

Industrial Relations Section  
Princeton University  
Working Paper #141\*  
November 1980

ESTIMATION OF A LABOUR SUPPLY MODEL  
WITH CENSORING DUE TO  
UNEMPLOYMENT AND UNDEREMPLOYMENT\*\*

by

John C. Ham  
Institute for Policy Analysis  
University of Toronto

## 1. Introduction

In recent years the estimation of labour supply functions has been one of the most active areas of research in labour economics, and this estimation, along with most policy applications stemming from it, is based on the crucial assumption that individuals can work as many hours as they desire. However two related developments have made this assumptions less tenable. First, the number of unemployed workers has grown over the past decade. Second, a substantial body of theoretical work has developed in which unemployment acts as a constraint on an individual's work-leisure choice.<sup>1</sup> Moreover, consideration of the possibility that unemployment represents a constraint naturally leads to a re-examination of whether the existence of a fixed work-week constrains workers in their hours decisions in the absence of unemployment. In this case workers may be underemployed in the sense that they are unable to work as many hours as they want even if they are not unemployed.

If the unemployed or underemployed are truly constrained it is straightforward to show the standard approach to estimating labour supply functions is inappropriate. As a result, in recent years several authors have proposed methods for handling this problem. However, all of these procedures are based on fairly restrictive assumptions (noted below); in particular they assume that these workers are constrained and/or that the constraints are exogenous and not a result of previous choice.

The purpose of this paper is to provide a new method of labour supply estimation in the presence of potentially constrained workers. This new method of estimation possesses three desirable features. First, it provides consistent estimates under much more general conditions than previous approaches;

most importantly the estimates are consistent independently of whether the unemployed and underemployed are truly constrained. Second, the results of this method may be used to test the hypothesis that the unemployed and underemployed are truly constrained. Third, the method introduced here may be used to estimate not only the incidence but also the duration of unemployment and underemployment.

The estimation procedure is an extension of Heckman's sample selection technique to the case where two correlated selection rules generate the sample. The procedure first estimates, by bivariate probit analysis, the factors affecting the probability of unemployment and underemployment. It then estimates labour supply behavior by excluding constrained workers from the sample and using the probit coefficients to correct for selectivity bias in the estimation. Finally, the paper employs a result from Hausman (1978) to test whether use of the standard least squares approach to labour supply estimation results in serious biases in parameter estimates.

The outline of the paper is as follows: Section 2 critically examines previous approaches to the problem of estimating labor supply functions when some workers are underemployed or unemployed. Section 3 discusses the extension of Heckman's procedure and illustrates how it may be used to obtain not only consistent estimates of the labour supply parameters but also consistent estimates of the equations describing the duration of unemployment and underemployment. Section 4 provides joint estimates of the factors affecting the probability of being unemployed or underemployed.

Section 5 employs the results of Section 3 and 4 to estimate the labour supply function. Significant evidence of selection bias from excluding underemployed workers is found while little evidence of selection bias from excluding unemployed workers is apparent. However, the two selection rule extension of

Heckman's approach produces coefficients with large standard errors. Since it was not possible to reject the null hypothesis that excluding unemployed workers did not cause selection bias, the labour supply equation was re-estimated considering only selection bias due to excluding the underemployed. As a result the precision of the estimates improved considerably, and significant evidence remains that excluding the underemployed leads to selection bias. The evidence from comparing the new estimates with those obtained from the traditional approach of ignoring constrained workers supports the hypothesis that these workers are constrained. First, the new estimates differ from the traditional estimates in the direction that would be predicted if the unemployed and underemployed were constrained. Second, the traditional method of least squares produces several coefficients which appear to be significantly biased; this is especially true for variables such as education and unearned income. Least squares estimates are found to compound the effect of a variable on desired labour supply with the effect of the variable on unemployment or underemployment. For example, least squares overestimates the effect of education on desired labour supply because increased education significantly reduces unemployment and underemployment.

Section 6 contains the estimates of the duration of unemployment and underemployment equations. While the results of this section are imprecise, (due to the relatively small samples used for estimation), they are plausible, indicating that the approach to estimating duration equations introduced here may prove fruitful when applied to larger data sets. Section 7 concludes the paper.

## 2. Previous Approaches to Unemployed and Underemployed Workers in Labour Supply Analysis

In the past the problem of underemployed and unemployed workers in labour

supply analysis has been handled in several ways. Researchers have assumed that these workers were not constrained, excluded them from the sample, assumed unemployed hours were exogenous or assumed that these workers were truly constrained. Each of these approaches has problems of its own and each will be discussed below.

The most common approach to unemployed and underemployed workers has been to ignore constraints and treat these workers as if they are not constrained. To see the kind of econometric problems that can arise in this situation consider the case of an individual who is truly unemployed or underemployed. Letting  $H_i^S$ ,  $w_i$ ,  $y_i$ , and  $R_i$  be the desired hours of work, wage rate, unearned income, and a vector of taste variables for individual  $i$ , write his labour supply function as

$$H_i^S = h(w_i, y_i, R_i) + \epsilon_i, \quad (1)$$

where  $\epsilon_i$  is an error term representing omitted variables and an approximation error resulting from assuming that all individuals have the same supply function. In what follows it will be assumed that  $\epsilon_i$  is homoskedastic and independent of  $w_i$ ,  $y_i$  and  $R_i$ . Next, define the vector  $X_i' = (1, w_i, y_i, R_i')$  and assume for the sake of simplicity that  $h(\cdot)$  is linear in its arguments. Then rewrite (1) as

$$H_i^S = X_i' \beta + \epsilon_i. \quad (2)$$

Equation (2) can of course be estimated by standard regression techniques given the necessary data. However, for the unemployed and underemployed only actual hours  $H_i$  are available. Now by definition  $H_i < H_i^S$ , which implies that  $H_i - H_i^S < 0$ . Equation (2) may be rewritten as

$$H_i = X_i' \beta + \epsilon_i + (H_i - H_i^S) = X_i' \beta + v_i . \quad (3)$$

In the unlikely event that  $H_i - H_i^S$  is independent of  $X_i$ , using  $H_i$  instead of  $H_i^S$  as a dependent variable will simply lower the constant term. On the other hand, if the unemployed or underemployed are more likely to be young or black, or to have relatively lower levels of schooling or unearned income, (as is usually thought to be the case), then the independent variables will be correlated with the new error term  $v_i$  and the resulting parameter estimates will be inconsistent. To see what form the inconsistency will take, simply differentiate (3) with respect to one of the  $X_i$  to obtain

$$\frac{\partial H_i}{\partial X_{ij}} = \beta_j - \frac{\partial (H_i^S - H_i)}{\partial X_{ij}} . \quad (4)$$

Now  $\frac{\partial H_i}{\partial X_{ij}}$  will be the least squares coefficient from a regression on the full sample, while  $\frac{\partial (H_i^S - H_i)}{\partial X_{ij}}$  is the effect of the  $j$ th variable on a worker's unemployment or underemployment. Thus, if these workers are truly constrained, and a variable decreases (increases) unemployment or underemployment, the least squares coefficient will overstate (understate) the true effect of this variable on desired labour supply.

An alternate approach to the problem is to remove the unemployed and underemployed from the sample. Wales and Woodland (1976), (1977) removed underemployed workers from the sample while DaVanzo, De Tray and Greenberg (1976) removed the unemployed from the sample.<sup>2</sup> However, censoring (removing) these workers from the sample will lead to inconsistent parameter estimates (Heckman (1979)) if the probability of being underemployed or unemployed is correlated with  $\epsilon_i$  and the independent variables. Moreover, under certain plausible assumptions the bias in the censored sample estimates will go in

the same direction as the bias in the full sample estimates; thus it is incorrect to argue that if the two sets of coefficients look the same, no problems arise from ignoring or censoring constrained workers.

An interesting approach to the problem of unemployed workers is provided by Rea (1974) and Kalachek, Mellow and Raines (1978). Rea's procedure ignores underemployment but does address unemployment, and may be intuitively explained in the following manner.<sup>3</sup> Assume desired hours for the unemployed equal actual hours plus some fraction  $\alpha$  of unemployed hours. Then

$$H_i^S = H_i + \alpha U_i, \quad (5a)$$

or

$$H_i = X_i' \beta - \alpha U_i + \epsilon_i, \quad (5b)$$

where  $U_i$  is individual  $i$ 's unemployed hours.<sup>4</sup> Equation (5) can be estimated by standard regression techniques, but such estimation presents several problems. First, because it ignores the underemployed, it is vulnerable to all of the biases described above. Second, it constrains  $\alpha$  to be the same for all workers and independent of  $X_i$ . Third, if the incidence or duration of unemployment is correlated with the independent variables and  $\epsilon_i$ , simultaneous equation bias will occur and estimates of  $\beta$  and  $\alpha$  will be inconsistent.

Finally, Ham (1980, Chapter 1) avoided some of the problems of previous studies by using a Tobit-like estimation scheme where workers faced an upper bound equation on the hours they could work. However this model is only valid if the underemployed and unemployed are truly constrained; otherwise it will produce inconsistent parameter estimates.

Of course, some readers may object to this treatment of previous studies on the grounds that either these constrained workers comprise too small a fraction

of the labor force to be of any practical importance, or that they are not truly constrained. The first objection is not valid, since in a sample of prime aged males taken from the University of Michigan's Panel Study of Income Dynamics (PSID), (the data set employed below), the fraction of the sample either underemployed or unemployed ranged from .192 in 1973 to .282 in 1970 over the period 1967-1974.<sup>5</sup>

The view that these workers are constrained is accepted by some economists and is rejected by others. As a result, it is important to have an estimation scheme which is valid independently of whether these workers are truly constrained. Fortunately, the method presented in Section 3 meets this requirement as well as avoiding the problems of previous approaches.

### 3. A Selection Rule Approach to Estimating Labour Supply Functions

The procedure used in this paper is to exclude unemployed and underemployed workers from the sample when estimating labour supply behavior and then to adjust the estimation scheme to avoid sample selection bias. This approach is an extension of Heckman's work to the case where workers are excluded from the sample on the basis of two correlated selection rules.<sup>6</sup> The paper estimates two rather than one selection rule(s), since the equation describing the probability of unemployment may differ from that describing underemployment (this is confirmed below) and Poirier (1980) has argued persuasively that combining two selection rules into one will produce a hybrid model and inconsistent parameter estimates. (Indeed, many economists appear to believe that unemployed, but not underemployed, workers are potentially constrained and therefore believe the two groups are made up of significantly different types of workers.)



The extension of Heckman's work is straight-forward, and may be found in Ham (1980) or Poirier (1980).<sup>7</sup> The goal is to estimate  $\beta$  in the equation

$$y_{1i} = X'_{1i} \beta + \varepsilon_{1i} \quad (6)$$

for the case where data on  $y_1$  (desired hours) are available only if the following two latent variables are both positive

$$y_{2i} = X'_{2i} \gamma_2 + \varepsilon_{2i} \quad (\text{no unemployment}) \quad (7)$$

and

$$y_{3i} = X'_{3i} \gamma_3 + \varepsilon_{3i} \quad (\text{no underemployment}) . \quad (8)$$

In (6), (7), (8),  $X_1$ ,  $X_2$ , and  $X_3$  are vectors of independent variables;  $\beta$ ,  $\gamma_2$ , and  $\gamma_3$  are vectors of coefficients; and  $\varepsilon_1$ ,  $\varepsilon_2$ ,  $\varepsilon_3$  are potentially correlated random variables. Exactly as in Heckman's analysis, the crucial issue is the expectation of  $y_{1i}$  conditional on  $y_{2i} > 0$ ,  $y_{3i} > 0$ . Assume that  $(\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}) \sim \text{iid } N(0, V)$ , where

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \rho \\ \sigma_{13} & \rho & 1 \end{pmatrix} .$$

Then,

$$E(y_{1i} | y_{2i} > 0, y_{3i} > 0) - X'_{1i} \beta = \sigma_{12} \lambda_{2i} + \sigma_{13} \lambda_{3i} , \quad (9)$$

where:  $\lambda_{2i} = \phi(Z_{2i}) \phi(Z_{2i}^*) / F(Z_{2i}, Z_{3i}, \rho)$  ;

$\lambda_{3i} = \phi(Z_{3i}) \phi(Z_{3i}^*) / F(Z_{2i}, Z_{3i}, \rho)$  ;

$$Z_{2i} = X'_{2i} \gamma_2 ; \quad Z_{3i} = X'_{3i} \gamma_3 ;$$

$$Z_{2i}^* = (Z_{3i} - \rho Z_{2i}) / (1 - \rho^2)^{1/2} ;$$

$$Z_{3i}^* = (Z_{2i} - \rho Z_{3i}) / (1 - \rho^2)^{1/2} ;$$

and  $\phi(\cdot)$ ,  $\Phi(\cdot)$  and  $F(\cdot)$  are the univariate standard normal density function, the univariate standard normal distribution function and the bivariate standard normal distribution function respectively.<sup>8</sup>

If  $\sigma_{12}$  and  $\sigma_{13}$  are not both equal to zero, the expectation in (9) is not equal to zero and least squares estimation of (6) will lead to exactly the same sort of specification error biases that Heckman described in the one selection rule case. Of course, given Heckman's results, the solution to the problem is immediate. Estimate the parameters of the selection rule equations by bivariate probit analysis and then use these parameter estimates to form consistent estimates  $\hat{\lambda}_{2i}$  and  $\hat{\lambda}_{3i}$  of  $\lambda_{2i}$  and  $\lambda_{3i}$ . Then least squares can be used to obtain consistent estimates of  $\beta$ ,  $\sigma_{12}$  and  $\sigma_{13}$  by estimating

$$y_{1i} = X'_{1i} \beta + \sigma_{12} \hat{\lambda}_{2i} + \sigma_{13} \hat{\lambda}_{3i} + \tilde{\epsilon}_{1i} , \quad (10)$$

$$\text{where } \tilde{\epsilon}_{1i} = \epsilon_{1i} + \sigma_{12}(\lambda_{2i} - \hat{\lambda}_{2i}) + \sigma_{13}(\lambda_{3i} - \hat{\lambda}_{3i}) .$$

However, the least squares estimates of  $\sigma_1^2$  and the standard errors will be inconsistent. Consistent estimators for these parameters are given in Appendix 2.

The present study uses this approach to estimate the desired labor supply, and it is important to note that the selection-rule approach is valid under much more general conditions than the previous approaches of simply using least squares on the full or censored samples. For example, if the selection rules are correctly specified this approach will produce consistent,

albeit inefficient, estimates even if none of the unemployed and underemployed are constrained. (When least squares is estimated on the full sample one implicitly assumes that these workers are not constrained.) The intuition is straight-forward: least squares on the full sample constrains the hours worked equation for the unemployed and underemployed to be the same as that for the nonconstrained workers. The selection rule approach imposes no such constraints on the parameter estimates. Clearly the selection rule approach is consistent if no selection bias is present, (the assumption made when least squares is used on the censored sample), since it simply estimates two parameters,  $\sigma_{12}$  and  $\sigma_{13}$ , which are truly zero.

Since the selection rule estimates nest least squares on the full sample, it is possible to use the lemma from Hausman (1978) to test whether any labor supply parameters are significantly biased from using least squares on the full sample. From Hausman's lemma, the variance of the difference between a selection rule and least squares-coefficient is simply the selection rule variance minus the least squares variance. Using this variance, it is possible to test whether any of the differences are significant. If a significant difference is found between the selection rule and full sample least squares estimates, this provides evidence supporting the hypothesis that the unemployed and underemployed are constrained, since if they were not constrained the labour supply estimates should not be affected by moving to the selection rule procedure.

The selection rule approach is also consistent with several different theories of unemployment. For example, if a search-theoretic explanation of unemployment (e.g. Mortenson (1970)) is accepted, the unemployment incidence equation should be interpreted as a reduced form version of the equation estimated by Kiefer and Neumann (1979). Alternatively, if one believes in a

model where workers voluntarily sort themselves into jobs with different probabilities of layoffs and underemployment and different compensating wage differentials (Abowd and Ashenfelter (1979)), the probit equations should be viewed as reduced-form sorting equations. Finally, the approach taken here is clearly consistent with a Keynesian theory (e.g. Malinvaud (1977)), where workers do not anticipate their unemployment but differ in their unemployment probabilities.

Moreover, the selection rule method can be used to form duration of unemployment and underemployment equations. Assume that the probit equations and the desired hours equation have been estimated. Then the expected desired hours for an unemployed worker can be written as

$$E(H^S | y_{2i} < 0) = X'_{1i} \hat{\beta} - \hat{\sigma}_{12} \lambda^*_{2i}, \quad (11)$$

$$\text{where } \lambda^*_{2i} = \phi(Z_{2i})/\phi(-Z_{2i}).$$

The worker's actual hours worked may be described by the equation<sup>9</sup>

$$E(H_i | y_{2i} > 0) = X'_{4i} \theta + \alpha_2 \lambda^*_{2i}, \quad (12)$$

where  $X'_{4i}$  is a vector of independent variables and  $\theta$  is a vector of parameters.

After estimating (12) and defining his true unemployment as the difference between the hours he wanted to work and the hours he actually worked, the unemployment duration equation becomes<sup>10</sup>

$$\begin{aligned} E(U_i | y_{2i} < 0) &= E(H^S_i - H_i | y_{2i} < 0) \\ &= X'_{1i} \hat{\beta} - X'_{4i} \hat{\theta} - (\hat{\sigma}_{12} + \hat{\alpha}_2) \lambda^*_{2i}. \end{aligned} \quad (13)$$

Estimates from (13) will differ from a traditional duration of unemployment equation. The latter measures the effect of the independent variables on the length of an unemployment spell. Equation (13) measures the effect of the independent variables on the unemployed hours that a worker could not compensate for by working longer hours after his unemployment spell ended (or

before it began, if the spell was anticipated).<sup>11</sup> Clearly, both methods of estimating the duration of unemployment equation convey useful information.

A similar equation may be estimated for the duration of underemployment. To the best of my knowledge, the equation presented below is the first attempt to measure the duration of underemployment.

Thus the approach taken here provides consistent estimation of the equations describing: desired labour supply; the incidence and duration of unemployment; and the incidence and duration of underemployment. Section 4 presents the estimates of the incidence equations. Section 5 contains the estimates of the labour supply function. The estimates of the duration equations are contained in Section 6.

#### 4. Estimating the Incidence of Unemployment and Underemployment

This section estimates the factors affecting the probability of an individual experiencing unemployment or underemployment. An individual is assumed to experience unemployment if

$$y_{2i} = X'_{2i} \gamma_2 + \epsilon_{2i} \quad (7')$$

is non-positive, and is assumed to experience underemployment if

$$y_{3i} = X'_{3i} \gamma_3 + \epsilon_{3i} \quad (8')$$

is non-positive. Assuming that  $(\epsilon_{2i} \ \epsilon_{3i})' \sim \text{iid } N(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$ , the bivariate probit likelihood function is

$$L = \prod_{y_2 > 0, y_3 > 0} F(Z_2, Z_3, \rho) \prod_{y_2 > 0, y_3 < 0} F(Z_2, -Z_3, -\rho) \\ \prod_{y_2 < 0, y_3 > 0} F(-Z_2, Z_3, -\rho) \prod_{y_2 < 0, y_3 < 0} F(-Z_2, -Z_3, \rho) \quad (14)$$

(In (14) the subscript  $i$  has been dropped for notational convenience).

The parameters of the bivariate probit function were estimated from a sample of 837 prime-aged males from the PSID. An individual was included in the sample if:

1. he was either black or white and between the ages of 25 and 50 in 1967;
2. he was not a member of the nonrandom poverty sub-sample in 1967;
3. he stayed in the sample for all the eight years 1967-1974;
4. he had not retired from the labor force.

Criterion 4 was adopted out of expediency and eliminated 49 individuals from the sample. It is clearly an example of truncating on the basis of a dependent variable, although it is a truncation of less than six per cent of the sample.<sup>12</sup>

Underemployed and unemployed workers were identified in the following manner. Each year the PSID contained retrospective questions on whether there was more work available (in the previous year) on a worker's job (any of his jobs) so that he could have worked more if he had wanted to, and if not, whether he had wanted to work more. A worker who answered no to the first question and yes to the second question was classified as underemployed. A worker was classified as unemployed if he experienced one or more days of unemployment during the year.<sup>13</sup> Neither of these classifications is ideal. First, a worker classified as underemployed may really want to work more hours, but only at an overtime premium. A much better question would ask a worker what his marginal wage was last year, and at what marginal wage would he have been willing to work more hours.<sup>14</sup> Second, the PSID combined days of unemployment with days on strike, and there is no straight-forward means of separating the two.

Keeping these potential data problems in mind, the next step was to choose the independent variables for the probit equations. All variables to be included in the labour supply function were also included in  $X_2$  and  $X_3$ . On the basis of previous studies and the relevant theory, the following variables were included in the labour supply function:

1. after-tax net wage;
2. after-tax unearned income;
3. dummy variable coded 1 if the individual suffered from a health limitation;
4. dummy variable coded 1 if the individual was married;
5. age;
6. education (in years);
7. dummy variable coded 1 if the individual was black;
8. number of children.

The use of the after-tax wage and unearned income raises a simultaneity problem, since both depend on hours worked. Two sets of instrumental variables were employed to deal with this issue.<sup>15</sup> First, lagged values of the after-tax wage and unearned income, along with the other independent variables in the labor supply equation, were used as instruments. Thus identification is being obtained by a strong independence - over-time assumption. Second, the current after-tax wage and unearned income, evaluated as if the husband worked the sample mean hours of 2351 and his wife did not work, were also used along with the other independent variables.<sup>16</sup> This is a modification of Rosen's (1976) suggestion, and is based on the assumption that the gross wage and unearned income are exogenous.<sup>17</sup> The latter assumption is a standard one in labour supply studies that deal with taxes (Burtless and Hausman (1978), Hurd (1976) and Wales and Woodland (1979)).

A dummy variable, coded 1 for individuals belonging to a trade union, was also included in  $X_2$  and  $X_3$  for two reasons. First, using a different data set, Medoff (1979) found that trade union members were more likely to experience unemployment, and this study investigates whether his result also applies to underemployment.<sup>18</sup> Second, the use of the union dummy may alleviate partially the problem in the unemployment data noted above.

Turn now to the results of estimating the incidence equations for unemployment and underemployment contained in Table 1.<sup>19</sup> Independent of which instruments are used, the correlation coefficient is estimated to be quite small, although positive and significantly different from 0.

Using either set of instruments, three variables significantly affect the probability of experiencing a spell of unemployment: trade union membership; education; and the number of children.<sup>20</sup> The interpretation of the union coefficient is unclear because of data problems. The result that increased education significantly reduces the probability of unemployment may reflect those with more schooling choosing (on average) occupations with less variance in employment. Finally, the result that the probability of unemployment significantly increases with the number of children is somewhat surprising, and there appears to be no theoretical explanation for it.<sup>21</sup>

Several variables significantly affect the probability of underemployment. In the equation using the lagged values as instruments, the higher the wage, holding age and education constant, the higher the probability of a worker being underemployed. This may reflect workers demanding a compensating wage differential in jobs with a relatively high probability of underemployment.<sup>22</sup> Independent of which set of instruments is used, an increase in unearned income also significantly reduces the probability of underemployment.



TABLE 1  
ESTIMATING THE PROBABILITY OF BEING UNEMPLOYED OR UNDEREMPLOYED  
BY BIVARIATE PROBIT ANALYSIS

Variable	Structural Equation 1 <sup>a/</sup>		Structural Equation 2 <sup>b/</sup>	
	Unemployed <sup>c/</sup>	Underemployed <sup>d/</sup>	Unemployed <sup>c/</sup>	Underemployed <sup>d/</sup>
Constant	-.0788 (.5682)	.1843 (.5525)	.0145 (.5660)	.0536 (.5353)
Wage (\$)	-.0414 (.0619)	.1170* (.0570)	.0010 (.0290)	-.0080 (.0360)
Unearned Income (0000's of \$)	-.2512 (.4402)	-1.8997* (.6312)	-.0737 (.3316)	-.7811** (.4006)
Trade Union Dummy (Union = 1)	.7402* (.1374)	.3176* (.1179)	.7207* (.1340)	.4229* (.1113)
Number of Children	.0784* (.0376)	.0581 (.0388)	.0777* (.0383)	.0515 (.0400)
Health Limitation Dummy	.0354 (.2675)	.1359 (.2250)	.0924 (.2502)	.0663 (.2205)
Marital Dummy (Married = 1)	-.4030 (.2790)	-.1351 (.2628)	-.4374 (.2737)	-.0056 (.2697)
Age (10 yr. units)	-.0884 (.0916)	-.1224 (.0878)	-.1199 (.0881)	-.0809 (.0834)
Education (years)	-.0541* (.0248)	-.0789* (.0224)	-.0642* (.0233)	-.0678* (.0205)
Race Dummy (Black = 1)	-.0143 (.2385)	.2937 (.2035)	.0303 (.2347)	.2701 (.2025)
$\rho$		.1986* (.0827)		.1944* (.0812)
Log L	-633.33		-638.32	

Note: Standard errors in parenthesis.

<sup>a/</sup> After-tax wage and unearned income from the previous year used as instruments.

<sup>b/</sup> After-tax wage and unearned income evaluated at mean hours used as instruments.

<sup>c/</sup> Dependent variable coded 1 if unemployed, 0 otherwise.

<sup>d/</sup> Dependent variable coded 1 if underemployed, 0 otherwise.

\* Significant at the 5% level.

\*\* Significant at the 10% level.

There are two possible explanations for this result. First, flexibility over hours of work may be a normal good. Second, unearned income may be acting as a proxy for the absence of constraints in the past. Individuals with substantial assets (and thus a high level of unearned income) are likely to be those who have worked long hours in the past; those who have worked long hours in the past are less likely to have been constrained. If the probability of underemployment is positively correlated over time, the unearned income coefficient is likely to be negative.<sup>23</sup> The significantly positive coefficient on trade union membership indicates that Medoff's result is applicable to underemployment. Finally, it is interesting to note that using both sets of instruments, increased years of schooling also reduces the probability of underemployment.

The results of this section indicate that ignoring or excluding the underemployed and unemployment may indeed lead to inconsistent estimates of important labour supply parameters. The worker's wage, unearned income, and education all significantly affect the probability of his being constrained, and this is a necessary condition for the coefficients of these variables to be inconsistent in a labour supply regression on the full or censored sample.

Moreover, in the case of the full sample least squares estimates, it is possible to predict the direction of the inconsistency that will occur if the unemployed and underemployed are truly constrained, since the probit coefficients may be used to classify which independent variables increase (or decrease) unemployment and underemployment.<sup>24</sup> The variables were classified in the following (arbitrary) manner. First, if a variable's coefficient had the same sign in both the unemployment and underemployment probit equations, and at least one of the estimated coefficients had an asymptotic normal statistic greater than 1, then the direction of the inconsistency was unambiguous.

Second, if a variable's coefficient had opposite signs in the probit equations, but one had an asymptotic normal statistic greater than 1 in absolute value, the latter coefficient was assumed to dominate. Otherwise, the direction of the inconsistency could not be predicted.

On the basis of these rules, (using either set of instruments), least squares on the full sample should overestimate the coefficients on unearned income, marital status, age and education while underestimating the coefficients for number of children and race. Further, when the lagged values are used as instruments, least squares should underestimate the wage coefficient.

#### 5. Alternate Estimates of the Labour Supply Function

In this section, the labour supply function is first estimated (using both sets of instruments) by three methods: least squares on the full sample; least squares on the censored sample; and the two selection rule approach on the censored sample. Consider the first two columns of Table 2 which contain the results of using least squares on the full and censored samples when the lagged values are used as instruments.

The constant is estimated at between 2380 and 2500 hours, the wage coefficient is very significantly negative and the unearned income coefficient is strongly positive. The race coefficient is negative but insignificant in both equations, while the number of children coefficient is very small and insignificant in the full sample regression. The age coefficient is significant at the 10% level while the education coefficient is very significantly positive.

Consider next the two-selection rule estimates in column three (before the standard errors have been corrected). The coefficient of the  $\lambda$  term for

TABLE 2  
ALTERNATE ESTIMATES OF THE LABOR SUPPLY FUNCTION  
(LAGGED VALUES USED AS INSTRUMENTS)

Variable	Method of Estimation			
	2SLS (Full Sample)	2SLS (Censored Sample)	2SLS 2 Selection Rule (Censored Sample - Uncorrected Standard Errors)	2SLS 2 Selection Rule (Censored Sample - Corrected Standard Errors)
Constant	2.3862* (.1820)	2.4910* (.1912)	3.5612* (.4351)	3.5612* (.8461)
Wage (\$)	-.1792* (.0161)	-.1848* (.0166)	-.1170* (.0349)	-.1170 (.0735)
Unearned Income (0000's of \$)	.8105* (.1125)	.6894* (.1042)	.1097 (.2834)	.1097 (.5939)
Race Dummy (Black = 1)	-.1308 (.0851)	-.1081 (.0973)	.1963 (.1582)	.1963 (.2909)
Number of Children	.0002 (.0132)	.0212 (.0136)	.0300* (.0149)	.0300 (.0312)
Health Limitation Dummy	-.3148* (.0850)	-.3022* (.0881)	-.2240* (.0935)	-.2240 (.1958)
Marital Dummy (Married = 1)	.1856** (.0974)	.2140* (.1009)	.2824* (.1125)	.2824 (.2284)
Age	-.0521** (.0302)	-.0550** (.0312)	-.1235* (.0474)	-.1235 (.0821)
Education	.0492* (.0074)	.0452* (.0082)	.0065 (.0212)	.0065 (.0423)
$\lambda_2$ (Unemployed)	-- --	-- --	.6563 (.4678)	.6563 (.8681)
$\lambda_3$ (Underemployed)	-- --	-- --	-1.8971* (.8079)	-1.8971 (1.5439)
R <sup>2</sup>	.213	.329	.331	.331
S.E.E.	.571	.506	.505	1.271
Sample Size	837	618	618	618

(Dependent Variable: Annual Hours Worked in 000's in 1971)

Note: Standard errors in parenthesis.

\* Significant at 5% level.

\*\* Significant at 10% level.

the underemployed is significantly different from zero while the coefficient of the  $\lambda$  term for the unemployed is not significantly different from zero.<sup>25</sup> Thus sample selection bias appears to be a more important problem when excluding underemployed workers than when excluding unemployed workers. The constant term is now more than 1000 hours greater than either least squares estimate. While this parameter is of little economic interest, its increase when the least squares estimates are replaced by the selection rule estimates supports the conjecture of Section 2 that ignoring constrained workers will lower the constant. The wage coefficient is one-third less in absolute value than the coefficient of either of the least squares estimates. Moreover, the income coefficient falls by more than a factor of six. The race coefficient is now positive, a net change of over 300 hours from the least squares estimates. The age coefficient has doubled in absolute value while the education coefficient has become very small. Further, only the marital status coefficient fails to move in the direction predicted by the constrained hypothesis.

However, once the corrected standard errors are used, the precision of the estimates falls to the point where they are of little value in making inferences. This increase in the estimated standard errors may be due to the correlation coefficient between the underemployment and labour supply equations being estimated at -1.5, (the underemployed have a below-average taste for leisure), since this parameter is quite important in the correction to the standard errors.

The labour supply results using the after-tax wage and unearned income evaluated at mean hours as instruments are similar to those obtained with the lagged values as instruments. First, the coefficient on the  $\lambda$  term for the unemployed is still much smaller than that for the underemployed, although neither is significantly different from zero. Second, again all but the

TABLE 3  
ALTERNATE ESTIMATES OF THE LABOR SUPPLY FUNCTION  
(AFTER-TAX WAGE AND UNEARNED INCOME EVALUATED  
AT MEAN VALUES USED AS INSTRUMENTS)

Variable	Method of Estimation			
	2SLS	2SLS	2SLS	2SLS
	(Full Sample)	(Censored Sample)	2 Selection Rule (Censored Sample - Uncorrected Standard Errors)	2 Selection Rule (Censored Sample - Corrected Standard Errors)
Constant	2.4077* (.1754)	2.4703* (.1891)	2.8932* (.3234)	2.8932* (.3463)
Wage (\$)	-.1369* (.0098)	-.1715* (.0122)	-.1683* (.0123)	-.1683 (.0135)
Unearned Income (0000's of \$)	.5086* (.0857)	.4206* (.0794)	.3242* (.1511)	.3242* (.1615)
Race Dummy (Black = 1)	-.1289 (.0825)	-.1263 (.0972)	-.0499 (.1375)	-.0499 (.1439)
Number of Children	.0014 (.0128)	.0235** (.0135)	.0351* (.0148)	.0351* (.0160)
Health Limitation Dummy	-.2968* (.0822)	-.3120* (.0879)	-.2863* (.0892)	-.2863* (.0951)
Marital Dummy (Married = 1)	.1505 (.0942)	.2109* (.1006)	.2001** (.1300)	.2001** (.1387)
Age	-.0671* (.0285)	-.0530** (.0306)	-.0689* (.0320)	-.0689* (.0346)
Education	.0457* (.0068)	.0463* (.0076)	.0291* (.0140)	.0291** (.0150)
$\lambda_2$ (Unemployed)	-- --	-- --	-.1528 (.5186)	-.1528 (.5394)
$\lambda_3$ (Underemployed)	-- --	-- --	-.4637 (.7325)	-.4637 (.7667)
R <sup>2</sup>	.252	.328	.327	.327
S.E.E.	.591	.506	.501	.577
Sample Size	837	618	618	618

(Dependent Variable: Annual Hours Worked in 000's in 1971)

Note: Standard errors in parenthesis.

\* Significant at 5% level.

\*\* Significant at 10% level.

marital coefficient move in the direction predicted by the probit equations, although the difference between the coefficients is, on average, smaller. Finally, the corrected standard errors for the two selection rule parameter estimates are still large, though not nearly as large as when the lagged values are used as instruments.

In an attempt to reduce the standard errors of the selection rule estimates, the hypothesis that sample selection bias is not significant when excluding the unemployed was tested and could not be rejected. Thus, the coefficient on  $\lambda_2$  was constrained to equal zero and the selection rule equations re-estimated, using Heckman's one selection rule approach.<sup>26</sup> Note that if the two-selection rule approach had not been estimated, but rather the study had begun by using a one-selection rule approach, it would have been necessary to impose the constraint  $\sigma_{12}=0$  without testing it.

From the results in the first two columns of Table 4, it is clear that imposing this constraint improves the precision of the selection rule estimates, especially when the lagged values are used as instruments. Again, all but one of the selection rule coefficients for both sets of instruments continue to move in the direction predicted by the probit coefficients, although now the differences are less pronounced. Further, the  $\lambda$  term for the underemployed is now significantly different from zero when the alternative instruments are used.

Columns 3 and 4 of Table 4 contain the estimated differences between the full sample least squares coefficients and the selection rule coefficients.<sup>27</sup> These results indicate that several coefficients may be biased by ignoring constrained workers. When the lagged values are used as instruments, the constant and the coefficients of unearned income, number of children and education are

TABLE 4

ONE SELECTION RULE ESTIMATES OF THE LABOUR SUPPLY FUNCTION AND DIFFERENCES  
WITH THE FULL SAMPLE LEAST SQUARES ESTIMATES

Variable	One Selection Rule Labour Supply Estimates		Differences with Full Sample Least Squares Estimates	
	Lagged Instruments	Alternate <sup>a/</sup> Instruments	Lagged Instruments	Alternate <sup>a/</sup> Instruments
Constant	3.1907* (.4015)	2.9258* (.3171)	.8045* (.3583)	.5181* (.2648)
Wage (\$)	-.1542* (.0281)	-.1683* (.0144)	.0249 (.0231)	-.0315* (.0107)
Unearned Income (0000's of \$)	.4169* (.2126)	.2987* (.1146)	-.3936* (.1808)	-.2100* (.0766)
Race Dummy (Black = 1)	.0385 (.1276)	-.0308 (.1148)	.1692** (.0954)	.0981 (.0802)
Number of Children	.0338** (.0187)	.0351* (.0167)	.0336* (.0133)	.0337* (.0109)
Health Limitation Dummy	-.2563* (.1103)	-.2844* (.0998)	.0584 (.0707)	.0124 (.0571)
Marital Dummy (Married = 1)	.2033** (.1226)	.2165** (.1134)	.0177 (.0752)	.0660 (.0638)
Age (10 yr. units)	-.0939* (.0430)	-.0698** (.0361)	-.0418 (.0307)	-.0027 (.0223)
Education (years)	.0138 (.0178)	.0273* (.0130)	-.0354* (.0162)	-.0185** (.0111)
$\lambda$ (underemployed)	-.8225* (.3818)	-.5922* (.2965)	--	--
R <sup>2</sup>	.343	.326	--	--
S.E.E.	.694	.616	--	--
Sample Size	618	618	--	--
X <sup>2</sup> Statistic	--	--	12.4	25.5

(Dependent Variable: Annual Hours worked in 000's in 1971)

Note: Corrected standard errors in parenthesis.

a After-tax wage and unearned income evaluated at mean hours used as instruments.

\* Significant at 5% level.

\*\* Significant at 10% level.



all significantly biased at the 5% level, while the difference in the coefficient for race is significant at the 10% level. When the instruments contain the after-tax wage and unearned income evaluated at the sample mean, the constant and the wage, unearned income and number of children coefficients are biased at the 5% level. The education coefficient is biased at the 10% level. The relevant chi-square statistics for testing the overall differences in the coefficients are 12.4 and 25.5 when the lagged values and mean values respectively are used as instruments. The former statistic indicates that the differences are significant only at the 20% level (approximately) while the latter statistic indicates that the differences are significant at the .5% level. Thus while the non-constrained model could not be decisively rejected for both sets of instruments, the results are favourable to the hypothesis that the unemployed and underemployed are constrained.

Before turning to the duration equations there are three caveats to the above results that should be noted. First, when the lagged values are used as instruments, the correlation coefficient between the underemployment equation and the labour supply equation is estimated at -1.18 (it is estimated at -.96 using the alternate instruments). While this estimate is not that much greater than one, its size is nevertheless disturbing. Second, while the labour supply results tend to improve when the selection rule approach is used, all the labour supply equations presented here have negative own substitution effects, a violation of the classical restrictions. Finally, while the estimates do not appear to be too sensitive to which instruments are used, it is important to try alternative instruments and alternative years of data. Carrying out this latter step, along with improving the precision of the selection rule estimates, should provide an important avenue for future research.

## 6. Estimating the Duration of Unemployment and Underemployment

Table 5 contains the estimates of the actual hour and duration equations when the lagged values are used as instruments. The estimates are quite imprecise, probably reflecting both the small sample sizes used to estimate the actual hours equations and the fact that the duration coefficients are the difference between the desired and actual hours coefficients.<sup>28</sup> However, the results do appear to be plausible. First, most coefficients in the duration equation have the same sign as the respective probit coefficients.<sup>29</sup> Second, the coefficients on  $\lambda^*$  indicate that an individual who is more likely to have an unemployment (underemployment) spell is also more likely to have a longer duration of unemployment (underemployment).<sup>30</sup> Larger data sets are necessary before this method can be fully evaluated and exploited.

## 7. Conclusion

This paper has presented and implemented a new method of labour supply estimation which is valid under much more general assumptions than previous approaches. This method is based on a generalization of Heckman's sample selection procedure. It estimates labour supply behavior for workers who are neither unemployed nor underemployed while correcting the estimates for selection bias. The empirical results suggest that selection bias is an important problem when excluding underemployed workers but not when excluding unemployed workers, and that simply censoring all constrained workers leads to significant biases in many labour supply parameters. The empirical results are also favourable to the hypothesis that the unemployed and underemployed are constrained; the difference between the full sample least squares estimates and the selection rule

TABLE 5  
ESTIMATING THE DURATION OF UNEMPLOYMENT  
AND UNDEREMPLOYMENT  
(LAGGED VALUES USED AS INSTRUMENTS)

	Actual Hours for		Duration of	
	Unemployed Workers (2SLS)	Underemployed Workers (2SLS)	Unemployment	Underemployment
Constant	2.6542* (.6304)	2.5704* (.6767)	.5365 (.7379)	.6203 (.7706)
Wage (\$)	-.0863 (.0578)	-.4163* (.1280)	-.0680 (.0641)	.2620* (.1259)
Unearned Income (0000's of \$)	1.0486 (.6780)	4.0254* (1.4386)	-.6317 (.7121)	-3.6085* (1.4263)
Race Dummy (Black = 1)	-.1971 (.2682)	-.1060 (.2529)	.2355 (.2960)	.1444 (.2518)
Number of Children	-.0242 (.0513)	-.0350 (.0455)	.0579 (.0546)	.0687 (.0452)
Health Limitation Dummy	-.3441 (.3334)	-.2824 (.2645)	.0878 (.3526)	.0261 (.2606)
Marital Dummy (Married = 1)	-.1674 (.3563)	.6705** (.3528)	.3707 (.3769)	-.4672 (.3501)
Age (10 yr. units)	.0485 (.1132)	.0077 (.1191)	-.0454 (.1200)	-.1016 (.1174)
Education $\lambda^*_{a}$ (years)	.0757* (.0322)	.0741** (.0430)	.0618** (.0364)	-.0602 (.0430)
	-.5913* (.2875)	-.5659 (.5311)	.5912* (.2875)	1.4137* (.5873)
R <sup>2</sup>	.188	-.417	--	--
S.E.E.	.835	.846	--	--
Sample Size	102	151	--	--

(Dependent variable measured in 000's of hours)

Note: Corrected standard errors in parenthesis. The calculation of the standard errors in columns 3 and 4 incorporates the fact that the desired hour and actual hour coefficients are based on the same bivariate probit coefficients.

a/ The respective reduced-form probit coefficients are substituted into equation (11).

\* Significant at the 5% level.

\*\* Significant at the 10% level.

estimates for all but one of the variables is in the direction that one would predict under the hypothesis that the unemployed and underemployed were constrained. Moreover, it appears that using least squares on the full sample and ignoring constrained workers leads to significant biases in several labour supply parameters. This paper has also presented and implemented a method of estimating the duration of unemployment and underemployment.

# APPENDIX 1

## EMPLOYMENT EXPERIENCE OF MALES, 1967-74 (AGED 25-50 IN 1967)

Year	Not-Constrained	Percent of Sample		
		Underemployed & not Unemployed	Unemployed & not Underemployed	Underemployed & Unemployed
1967	79.7	9.3	7.4	3.6
1968	78.1	12.4	6.3	3.2
1969	75.8	14.4	6.1	3.6
1970	71.8	14.3	8.1	5.7
1971	73.8	14.0	8.1	4.1
1972	79.7	11.9	6.2	2.2
1973	80.8	10.5	6.1	2.6
1974	72.6	13.1	9.6	4.8

## APPENDIX 2

CONSISTENT ESTIMATION OF THE VARIANCE OF THE LABOR SUPPLY  
RESIDUAL AND THE VARIANCE-COVARIANCE MATRIX OF  
THE PARAMETER ESTIMATES

This appendix contains the expressions for the consistent estimators of  $\sigma_1^2$  and for the variance-covariance matrix of the two-selection rule parameter estimates.

The censoring causes the error terms in the censored sample to be heteroskedastic. Using Tallis (1961)

$$\begin{aligned}\sigma_{1i}^2 &= E(\varepsilon_{1i}^2 | y_{2i} > 0, y_{3i} > 0) = \sigma_1^2 - \sigma_{12}^2 Z_{2i} \lambda_{2i} - \sigma_{13}^2 Z_{3i} \lambda_{3i} \\ &\quad + \mu_i [2\sigma_{12}\sigma_{13} - \rho(\sigma_{12}^2 + \sigma_{13}^2)] - (\sigma_{12}\lambda_{2i} + \sigma_{13}\lambda_{3i})^2 \\ &= \sigma_1^2 + v_i\end{aligned}\tag{A-1}$$

where  $\mu_i = f(Z_{1i}, Z_{2i}, \rho)/F(Z_{1i}, Z_{2i}, \rho)$ , and  $f(\cdot)$  is the bivariate standard normal density function.

As a result, the least squares estimator of  $\sigma_1^2$  will be inconsistent. However, one may use the following generalization of Heckman's estimator for  $\sigma_1^2$

$$\hat{\sigma}_1^2 = (S - \sum \hat{v}_i)/N, \tag{A-2}$$

where  $S$  is the sum of squared residuals,

$N$  is the size of the censored sample, and

$\hat{v}_i$  equals  $v_i$  after parameter estimates have been substituted for their true values.

Calculating the correct variance-covariance matrix of the parameter estimates is a simple generalization of Heckman or Lee, Maddala and Trost (1980). Following the latter, write the residual of (10) as

$$\tilde{\epsilon}_{1i} = \epsilon_{1i} + \sigma_{12}(\lambda_{2i} - \hat{\lambda}_{2i}) + \sigma_{13}(\lambda_{3i} - \hat{\lambda}_{3i}) . \quad (A-3)$$

Then define a  $1 \times 1$  vector  $\delta' = (\gamma_1', \gamma_2', \rho)$  and take first-order approximations of  $\hat{\lambda}_{2i}$  and  $\hat{\lambda}_{3i}$

$$(\lambda_{2i} - \hat{\lambda}_{2i}) = G_{2i}'(\delta - \hat{\delta}) \quad (A-4a)$$

$$\text{and } (\lambda_{3i} - \hat{\lambda}_{3i}) = G_{3i}'(\delta - \hat{\delta}) \quad (A-4b)$$

where  $G_{2i} = \frac{\partial \lambda_{2i}}{\partial \delta}$  and  $G_{3i} = \frac{\partial \lambda_{3i}}{\partial \delta}$  are  $1 \times 1$  vectors.

Now let  $X_i^{*'} = (X_i', \hat{\lambda}_{2i}, \hat{\lambda}_{3i})$  and  $\beta^{*'} = (\beta', \sigma_{12}, \sigma_{13})$  be  $1 \times K$  vectors and  $X^{*'} = (X_1^{*'}, \dots, X_N^{*'})$  be the  $K \times N$  matrix of observations on the independent variables. Further, define the  $1 \times N$  vectors  $\tilde{\epsilon}_1' = (\tilde{\epsilon}_{11}, \dots, \tilde{\epsilon}_{1N})$  and  $\epsilon_1' = (\epsilon_{11}, \dots, \epsilon_{1N})$  as well as the  $1 \times 1$  vector  $C_i = (\sigma_{12} G_{2i} + \sigma_{13} G_{3i})$  and the  $1 \times N$  matrix  $C' = (C_1, \dots, C_N)$ .

Then<sup>31</sup>  $(\hat{\beta}^* - \beta^*) = (X^{*'} X^*)^{-1} X^{*'} \tilde{\epsilon}_1$

$$\stackrel{D}{=} (X^{*'} X^*)^{-1} X^{*'} [\epsilon_1 + C(\delta - \hat{\delta})] . \quad (A-5)$$

Ignoring covariance terms (Lee et. al. (1980), p.500) yields

$$\begin{aligned} \text{Var}(\hat{\beta}^*) &= (X^{*'} X^*)^{-1} X^{*'} [\text{var}(\epsilon_1) + \text{Cvar}(\hat{\delta}) C'] X^* (X^{*'} X^*)^{-1} \\ &= (X^{*'} X^*)^{-1} [X^{*'} \text{var}(\epsilon_1) X^* + X^{*'} \text{Cvar}(\hat{\delta}) C' X^*] (X^{*'} X^*)^{-1} \\ &= (X^{*'} X^*)^{-1} [B_1 + B_2 \text{var}(\hat{\delta}) B_2'] (X^{*'} X^*)^{-1} \end{aligned} \quad (A-6)$$

$$\text{where } B_1 = \sum_{i=1}^N \sigma_{li}^2 X_i^* X_i^{*'} \\ \text{and } B_2 = \sum_{i=1}^N X_i^* C_i'$$

Equation (A-5) is a computationally efficient expression for calculating the variance-covariance matrix because the largest matrix it involves storing in the computer has dimension  $l \times K$ .



## APPENDIX 3

## ESTIMATING THE DURATION OF EMPLOYMENT

## AND UNDEREMPLOYMENT

(AFTER-TAX WAGE AND UNEARNED INCOME EVALUATED AT

MEAN HOURS USED AS INSTRUMENTS)

	Actual Hours for		Duration of	
	Unemployed Workers (2SLS)	Underemployed Workers (2SLS)	Unemployment	Underemployment
Constant	2.6507* (.5882)	2.4033* (.4561)	.2751 (.6573)	.5224 (.5528)
Wage (\$)	-.0517* (.0171)	-.1145* (.0362)	-.1166* (.0226)	-.0538 (.0387)
Unearned Income (0000's of \$)	.1299 (.5530)	.2259 (.4285)	.1688 (.5649)	.0728 (.4418)
Race Dummy (Black = 1)	-.2299 (.2443)	.0516 (.1568)	.1991 (.2692)	-.0824 (.1916)
Number of Children	.0031 (.0420)	.0018 (.0298)	.0319 (.0453)	.0332 (.0337)
Health Limitation Dummy	-.3717 (.2884)	-.1138 (.1727)	.0872 (.3060)	-.1706 (.1971)
Marital Dummy (Married = 1)	-.3399 (.2952)	.1927 (.2386)	.5564** (.3163)	.0238 (.2619)
Age (10 yr. units)	-.0684 (.0983)	-.1185** (.0696)	-.0013 (.1040)	.0487 (.0776)
Education (years)	.0725* (.0305)	.0413** (.0234)	-.0452 (.3293)	-.0140 (.0265)
$\lambda^*_{a/}$	-.4563** (.2588)	.0533 (.2734)	.4563* (.2588)	.6456 (.3985)
R <sup>2</sup>	.258	.177	--	--
S.E.E.	.742	.526	--	--
Sample Size	102	151	--	--

(Dependent variable measured in 000's of hours)

Note: Corrected standard errors in parenthesis. The calculation of the standard errors in columns 3 and 4 incorporates the fact that the desired hour and actual hour coefficients are based on the same bivariate probit coefficients.

a/ The respective reduced-form probit coefficients are substituted into equation (11).

\* Significant at the 5% level.

\*\* Significant at the 10% level.

### Footnotes

- \* This paper is also being circulated as Institute for Policy Analysis Working Paper No. 8108 University of Toronto.
- \*\* An earlier version of the paper was presented at the Fourth World Congress of the Econometric Society in Aix-en-Provence, France, August 1980. This paper represents an extensive revision of Chapter 2 of my dissertation in the Department of Economics, Princeton University. I am indebted to my thesis committee, O. Ashenfelter, R. Quandt and J. Brown, for excellent comments and suggestions. I would also like to thank R. Blundell, G. Chamberlain, R. Gordon, J. Hausman, S. Nickell, D. Poirier and S. Rosen for helpful discussions or correspondence, as well as M. Gunderson, C. Hsiao, G. MacDonald, S. Rea and R. Winter for very useful comments on earlier drafts of the paper. None of the above are responsible for any errors. I also gained from discussions in seminars at Cornell University, Princeton University, Queen's University, and the University of Toronto. Generous financial support was provided by the Canada Council, Princeton University and the University of Toronto.
1. Representative examples of such work are Abowd and Ashenfelter (1979), Baily (1974), Benassy (1976), Dreze (1975), Hahn (1978) and Malinvaud (1977). Of course, these models differ substantially in their normative implications, since in some models the unemployment is anticipated while in others it is not.
  2. Wales and Woodland and Ham (1980, chapters 1 and 3) also considered the overemployed (those who wish to work fewer hours on their job). These workers are numerically much less important than the underemployed and unemployed, and are ignored here to keep the necessary computations within the feasible range.
  3. In fairness to Rea, Kalachek et.al., and DaVanzo et.al., their data sets did not allow them to identify underemployed workers.
  4. See Ashenfelter (1980) for an elegant derivation of a similar equation in the context of a system of commodity demand and labour supply equations.
  5. See Appendix 1 for more details. The methods used to identify unemployed and underemployed workers are described in Section 4.
  6. Catsiapis and Robinson (1978) consider the case of two independent selection rules.
  7. Poirier also provides an interesting problem where a two selection rule procedure is appropriate.
  8. Expressions for this conditional expectation may be found in Amemiya (1974b), Lee (1979) and Tallis (1961).
  9. Since the coefficients  $\theta$  are defined only for the case of  $y_{2i} < 0$ , it is a matter of choice whether or not to include the term  $\lambda_{2i}^*$  (Poirier and Ruud (1980)).

10. It would be preferable to estimate duration equations for the three mutually exclusive groups of workers: unemployed and not underemployed, underemployed and not unemployed; unemployed and underemployed. However, in the current sample the small number of workers in each category makes such a procedure impractical.
11. A simple example may clarify this distinction. Consider a worker who, at the beginning of the year, intends to work 200 hours per month or 2400 hours per year. He is unemployed for the first month and a half of the year, but works 2200 hours during the rest of the year to compensate for the unemployment. Then the traditional measure of the duration of unemployment is 300 hours, while the measure used here is only 200 hours.
12. Ham (1980, chapter 1) accounted for this truncation and found that it changed the labour supply estimates very little.
13. I am viewing the underemployed as being constrained in terms of their hours of work per week and the unemployed as being constrained in terms of their weeks of work per year, so that it makes sense for a worker to be unemployed but not underemployed. The estimation method proposed in Section 3 remains valid whether or not this interpretation is correct.
14. Simple nonlinearities in the budget constraint (such as an overtime premium) will not cause a worker to be classified as underemployed, since any worker so classified stated that he could not work any more hours on his job.
15. Two other possible procedures were not used. One possibility would be to use quadratic terms and cross-products of the other independent variables as instruments. Unfortunately, to implement the selection-rule procedure, one would have to substitute in the quadratic and cross-product terms in the probit equations and estimate reduced-form equations (Lee, Maddala and Trost (1980)). Carrying out the substitution yields a bivariate probit problem in over fifty parameters, and thus this approach was not used here. Alternatively, one could assume that variables such as age and education enter the wage and unearned income equations but not the labour supply equation. Since this procedure seems at least as arbitrary as the ones employed here, and because a principal goal of this study is to examine the effect of the selection rule procedure on the coefficients of variables such as age and education, this approach was not adopted either.
16. There remains one problem for correction in future work. By not including the wife's wage in the supply equation, I am ignoring any family labour supply considerations. Handling this problem correctly is quite difficult. Whether or not the wife works is endogenous, and the labour supply function will differ between husbands with working and nonworking wives. Olsen (1977) contains a complete discussion of this problem and a means of handling it.
17. In the labour supply equation it is better to use these variables as instruments rather than simply substituting them in for the actual values of the after-tax wage and unearned income. Use of the latter procedure could yield inconsistent results since the deviations between predicted and actual values of the after-tax wage and unearned income would not be orthogonal to the other independent variables.

18. Note that the union variable may be endogenous and therefore a significant coefficient on this variable does not necessarily indicate that unions cause underemployment. This result may simply reflect a higher demand for union services by workers experiencing a high level of unemployment or underemployment.
19. Equation (14) was maximized using the GRADX algorithm of Goldfeld and Quandt (1972). The likelihood function was evaluated using a bivariate normal integration program written by J. Hausman. I first attempted to maximize (14) using numerical derivatives but could not achieve convergence. Convergence was obtained using analytical first derivatives and the Berndt et.al. (1974) approximation to the second derivate matrix. Amemiya (1974a) contains several identities which were helpful in calculating the derivatives.
20. In Table 1, coefficients which increase (decrease) unemployment or underemployment have a positive (negative) sign. The estimates in Table 1 are structural estimates. Reduced-form versions of the probit equations were estimated for use in implementing the selection rule procedure.
21. Nickell (1979) obtained the same result and contains an extensive discussion of why more children may lead to a higher probability of unemployment. He concluded that this result was something of a puzzle.
22. The wage may be significant in the underemployment equation but not in the unemployment equation because underemployed workers are ineligible for unemployment insurance while unemployed workers are eligible.
23. I would like to thank Mark Rosenzweig for making this point.
24. This, of course, assumes that a variable's effect on the duration of unemployment or underemployment is in the same direction as its effect on the incidence. This would be satisfied, for example, if unemployment and underemployment followed a Tobit structure.
25. Since the null hypothesis of no selection bias is being tested, the use of the uncorrected standard errors is appropriate.
26. In order to maintain internal consistency,  $\rho$  was also constrained to equal zero.
27. The labour supply equation is exactly identified using either set of instruments, and thus the two-stage estimator will be asymptotically efficient.
28. The duration estimates using the alternate instruments, which are equally imprecise, have been placed in Appendix Three.
29. It is interesting to note that the marital status coefficient does change sign in the unemployment duration equation, which is consistent with the selection rule estimator producing a larger coefficient than the full sample least squares estimator for this variable.
30. If the union variable is included as a separate independent variable, severe multicollinearity problems arise. Since the data set used here is too small to precisely estimate the duration equations, the issue of the most appropriate functional form is left for future research.
31. The expression  $Y \overset{D}{=} X$  means that  $Y$  has the same asymptotic distribution as  $X$ .

## References

1. Abowd, J. and Ashenfelter, O. (1979), "Unemployment and Compensating Wage Differentials", Industrial Relations Section, Princeton University, Working Paper No.120, March.
2. Ashenfelter, O. (1980), "Unemployment as Disequilibrium in a Model of Aggregate Labor Supply", Econometrica, 48, 547-564.
3. Amemiya, T. (1974a), "Bivariate Probit Analysis: Minimum Chi-Square Methods", Journal of The American Statistical Association, 69, 940-44.
4. \_\_\_\_\_ (1974b), "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal", Econometrica, 42, 999-1012.
5. Baily, M. (1974), "Wages and Employment Under Uncertain Demand", Review of Economic Studies, 41, 37-50.
6. Benassy, J.P. (1976), "The Disequilibrium Approach to Monopolistic Price Setting and General Monopolistic Equilibrium", Review of Economic Studies, 43, 69-81.
7. Berndt, E.B., Hall, B., Hall, R. and Hausman, J. (1974), "Estimation of Inference in Nonlinear Structural Models", Annals of Economic and Social Measurement, 3, 653-665.
8. Burtless, G. and Hausman, J. (1978), "The Effect of Taxation on Labor Supply: Evaluating the Gary Income Maintenance Experiment", Journal of Political Economy, 86, 1103-1130.
9. Catsiapas, G. and Robinson, C. (1978), "Sample Selection Bias with Two Selection Rules: An Application to Student Aid Grants", Department of Economics, University of Western Ontario, Working Paper No. 7833, Fall.
10. DaVanzo, J., De Tray, D. and Greenberg, D. (1976), "The Sensitivity of the Male Labor Supply Estimates to the Choice of Assumptions", Review of Economics and Statistics, 58, 313-325.
11. Dreze, J. (1975), "Existence of an Equilibrium Under Price Rigidity and Quantity Rationing", International Economic Review, 16, 301-320.
12. Goldfeld, S. and Quandt, R. (1972), Nonlinear Methods in Econometrics, Amsterdam: North Holland.
13. Hahn, F.H. (1978), "On Non-Walrasian Equilibria", Review of Economic Studies, 45, 1-17.
14. Ham, J. (1980), Three Empirical Essays on Constraints in the Labor Market, unpublished Ph.D. thesis, Princeton University, May.

15. Hausman, J. (1978), "Specification Tests in Econometrics", Econometrica, 46, 1251-1272.
16. Heckman, J. (1979), "Sample Selection Bias as a Specification Error", Econometrica, 47, 153-162.
17. Hurd, M. (1976), "The Estimation of Nonlinear Labor Supply Functions with Taxes from a Truncated Sample", Technical Report No. 217, Institute for Mathematical Studies on the Social Sciences, Stanford University, September.
18. Kalacheck, E., Mellow, W. and Raines, F. (1978), "The Male Labor Supply Function Reconsidered", Industrial and Labor Relations Review, 31, 365-367.
19. Kiefer, R. and Neumann, G. (1974), "An Empirical Job-Search Model, With a Test of the Constant Reservation-Wage Hypothesis", Journal of Political Economy, 87, 89-108.
20. Lee, L.F. (1979), "On the First and Second Moments of the Truncated Multi-Normal Distribution and a Simple Estimator", Economic Letters, 3, 165-169.
21. \_\_\_\_\_, Maddala, G.S. and Trost, R. (1980), "Asymptotic Covariance Matrices of Two-State Probit and Two-Stage Tobit Methods for Simultaneous Equation Models with Selectivity", Econometrica, 48, 491-503.
22. Malinvaud, E. (1977), The Theory of Unemployment Reconsidered, New York: Halsted.
23. Medoff, J.L. (1979), "Layoffs and Alternatives Under Trade Unions in U.S. Manufacturing", American Economic Review, 69, 380-395.
24. Mortenson, D. (1970), "Job Search, The Duration of Unemployment and the Phillips Curve", American Economic Review, 60, 847-862.
25. Nickell, S. (1979), "A Picture of Male Unemployment in Britain", Industrial Relations Section, Princeton University, Working Paper No. 123, June.
26. Olsen, R. (1977), An Econometric Model of Family Labor Supply, unpublished Ph.D. Thesis, University of Chicago, December.
27. Poirier, D. (1980), "Partial Observability in Bivariate Probit Models", Journal of Econometrics, 12, 209-218.
28. \_\_\_\_\_, and Ruud, P. (1980), "On the Appropriateness of Switching Regression Models with Endogenous Switching", unpublished paper, University of Toronto, February.
29. Rea, S. (1974), "Unemployment and the Supply of Labor", Journal of Human Resources, 9, 279-289.
30. Rosen, H. (1976), "Taxes in a Labor Supply Model with Joint Wage-Hours Determination", Econometrica, 44, 485-507.

31. Tallis, G. (1961), "The Moment Generating Function of the Truncated Multi-Normal Distribution", Journal of the Royal Statistical Society, 23, Series b, 223-229.
32. Wales, T. and Woodland, A. (1976), "Estimation of Household Utility Functions and Labor Supply Response", International Economic Review, 17, 397-410.
33. \_\_\_\_\_ (1977), "Estimation of the Allocation of Time for Work, Leisure and Housework", Econometrica, 45, 115-132.
34. \_\_\_\_\_ (1979), "Labour Supply and Progressive Taxes", Review of Economic Studies, 46, 83-95.