects for each variable can be combined and evaluated at the sample mean. The result of such combination is displayed in column 5. By and large there is close agreement between the coefficients of column 3, which are estimates of the combined effect, and the coefficients in column 5. In particular, the estimates of the wage elasticities are in very close agreement.

The conclusion to be drawn from the labor supply analysis is that traditional methods of estimating labor supply functions give a downward biased (in absolute value) estimate of the true effect of economic variables. The estimates presented here reveal a strong behavioral response to wage change which is not discordant with previous estimates, but which casts a new light on their interpretation.

The estimated labor supply elasticity reported in column 3 (1.99) is quite similar to an estimate of 2.3 reported by Harvey Rosen who uses the same data set.<sup>1</sup> Rosen's estimated elasticity combines the effect of a wage change on/ the labor supply of already working women with the effect of a wage change in altering

---

<sup>1</sup>See Rosen (1976).



the composition of the sample of working women. Thus his estimate is compared with the estimate in column 3, and is an understatement of the effect of a wage change on the labor supply of women already working.

A final feature of the estimates presented in column 4 is worth noting. The coefficient on " $\lambda$ " is large and positive. This suggests that unmeasured factors that raise the probability of participation also tend to increase the volume of labor supplied to the market. The sign of the correlation is in accord with that predicted by the simple model of equations (11a) and (11b).

C. Tests of the Simple Model of Equations (11a)-(11b) and a Revised Model

In this section, some informal tests of the simple model of equations (11a) and (11b) are conducted. Most, but not all, of the restrictions predicted by the model are in accord with the data. An expanded version of the simple model is offered that allows for the effect of variation in the availability of informal day care arrangements that is documented elsewhere (Heckman, 1974) as well as variation in the fixed costs of work examined by Cogan (1975) and Hanoch (1975). The structure of this section is as follows. First, informal tests are discussed. Then, a revised model is offered.

As previously noted (page 21), one implication of the simple model of equations (11a) and (11b) is proportionality between the estimates of  $\beta_2/\sigma_{22}^{1/2}$  from the probit function and the parameters of the labor supply function  $\beta_3$  (i.e.,  $\beta_3 = (-\frac{\sigma_{22}^{1/2}}{\gamma})\beta_2/\sigma_{22}^{1/2}$ ). The constant of proportionality is predicted to be positive.

The ratios of the probit coefficient estimates (taken from column 2 of Table 1) to the hours of work coefficient estimates (presented in column

4 of Table 3) are displayed in Table 4. Given the sampling error in estimating these coefficients, the ratios are remarkably close to each other. The agreement is closer yet if one examines only the ratios of coefficients that are statistically significant in both equations. These ratios are denoted by an asterisk.

A second test of the simple model is available. From equation (7b) and (11b), one can write the hours of work function as

$$h_i = \frac{\sigma_{22}^{1/2}}{\gamma} (-Z_i + \lambda_i) + \frac{1}{\gamma} v_{2i}.$$

The variables in parentheses can be estimated from the probit coefficients. From equation (8c), the variance in the residual in the second term is given

by

$$E(v_{2i}^2) = \left(\frac{\sigma_{22}^{1/2}}{\gamma}\right)^2 (1 + \lambda_i Z_i - \lambda_i^2).$$

TABLE 4  
RATIO OF PROBIT COEFFICIENTS ( $\beta_2/\sigma_{22}^{1/2}$ ) TO LABOR SUPPLY  
FUNCTION COEFFICIENTS ( $\beta_3$ )

Intercept	Nbr. of Children Less than 6	Assets	Husband's Hourly Wage Rate	Wife's labor Market Experience	Wife's Education
$.23 \times 10^{-3}$	$.53 \times 10^{-3*}$	$.19 \times 10^{-3}$	$.61 \times 10^{-3*}$	$.36 \times 10^{-3}$	$.62 \times 10^{-3*}$

\* Denotes a ratio of coefficients that are statistically significant at conventional levels in both relationships.

TABLE 5  
ESTIMATES OF EQUATION 15

k (From regression coefficient)	Standard Error of Estimate (From regression residuals)
1424 (t stat. is 46)	1221

TABLE 6  
LABOR SUPPLY COMPARISONS WITH THE "TOBIT" MODEL

	"Tobit" Estimates	Estimates from Table 3, Column 4
Nbr. of children less than 6	-658.5 (9.5)	-925.1 (2.78)
Assets (\$)	$.24 \times 10^{-2}$ (.9)	$.3 \times 10^{-2}$ (.2)
Husband's hourly wage rate (\$/hr.)	-201.4 (7.6)	-275.4 (2.5)
Wife's experience* yrs.	87.02 (5.1)	128.9 (3.4)
Wife's education yrs.	88.0 (5.1)	119.5 (2.5)
Intercept	-669 (2.0)	-1755
Estimated standard error of regression ( $\sigma_{22}^{1/2}/\gamma$ )	1409	. . . .
Ln likelihood	-7595.53	. . . .

\* Predicted experience as defined in Table 1.

Thus another test of the simple model can be conducted. Run the weighted interceptless regression

$$(15) \quad h_1 w_1 = k(-Z_1 + \lambda_1) w_1 + \varepsilon_1$$

where  $w_1 = (1 + \lambda_1 Z_1 - \lambda_1^2)^{-1/2}$ . The regression coefficient is a consistent estimator of  $\sigma_{22}^{1/2}/\gamma$  which is the square root of the residual variance in the estimated equation. A test of the simple model is to compare the square root of the estimated residual variance with the regression coefficient. This comparison is made in Table 5. The agreement between the two estimates is remarkably close. Moreover, as previously discussed in the first paragraph of this section, the constant of proportionality estimated in Table 4 is the inverse of  $\sigma_{22}^{1/2}/\gamma$ . Using  $.6 \times 10^{-3}$  as an estimate of the average ratio, another estimate of  $\sigma_{22}^{1/2}/\gamma$  is 1666, again a number close to the regression coefficient estimate.

Other informal tests of the simple model are possible. If equation (11b) is the labor supply function, the model of Tobin, discussed in section one ("Tobit") is an appropriate description of the labor supply function. Estimated Tobit coefficients are displayed in Table 6. "Tobit" underestimates (in absolute value) the coefficients of the labor supply function.<sup>1</sup> Except for the intercept term, each "Tobit" coefficient is about seven-tenths of the corresponding coefficient of the unrestricted labor supply estimates, reproduced in the second column of Table 6. Note, however, that the estimated standard error of the Tobit regression (1409) is remarkably close to the previous estimate (1424) obtained from equation (15).

<sup>1</sup>This understatement of Tobit coefficients suggests that the Shultz (1975) estimates of female labor supply elasticities, based on Tobit, are downward biased.

At this point it may be helpful to take stock of what has been learned. The simple model is almost right. The "Tobit" estimates of the slope coefficients are smaller (in absolute value) than the initial consistent estimates, but by a constant of proportionality (.7). The only discordance in this pattern comes in the estimates of the intercept terms. The Tobit intercept is disproportionately larger than the unrestricted labor supply intercept than the ratio of slope coefficients would suggest is appropriate.

The model of equations (11a)-(11c) and (12) may be modified slightly to rationalize this pattern. First review the economics of the simple model. To focus ideas suppose the wage function (corresponding to equation (11a)) is

$$(16a) \quad W = \alpha_0 + \alpha_1 E + U$$

while the reservation wage function (i.e., value of time) is

$$(16b) \quad W^* = \gamma_0 + \gamma_1 h + \gamma_2 A + \epsilon$$

where  $\epsilon$ ,  $U$  are disturbances,  $E$  is (exogenous) experience,  $A$  is asset income and  $h$  is time not spent at home. The simple theory assumes that  $\alpha_1 > 0$ ,  $\gamma_1 > 0$ ,  $\gamma_2 > 0$ . A woman works if  $W > W^*$  at zero hours of work, i.e.,

$$(17) \quad \alpha_0 + \alpha_1 E + U > \gamma_0 + \gamma_2 A + \epsilon.$$

Her labor supply function is obtained by equating (16a) and (16b) when inequality (17) is met, i.e.,

$$(18) \quad h = \frac{1}{\gamma_1} (\alpha_0 + \alpha_1 E - \gamma_0 - \gamma_2 A) + \frac{U - \epsilon}{\gamma_1}.$$

Suppose that there are work related costs such as day care and other household expenses. Recent evidence (Heckman, 1974b) suggests that some women have access to limited quantities of low cost day care and other household services from friends and neighbors. An analytically simple way to characterize such limited availability of low cost substitutes is to view it as an augmentation of the woman's time budget that expands available time by less than one hour for each hour worked up to some given number of working hours. A consequence of the limited availability of low cost substitutes is a discontinuity in the labor supply function at the given number of hours.

Figure 2 illustrates this case. The solid line <sup>AA'</sup> is the labor supply curve for a woman of given characteristics who uses market substitutes for her time. Market substitutes for the wife's household input are assumed to

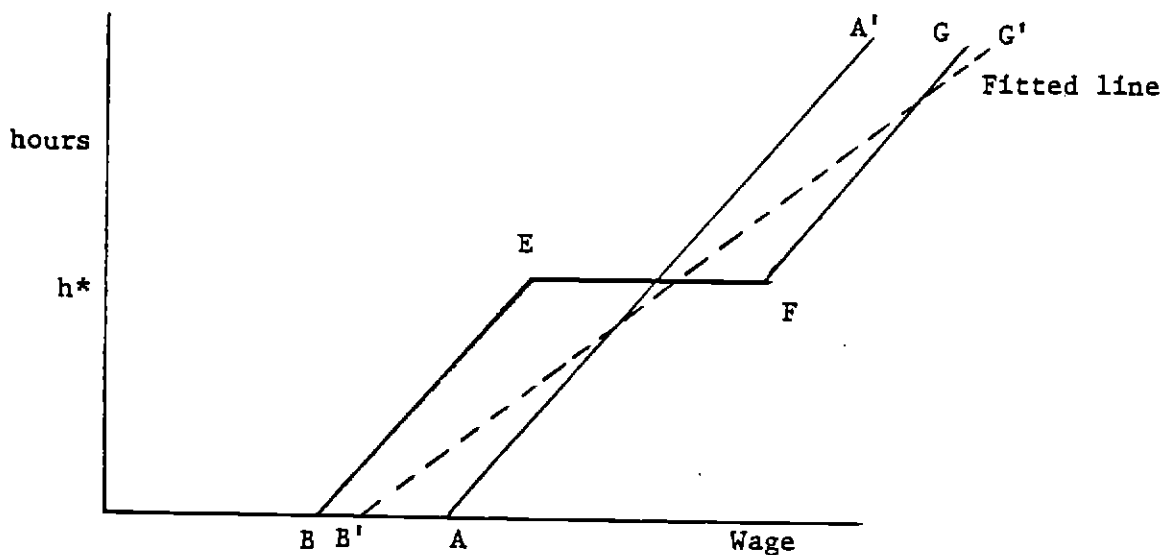


Figure 2

be available at fixed marginal prices. The reservation wage for such women is given by A.<sup>1</sup> The line, BEFG, illustrates a labor supply curve for

<sup>1</sup>As recently discussed by Brumm (1976), no unique reservation wage is defined if there are work related costs. Nonetheless, the extension of the such

a woman who has access to informal sources that (imperfectly) replace her time at home up to  $h^*$  hours. Note that this woman has a lower reservation wage,  $B$ , than the other woman, but that beyond  $h^*$  hours, it takes a greater wage rate to induce her to work more hours. This is so because of the wealth effect that arises from her access to low cost sources, and from the assumption that leisure is a normal good. The population of all women contains a mixture of women with the two types of labor supply functions.

In a general model it is plausible that both slopes and intercepts of labor supply functions are affected by the limited availability of low cost substitutes for the woman's time. Before more elaborate models are explored it is useful to examine more fully the implications of the simple model depicted in Figure 2.

Given a distribution of the two types of labor supply functions in the population, and given that some women with a "broken" labor supply function have hours of work in excess of  $h^*$ , the average reservation wage in the population is less than the average of the intercepts of the labor supply functions. In a model that ignores interpersonal variation in the cost of household substitutes, the two measures coincide so that the average of the intercepts is the average of the reservation wages. An important consequence of the inequality of reservation wages and intercepts is that any empirical procedure that constrains the intercept of the labor supply equation to be the reservation wage understates the effect of wages (and other variables) on labor supply. This insight is important because the model of equations (11a)-(11b) and the Tobit model both impose this constraint on the data.

To establish this result intuitively, note that using equations (16a) and (16b), equation (18) can be written as

---

labor supply curve to  $A$  defines a wage that plays the same role as the reservation wage in a model with no work related costs.

$$h = \frac{1}{\gamma} (w - w^*)$$

so that the wage that just induces a woman to work a positive number of hours is the reservation wage  $w^*$ . In terms of the notation of Figure 2,  $A = w^*$ . If one constrains the intercept of the labor supply curve to be the reservation wage, when the "broken line" function BEFG describes some or all of the data, one underestimates the response of hours to wage rates as well as the intercept in the labor supply equation. See the dashed line B'G' in Figure 2. One can prove that "Tobit" and the model of equation (11b) impose this constraint on the fitted function.

By way of contrast with these results, it is helpful to consider a model with fixed costs of work. For simplicity consider money costs of work. As both Cogan (1975) and Hanoch (1975) have shown, the effect of fixed costs of work on labor supply is that women who work at all must work a minimum number of hours, say  $\hat{h}$ , to recoup the fixed costs. The reservation wage is raised over a case without fixed costs. This model is depicted in Figure 3. Here the standard labor supply function for a woman of given characteristics is indicated by a solid line while the modified labor supply function is indicated by a dashed line. As before, "A" denotes the reservation wage in the standard case while "C" denotes the reservation wage in the case of

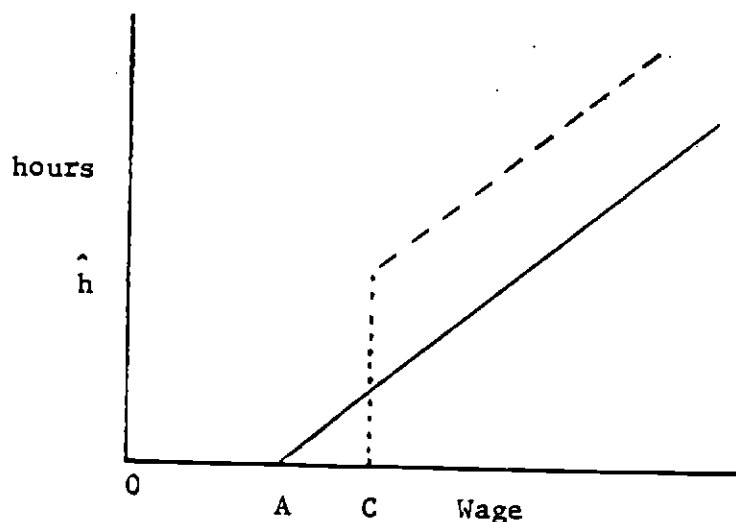


Figure 3



fixed costs of work. In the presence of fixed costs, the reservation wage is greater and the level of the supply function is higher reflecting the assumption that leisure is a normal good, and that fixed costs subtract from income. Unlike the situation in the previous case, the average of the reservation wages exceeds the average of the intercepts. Thus any model that constrains the intercept of the labor supply function to equal the average of the reservation wages overstates the effect of wages (and other variables) on labor supply.<sup>1</sup>

The empirical results reported in Tables 4 and 6 favor the model of differential access to low cost substitutes for time in the home over a model with fixed costs in a dominant role. "Tobit" underestimates the response of labor supply to a change in economic variables, and the Tobit intercept is higher than the intercept of the unrestricted labor supply function, an implication of a model in which the true reservation wage is less than the intercept of the labor supply equation. Finally, note that the only modification required to make equations (11a)-(11b) consistent with data is relaxation of the interequation proportionality of intercepts. The corresponding interequation slope coefficients are related by a common factor of proportionality.<sup>2</sup>

---

<sup>1</sup>This argument is made by both Cogan (1975) and Hanoch (1975). These effects would also arise if employers offered "tied" packages of wages and hours if each individual had his own "best" minimum hours offer in the market. In this case, the labor supply function for the tied case would coincide over the relevant range with the standard labor supply function, but the average of the intercepts would understate the reservation wage.

<sup>2</sup>Another implication of the modified model that the correlation between the disturbances of the labor supply function and the participation equation need not be unity because there is a source of variation in the labor supply function that does not effect the participation equation (maximum availability of close substitutes  $h^*$ ).

The choice of the dependent variable crucially affects the outcome of such tests. In results not reported here, use of the conventional measure of labor supply defined as the product of "usual hours per week" and "usual weeks" leads to precisely opposite implications, i.e., one would accept a model of fixed costs. But as previously noted, the standard measures induce the illusion of fixed costs via reporting error that overstates the extent of labor supply and the frequency of occurrence of standard reporting intervals so that empirical analyses based on this measure of labor supply yield misleading conclusions.

Summary and Conclusions

In this paper, the bias that results from using nonrandomly selected data is discussed within the specification error framework of Griliches and Theil. A computationally tractable technique is discussed that enables economists to utilize simple regression techniques to estimate behavioral functions free of selection bias. Asymptotic properties of the estimator are developed.

A model of female labor supply and wage rates is estimated with this technique. The empirical results suggest that selection bias is an important problem in estimating labor supply functions but is less important in estimating wage functions. Very high estimates of the elasticity of female labor supply are derived but these are shown to be consistent with conventional estimates that ignore selection bias. The labor force experience of the wife is shown to be an endogenous variable in labor supply equations but not in wage functions.

Some informal tests of the model of Heckman (1974a) are presented. Many implications of the model appear to be in accord with the data but an expanded model that introduces the notion of limited household accessibility to low cost substitutes for the wife's time appears to fit the data better. With a proper measure of labor supply, the implications of a model with fixed costs of work in a dominant role are rejected by the data.

APPENDIX A\*

The Asymptotic Distribution of Estimators  
Based on an Estimated  $\lambda_1$

For notational convenience rewrite equation (7a) in the text

$$(A1) \quad Y_{1i} = X_{1i}\beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda_1 + v_{1i}$$

$$i = 1, \dots, T_1 < T$$

When an estimated value of  $\lambda_1$  is used in place of the true  $\lambda_1$ , equation (7a) is modified to read

$$(A2) \quad Y_{1i} = X_{1i}\beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \hat{\lambda}_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} (\lambda_1 - \hat{\lambda}_1) + v_{1i}$$

Estimates of  $\lambda_1$  are taken from probit functions fit on the full sample of T observations. Thus, we know

$$\sqrt{T} \left( \hat{\beta}_2 / \sigma_{22}^{1/2} - \frac{\beta_2}{\sigma_{22}^{1/2}} \right) = N(0, \Sigma),$$

Since  $\lambda_1 = \lambda(Z_1) = \lambda(-X_{21}\beta_2 / \sigma_{22}^{1/2})$ , and since  $\lambda$  is a twice continuously differentiable function of  $Z_1$ ,

$$\text{plim } \hat{\lambda}_1 = \text{plim } \lambda(\hat{Z}_1) = \text{plim } \lambda(-X_{21}\hat{\beta}_2 / \sigma_{22}^{1/2}) = \lambda_1$$

and

$$\sqrt{T} (\hat{\lambda}_1 - \lambda_1) = N(0, \left(\frac{\partial \lambda}{\partial Z_1}\right)^2 X_{21} \Sigma X_{21}')$$

---

\*Remarks by Takeshi Amemiya stimulated this section. Of course, he is not responsible for any errors.

It is useful to bear in mind that the sample size for the probit estimates (T) exceeds the sample size used to estimate equation (A2) ( $T_1$ ).

The fact that  $\hat{\lambda}_1$  converges in probability to  $\lambda_1$  does not imply that least squares estimates of  $(\beta_1, \sigma_{12}/\sigma_{22}^{1/2})$  based on  $\hat{\lambda}_1$  converge in distribution to an estimate based on a known  $\lambda_1$ , but convergence, in probability does imply consistency of the least squares estimator.

There are three cases to consider: Case 1:  $\sigma_{12} = 0$ . Case 2;  $\lambda_1 =$  constant. Case 3: The general case.

Case One is an important special case. Clearly, under the null hypothesis that  $\sigma_{12} = 0$  (i.e., that there is no selection bias) the regression estimator based on an estimated  $\lambda_1$  converges in distribution to the estimator based on known  $\lambda_1$ . This result is useful in practical work because of the important nature of the null hypothesis.

Case 2 is uninteresting. If no variables affect the probability of sample selection, it is not possible to estimate  $\sigma_{12}/\sigma_{22}^{1/2}$  as long as  $\beta_1$  contains an intercept term. If the intercept is suppressed, Case 2 is a special case of Case 3 below.

Case 3 is the general case and its development will require some argument. Sampling is assumed to be "fixed in repeated samples" sampling.<sup>1</sup> There are L distinct data patterns  $(X_{1i_\ell}, X_{2i_\ell})$ ,  $\ell = 1, \dots, L$ . Assume L large enough so that all parameters of equation (7a) can be estimated as sample size increases (this requires at least  $K_1 + 1$  distinct  $X_{2i_\ell}$  configurations and at least  $K_2$  distinct  $X_{1i_\ell}$  patterns).

For each configuration, indexed by  $\ell$ , the number of observations becomes large ( $T_\ell \rightarrow \infty$ ). But suppose that  $\lim_{\substack{T \rightarrow \infty \\ T_\ell \rightarrow \infty}} \frac{T_\ell}{T} \rightarrow F_\ell$ ,  $0 < F_\ell < 1$ . Given

<sup>1</sup>The ensuing propositions are true for more general data schemes but "fixed in repeated samples" sampling is conceptually simple and appropriate to the survey data analyzed in the body of the paper.

the sample selection rule, only some fraction of the observations ( $\phi(Z_{i_1})$ ) have observations on  $Y_{i_1}$ .

For data configuration  $\ell$ , the error term of equation (A2) may be written as

$$\frac{\sigma_{12}}{(\sigma_{22})^{1/2}} (\lambda_{i_\ell} - \hat{\lambda}_{i_\ell}) + V_{1i_\ell}.$$

We seek the distribution of the terms

$$(A3) \quad \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \frac{\sum_{i_\ell=1}^{T_\ell} \hat{\lambda}_{i_\ell} (\lambda_{i_\ell} - \hat{\lambda}_{i_\ell}) \phi(Z_{i_\ell})}{\sum_{i_\ell=1}^{T_\ell} \sqrt{T_\ell}} + \frac{\sum_{i_\ell=1}^{T_\ell} \hat{\lambda}_{i_\ell} V_{1i_\ell} \phi(Z_{i_\ell})}{\sum_{i_\ell=1}^{T_\ell} \sqrt{T_\ell}}.$$

Note that  $\phi(Z_{i_\ell})$  appears in the expression because only proportion  $\phi(Z_{i_\ell})$  of the random sample of  $T_\ell$  observations generate an observed  $Y_{1i_\ell}$ .

It is straightforward to prove that the second term in equation (A3) converges in distribution to the random variable

$$\phi(Z_{i_\ell}) \frac{\sum_{i_\ell=1}^{T_\ell} \lambda_{i_\ell} V_{1i_\ell}}{\sum_{i_\ell=1}^{T_\ell} \sqrt{T_\ell}}.$$

The interesting aspect of (A3) is that the first term converges to a proper normal random variable if  $\sigma_{12} \neq 0$ . This is so because of the explicit exact dependence built into successive values of  $\lambda_{i_1} - \hat{\lambda}_{i_1}$  that are all based on the same set of probit coefficient estimates.

In fact,

$$\frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \frac{\sum_{i_\ell=1}^{T_\ell} \hat{\lambda}_{i_\ell} (\lambda_{i_\ell} - \hat{\lambda}_{i_\ell})}{\sum_{i_\ell=1}^{T_\ell} \sqrt{T_\ell}} \phi(Z_{i_\ell}) \sim N(0, \frac{\sigma_{12}^2}{\sigma_{22}} \Omega_{i_\ell i_\ell'})$$

where

$$\Omega_{i_{\ell} i_{\ell}'} = (\lambda_{i_{\ell}})(\lambda_{i_{\ell}'}) \left(\frac{\partial \lambda}{\partial Z_{i_{\ell}}}\right) \left(\frac{\partial \lambda}{\partial Z_{i_{\ell}'}}\right) \phi(Z_{i_{\ell}}) \phi(Z_{i_{\ell}'}) (F_{\ell} F_{\ell}')^{1/2} X_{2i_{\ell}} \Sigma X_{2i_{\ell}'}$$

Then it is the case that

$$\frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \quad \begin{matrix} L \\ \Sigma \\ \ell=1 \end{matrix} \quad \begin{matrix} L \\ \Sigma \\ \ell'=1 \end{matrix} \quad \frac{\hat{\lambda}_{i_{\ell}} (\lambda_{i_{\ell}} - \hat{\lambda}_{i_{\ell}'})}{\sqrt{T_{\ell}}} \quad \frac{\phi(Z_{i_{\ell}})}{\sqrt{T}} \quad \frac{\sqrt{T_{\ell}}}{\sqrt{T}}$$

$$N\left(0, \frac{\sigma_{12}^2}{\sigma_{22}} \quad \begin{matrix} L & L \\ \Sigma & \Sigma \end{matrix} \quad \Omega_{i_{\ell} i_{\ell}'}\right)$$

where the variance does not converge to zero.

The two terms in (A3) are asymptotically independent. (The proof is trivial and hence is omitted.) Now we derive the distribution of the least squares estimators based on estimated  $\lambda_i$ .

Define

$$B = \begin{pmatrix} \frac{\Sigma X_{1i}' X_{1i} \phi(Z_i)}{T} & \frac{\Sigma \phi(Z_i) X_{2i}' \lambda_i}{T} \\ \frac{\Sigma X_{1i} \lambda_i \phi(Z_i)}{T} & \frac{\Sigma \lambda_i^2 \phi(Z_i)}{T} \end{pmatrix}^{-1}$$

$$= \text{plim} \begin{pmatrix} \frac{\Sigma X_{1i}' X_{1i} \phi(Z_i)}{T} & \frac{\Sigma \phi(Z_i) X_{1i}' \hat{\lambda}_i}{T} \\ \frac{\Sigma X_{1i} \phi(Z_i) \hat{\lambda}_i}{T} & \frac{\Sigma \hat{\lambda}_i^2 \phi(Z_i)}{T} \end{pmatrix}^{-1}$$

Under the previous assumptions, this matrix exists and is of full rank ( $K_1 + 1$ ).

Then

$$\sqrt{T} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \frac{\hat{\sigma}_{12}}{\sigma_{22}^{1/2}} - \frac{\sigma_{12}}{\sigma_{22}^{1/2}} \end{pmatrix} =$$

$$B \begin{pmatrix} \frac{\sum X'_{1i} V_{1i} \phi(Z_i)}{\sqrt{T}} \\ \frac{\sum \hat{\lambda}_i V_{1i} \phi(Z_i)}{\sqrt{T}} \end{pmatrix} + \frac{\sigma_{12}^2}{\sigma_{22}} \begin{pmatrix} \frac{\sum X'_{1i} (\lambda_i - \hat{\lambda}_i) \phi(Z_i)}{\sqrt{T}} \\ \frac{\sum \hat{\lambda}_i (\lambda_i - \hat{\lambda}_i) \phi(Z_i)}{\sqrt{T}} \end{pmatrix}$$

So

$$(A4) \quad \sqrt{T} \begin{pmatrix} \hat{\beta}_1 & - & \beta_1 \\ \hat{\sigma}_{12} & - & \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \end{pmatrix} \sim N(0, B \psi B')$$

where

$$\psi = \sigma_{11} \begin{pmatrix} \frac{\sum X'_{1i} n_i \phi(Z_i) X_{1i}}{T} & \frac{\sum X'_{1i} n_i \phi(Z_i) \lambda_i}{T} \\ \frac{\sum \lambda_i n_i \phi(Z_i) X_{1i}}{T} & \frac{\sum \lambda_i^2 n_i \phi(Z_i)}{T} \end{pmatrix} + \frac{\sigma_{12}^2}{\sigma_{22}} \begin{pmatrix} \sum_{\ell=1}^L \sum_{\ell'=1}^{L'} X'_{1i_\ell} X_{1i_{\ell'}} \Theta_{i_\ell i_{\ell'}} & \sum_{\ell=1}^L \sum_{\ell'=1}^{L'} X'_{1i_\ell} \Pi_{i_\ell i_{\ell'}} \\ \sum_{\ell=1}^L \sum_{\ell'=1}^{L'} X_{1i_\ell} \Pi_{i_\ell i_{\ell'}} & \sum_{\ell=1}^L \sum_{\ell'=1}^{L'} \Omega_{i_\ell i_{\ell'}} \end{pmatrix}$$

$$\text{where } n_i = (1 + \frac{\sigma_{12}^2}{\sigma_{22}} (Z_i \lambda_i - \lambda_i^2))$$

$$\Pi_{i_\ell i_{\ell'}} = - \left( \frac{\partial \lambda}{\partial Z_{i_\ell}} \right) \left( \frac{\partial \lambda}{\partial Z_{i_{\ell'}}} \right) \lambda_{i_\ell} (F_\ell F_{\ell'})^{1/2} \phi(Z_{i_\ell}) \phi(Z_{i_{\ell'}}) X_{2i_\ell} \sum X'_{2i_{\ell'}}$$

$$\Theta_{i_\ell i_{\ell'}} = \left( \frac{\partial \lambda}{\partial Z_{i_\ell}} \right) \left( \frac{\partial \lambda}{\partial Z_{i_{\ell'}}} \right) (\lambda_{i_\ell}) (\lambda_{i_{\ell'}}) (F_\ell F_{\ell'})^{1/2} \phi(Z_{i_\ell}) \phi(Z_{i_{\ell'}}) \cdot (X_{2i_\ell} \sum X'_{2i_{\ell'}})$$

and  $\Omega_{i_\ell i_{\ell'}}$  has been previously defined.



Under the null hypothesis of no sample selection bias,  $\frac{\sigma_{12}^2}{\sigma_{22}}$ , and  $B \psi B'$  is the standard least squares estimator for the variance-covariance matrix of the regression coefficients. In the general case, the parameters of  $\psi$  are estimable so that  $B \psi B'$  is estimable.

One may also derive approximate GLS estimators that do not converge in distribution to the GLS estimators with known  $\lambda_1$ . The essential ingredients of their derivation are available from previous results. Let  $T_{1\ell}$  denote the number of observations on configuration  $\ell$  that appear in the selected sample. Array the observations so that the first  $T_{11}$  observations are on configuration 1, etc. Then for a sample of size  $T_1$ , one may write

$$K = E(V_1 V_1') = \sigma_{11} \begin{pmatrix} n_1 I_{T_{11}} & & 0 \\ & n_2 I_{T_{12}} & \\ & 0 & n_L I_{T_{1L}} \end{pmatrix} + \frac{\sigma_{12}^2}{\sigma_{22}} \begin{pmatrix} \omega_{11} i_1 i_1' & \omega_{12} i_1 i_2' & \omega_{1L} i_1 i_L' \\ \omega_{21} i_2 i_1' & \omega_{22} i_2 i_2' & \dots \\ \dots & \dots & \omega_{LL} i_L i_L' \end{pmatrix}$$

where  $\omega_{\ell\ell'} = \left(\frac{\partial \lambda}{\partial Z_\ell}\right) \left(\frac{\partial \lambda}{\partial Z_{\ell'}}\right) X_{2i_\ell} \frac{\Sigma}{T_1} X_{2i_{\ell'}}'$ ,

and  $i_\ell$  is a  $T_{1\ell} \times 1$  vector of "1s". Now  $\Sigma$  is a positive definite matrix so

$$\Sigma = PP'$$

where  $P$  is an orthogonal matrix.

Then define

$$\psi_\ell = \left(\frac{\partial \lambda}{\partial Z_\ell}\right) \frac{X_{2i_\ell} P}{\sqrt{T_\ell}}.$$

so that  $K$  may be written as

$$K = \sigma_{11} \begin{pmatrix} n_1 I_{T_{11}} & & 0 \\ & \ddots & \\ 0 & & n_L I_{T_{LL}} \end{pmatrix} + \frac{\sigma_{12}^2}{\sigma_{22}} \begin{pmatrix} i_1 \otimes \psi_1 \\ i_2 \otimes \psi_2 \\ \vdots \\ i_L \otimes \psi_L \end{pmatrix} \begin{pmatrix} i_1' \otimes \psi_1' & \dots & \dots & \dots & i_L' \otimes \psi_L' \end{pmatrix}$$

and  $K^{-1}$  may be written as

$$K^{-1} = \sigma_{11}^{-1} \begin{pmatrix} I_{T_{11}} n_1^{-1} & & 0 \\ & \ddots & \\ 0 & & I_{T_{LL}} n_L^{-1} \end{pmatrix} + \gamma \begin{pmatrix} i_1 \otimes \psi_1^* \\ i_2 \otimes \psi_2^* \\ \vdots \\ i_L \otimes \psi_L^* \end{pmatrix} \begin{pmatrix} i_1' \otimes \psi_1^{*'} & & i_2' \otimes \psi_2^{*'} & \dots & \dots \end{pmatrix}$$

where  $\psi_\ell^* = (\psi_\ell) (n_\ell^{-1}) (\sigma_{11}^{-1})$

and  $\gamma = -\frac{\sigma_{12}^2}{\sigma_{22}} \left( 1 + \frac{\sigma_{12}^2}{\sigma_{22}\sigma_{11}} \sum_{\ell=1}^L T_{1\ell} \omega_{\ell\ell} n_\ell^{-1} \right)$

(see, e.g., Graybill (1969), Thm. 8.3.3).

Let the  $1 \times (K_1 + 1)$  data vector for observation  $i$  ( $X_{i1}, \hat{\lambda}_i$ ), be denoted by  $\gamma_i$ . The observation matrix is the  $T_1 \times (K_1 + 1)$  matrix  $R$ . Then the appropriate GLS estimator is

$$\tilde{\beta}_{1GLS} = (R'K^{-1}R)^{-1}R'K^{-1}Y_1.$$

Then

$$\sqrt{T} (\tilde{\beta}_{1GLS} - \beta_1) \sim N(0, R' K^{-1} R)$$

Note that the elements of  $K$  are estimable.

## APPENDIX B.1

In the 1967 National Longitudinal Survey of the work experience of women age 30-44, 5083 observations are available. The following sample selection criteria were imposed to reach a usable sample of 1735 women, 812 of whom work in 1966. The number of observations failing to meet a criteria is given in the column to the right of the rejection criterion. Observations may be rejected for any of the reasons listed, and a given observation may be rejected for several reasons.

(1) Nonwhite	1477
(2) Married spouse present	1019
(3) Farmers	252
(4) Missing husband's income	421
(5) Missing annual hours of husband (Including no work group)	336
(6) Missing wife's experience	301
(7) Missing wage data on wife	126

Assets were assigned in 165 cases. An equation is fit on the available 1570 observations. The equation is

$$\begin{aligned}
 \text{Assets (1967)} = & -6891 + 73 (\text{wife's experience}) \\
 & + 1647 (\text{wife's education}) + 466.4 (\text{number of children less than six}) \\
 & + 806.8 (\text{husband's education}) + 2040 (\text{husband's age}) - 17.475 \\
 & (\text{husband's age squared}).
 \end{aligned}$$

APPENDIX B.2

Sample Means of the Data Used in the Analysis

(From 1967 National Longitudinal Survey of the Work Experience of Women 30-44)

	<u>Workers</u>	<u>Total Sample</u>
Number of Observations	812	1735
Number of children less than 6	.312	.565
Assets (\$)	11,711	11,974
Husband's 1966 hourly wage rate (\$/hr)	3.45	3.73
Wife's education (yrs.)	11.42	11.29
Labor force experience (yrs.)	10.63	7.80
Wife's annual hours worked	1289	. . .
Wife's hourly wage rate	2.12	. . .
$\lambda$	.6412	1.12

## APPENDIX C.1

The histograms for reported weeks worked in 1966, reported annual hours worked, and estimated hours worked based on a division of 1966 earnings by a questionnaire wage rate are displayed in that order. The ordinate gives the number of observations for the value of the variable displayed on the abscissa.









## REFERENCES

- Amemiya, T. "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," Econometrica 42, No. 6 (November 1974): 999-1011.
- \_\_\_\_\_. "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica 41, No. 6 (1973): 997-1017.
- Becker, G. S. "A Theory of Marriage: Part Two," Journal of Political Economy 82, No. 2, Part II (1974): 511-526.
- Brumm, Harold. "Comment on 'Shadow Prices, Market Wages and Labor Supply,'" IDA, June 1976.
- Cain, Glenn and H. Watts. Income Maintenance and Labor Supply. Markham, 1973.
- Cogan, John. "Labor Supply with Time and Money Costs of Participation," The Rand Corporation (August, 1976).
- Cragg, J. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica 39, No. 5 (September 1971): 829-844.
- Gronau, Reuben. "The Effect of Children on the Housewife's Value of Time," Journal of Political Economy (March/April 1973).
- \_\_\_\_\_. "Wage Comparisons—A Selectivity Bias," Journal of Political Economy (November/December 1974): 1119-1144.
- Hanoch, Giora. "Hours and Weeks in the Theory of Labor Supply," The Rand Corporation, R-1787 (July 1976).
- Heckman, James H. (1974a), "Shadow Prices, Market Wages and Labor Supply," Econometrica (July 1974): 679-694.
- \_\_\_\_\_. (1974b). "Effects of Child Care Programs on Women's Work Effort," Journal of Political Economy (April 1974): 491-519.
- \_\_\_\_\_. "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 (Dec., 1976).
- \_\_\_\_\_. "Dynamic Models of Female Labor Supply," Unpublished paper, University of Chicago, March 1977.

- Johnson, N. and S. Kotz. Distributions in Statistics: Continuous Multivariate Distributions. New York: Wiley, 1972.
- Lewis, H. Gregg. "Comments on Selectivity Biases in Wage Comparisons," Journal of Political Economy (November/December 1974): 1145-1155.
- \_\_\_\_\_. Unionism and Relative Wages. Chicago: University of Chicago Press, 1963.
- Mincer, Jacob. "Labor Force Participation of Married Women," in Aspects of Labor Economics, National Bureau of Economic Research, 1962.
- Nelson, Forrest. "Censored Regression Models with Unobserved Stochastic Censoring Thresholds," National Bureau of Economic Research Working Paper 63, 1975.
- Nerlove, M. and J. Press. "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data," Discussion Paper #1, Center for Statistics and Probability, June 1976, Northwestern University.
- Rosen, Harvey. "Taxes in a Labor Supply Model with Joint Wage-Hours Determination," Econometrica 44, No. 3 (May 1976): 485-508.
- Rosen, Sherwin. "Comment on Heckman's Paper," Journal of Political Economy (March/April 1974).
- Rothenberg, T. and C. Leenders, "Efficient Estimation of Simultaneous Equation Systems," Econometrica 32, No. 1 (January 1964): 57-76.
- Schultz, T. Paul. "Estimating Labor Supply Equations for Married Women," The Rand Corporation R-1265-NIH/EDA (February 1975).
- Shea, et al. Dual Careers: A Longitudinal Study of Labor Market Experience. Columbus, Ohio: Center for Human Resource Research, The Ohio State University, May 1970.
- Tobin, J. "Estimation of Relationships for Limited Dependent Variables," Econometrica 26 (1958): 24-36.
- Wu, D. "Alternative Tests of Independence between Stochastic Regressions and Their Disturbances," Econometrica 41 (1973): 733-750.