NBER WORKING PAPER SERIES

CENSORED REGRESSION MODELS WITH UNOBSERVED STOCHASTIC CENSORING THRESHOLDS

Forrest D. Nelson*

Working Paper No. 63

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
Cambridge, Massachusetts 02139

December 1974

Preliminary: Not for quotation

*NBER Computer Research Center. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

Abstract

The "Tobit" model is a useful tool for estimation of regression models with a truncated or limited dependent variable, but it requires a threshold which is either a known constant or an observable and independent variable. The model presented here extends the Tobit model to the censored case where the threshold is an unobserved and not necessarily independent random variable. Maximum likelihood procedures can be employed for joint estimation of both the primary regression equation and the parameters of the distribution of that random threshold. The appropriate likelihood function is derived, the conditions necessary for identification are revealed, and the particular estimation difficulties are discussed. The model is illustrated by an application to the determination of a housewife's value of time.

Contents

Tables

INTRODUCTION

Of concern in this paper are appropriate estimation techniques for relationships involving a "censored" dependent variable. That is we wish to estimate parameters of a regression model when data on the dependent variable are incomplete in the sense that the variable is observed only when it's value exceeds (or falls short of) some censoring threshold. The model may be written as

(1) $$Y_i = \beta' X_i + u_i \quad \text{if RHS} \geq T_i$$

(2) $$Y_i = \text{n.a.} \quad \text{if RHS} < T_i$$

The distinction between this model and the tobit or limited dependent variable model considered by Tobin [6] should be carefully noted. The tobit model is a truncated variable model with equation (2) replaced by

$$Y_i = T_i \quad \text{if RHS} < T_i$$

and requires that we know both which observations are truncated and the value of the threshold T_i for at least those truncated observations. In the censored model the actual value of the threshold will not generally be known for any observations.

As in the tobit model the threshold censoring results in a non-zero expectation of the disturbance term within the subset of non-censored observations so that least squares will yield biased parameter estimates. It would thus appear that maximum likelihood estimation is more appropriate.

Derivation of the likelihood function requires a specification of the behavior of the unobserved threshold.

Part I of this paper treats the estimation problem when the threshold is assumed to be the unobserved endogenous variable of a second regression relationship. The likelihood function is derived and the model is compared with simple probit and tobit models to highlight certain features and difficulties such as conditions necessary for identification of parameters. The difficulties of obtaining estimates for the model are discussed and the results of some limited similation experiments are presented for some indication of the performance of the estimators.

Part II illustrates the model with an application to the determination of the value of a housewife's time. Following Gronau [3] and Heckman [4] the housewife's market wage is the censored dependent variable and the value of her time at home is the threshold variable. It is argued that the censored model discussed here is the appropriate one to use for estimation under the assumptions invoked by Gronau rather than the probit analysis model he employed. The relationship to Heckman's model, in which the two equations are simultaneous, is also discussed.

I. The Censored Dependent Variable Model

The model to be considered here is

(3) $$y_{1t} = \beta_1' X_{1t} + u_{1t}$$

(4) $$y_{2t} = \beta_2' X_{2t} + u_{2t}$$

(5) $$\begin{aligned} Y_t &= y_{1t} \text{ if } y_{1t} \geq y_{2t} \\ &= 0 \quad \text{if } y_{2t} > y_{1t} \end{aligned}$$

Y_t is the censored dependent variable which, for convenience only, is assigned the value zero from censored observations. y_{1t} and y_{2t} are latent (i.e., not directly observable) endogenous variables and X_{1t} and X_{2t} are perhaps overlapping vectors of observable exogenous variables which may include the constant unity. u_{1t} and u_{2t} are random disturbances assumed here to follow a bivariate normal distribution with a zero mean vector and unknown variances and covariance , σ_1^2, σ_2^2 and σ_{12}. Both disturbances are assumed to be independent across observations and independent of X_{1t} and X_{2t}. From a sample of T observations on Y_t, X_{1t} and X_{2t} we require estimates of the vectors β_1 and β_2 and the scalars σ_1^2, σ_2^2 and σ_{12}.

For notational convenience let Ψ_1 and Ψ_2 denote the subsets of censored and non-censored observations respectively. That is, if Ψ is the set of integers {1, ..., T} then Ψ_1 is the subset of Ψ corresponding to $y_{1t} < y_{2t}$ and Ψ_2 is the subset corresponding to $y_{1t} \geq y_{2t}$. Determination of the subsets Ψ_1 and Ψ_2 should be obvious from an inspection of the data. The subscript t will be deleted in what follows for ease of notation.

Clearly ordinary least squares is not the appropriate estimation procedure for even β_1 and σ_1^2 over the subsample Ψ_2. The method of censoring implies that observations with an algebraically small value for u_1 are more likely to be censored than observations with relatively large values for u_1. Thus the expected value of u_1 over the subsample Ψ_2 is not zero and OLS will yield biased estimates. Moreover, the censoring induces a correlation between u_1 and X_1 within the non-censored subsample.

Maximum likelihood appears to be a more reasonable estimation technique for this model. To formulate the likelihood function the distribution of Y must be derived from the distribution of u_1 and u_2. Y takes on the value 0 when $y_1 < y_2$, or when

$$u_1 - u_2 < \beta_2' X_2 - \beta_1' X_1.$$

Defining $V = u_1 - u_2$, it is obvious that V follows a univariate normal distribution with mean zero and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$. The probability that Y equals zero is thus given by

$$\text{(6)} \qquad \Pr(Y=0) = \Pr(v < \beta_2' X_2 - \beta_1' X_1) = P\left(\frac{\beta_2' X_2 - \beta_1' X_1}{\sigma}\right)$$

where P(A) represents the unit normal distribution function, $P(A) = \int_{-\infty}^{A} \frac{1}{\sqrt{2\pi}} \exp(-a^2/2)\, da$. The expression in (6) is the appropriate measure of probability for Y for observations in the set Ψ_1. For observations in the set Ψ_2 we know that $y_1 = Y$ while $y_2 < Y$. Letting $f(y_1 - \beta_1' X_1, y_2 - \beta_2' X_2)$ be the bivariate normal density function for u_1 and u_2 we obtain

(7) $$\int_{-\infty}^{Y-\beta_2' X_2} f(Y-\beta_1' X_1, u_2)\, du_2$$

as the appropriate probability measure for Y for observations in Ψ_2. Using (6) and (7) the likelihood function may be written as

(8) $$L(\beta_1, \beta_2, \sigma_1, \sigma_2, \sigma_{12} \mid Y, X_1, X_2) =$$

$$\prod^{\Psi_1} P\left(\frac{\beta_2' X_2-\beta_1' X_1}{\sigma}\right) \cdot \prod^{\Psi_2} \int_{-\infty}^{Y-\beta_2' X_2} f(Y-\beta_1' X_1, u_2)\, du_2$$

If we assume $\sigma_{12}=0$ this likelihood simplifies to

(9) $$L(\beta_1, \beta_2, \sigma_1, \sigma_2 \mid Y, X_1, X_2) =$$

$$\prod^{\Psi_1} P\left(\frac{\beta_2' X_2-\beta_1' X_1}{(\sigma_1^2+\sigma_2^2)^{1/2}}\right) \cdot \prod^{\Psi_2} \frac{1}{\sigma_1} Z\left(\frac{Y-\beta_1' X_1}{\sigma_1}\right) \cdot P\left(\frac{Y-\beta_2' X_2}{\sigma_2}\right)$$

where Z represents the unit normal density function.

Like the likelihood function for the tobit model, (8) and (9) include both density and distribution functions and yield nonlinear normal equations so that iterative maximization procedures are required for obtaining estimates. As will be shown below implementation of such procedures for the censored model is more difficult than for the tobit and probit models. Several other aspects of the model will also be considered including the marginality of the information in a sample with respect to identification of the parameters, the inseparability of the model which necessitates simultaneous estimation of both equations, and methods of obtaining initial estimates to start the iterative maximization procedure.

It is useful to first consider a decomposition of the model into the related tobit and probit models. As was suggested above, the tobit model requires observations on the threshold variable. Suppose that y_2 was observable. Then the likelihood function would be written as

$$(10) \qquad L = \prod^{\Psi_1} \int_{-\infty}^{y_2} f(y_1-\beta_1' X_1, y_2-\beta_2' X_2)\, dy_1 \cdot \prod^{\Psi_2} f(Y-\beta_1' X_1, y_2-\beta_2' X_2)$$

If in addition u_1 and u_2 were independent the likelihood would factor to

$$(11) \qquad L = \prod^{\Psi_1} \int_{-\infty}^{y_2} f(y_1-\beta_1' X_1)\, dy_1 \cdot \prod^{\Psi_2} f(Y-\beta_1' X_1) \cdot \prod^{\Psi} f(y_2-\beta_2' X_2)$$

allowing estimation of equation (3) by tobit analysis and equation (4) by OLS separately.* Clearly the lack of observations on y_2 in the censored model prevents estimation by tobit analysis. One might proceed instead to obtain consistent estimates of y_2 and then apply the tobit model as above using these estimates but, as will be seen, such estimates may be impossible to obtain and even then the quality of the resulting parameter estimates might diminish considerably.

It is possible to estimate the censored model directly by discarding the observations on Y, the only endogenous information retained being the separation of the sample into the two subsets Ψ_1 and Ψ_2. That is the endogenous variable retained is an indicator variable, say I, defined by

---

* Even if $\sigma_{12} \neq 0$ we might proceed to estimate the two equations separately arguing, by analogy to the "seemingly unrelated regressions" problem, that this sacrifices only efficiency. It is not clear, however, that the analogy holds. Separate estimation might lead in this case to inconsistent estimates.

(12) $$I_t = 1 \text{ if } t \in \Psi_2 \quad (y_1 \geq y_2)$$
$$= 0 \text{ if } t \in \Psi_1 \quad (y_1 < y_2)$$

The resulting likelihood function, conditional now on X_1, X_2 and I, is

(13) $$L = \overset{\Psi_1}{\Pi} P\left(\frac{\beta_2' X_2 - \beta_1' X_1}{\sigma}\right) \cdot \overset{\Psi_2}{\Pi}\left[1 - P\left(\frac{\beta_2' X_2 - \beta_1' X_1}{\sigma}\right)\right].$$

where, as before, $\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2 - 2\sigma_{12}$. The difficulty here is in the identification of the parameters. $(\beta_2' X_2 - \beta_1' X_1)/\sigma$ is observationally equivalent to $(\kappa\beta_2' X_2 - \kappa\beta_1' X_1)/\kappa\sigma$, where κ is any scalar other than zero. Thus we cannot identify σ, let alone it's separate components σ_1, σ_2 and σ_{12}, and can estimate the slope coefficients only up to a scalar multiple, (β_{ij}/σ). Furthermore if X_1 and X_2 overlap with common variables, for example if both equations include an intercept term, the corresponding coefficients would also not be separately estimable - we could only estimate their difference up to the scalar multiple $(\frac{\beta_{2j} - \beta_{1k}}{\sigma})$. Obviously the endogenous variable I by itself does not provide sufficient information to identify all parameters of the model.

Consider next the situation whey y_1 is observable for all observations instead of just those in the set Ψ_2. The likelihood function relevant here is

(14) $$L = \overset{\Psi_1}{\Pi} \int_{y_1}^{\infty} f(y_1 - \beta_1' X_1, y_2 - \beta_2' X_2)\, dy_2 \cdot \overset{\Psi_2}{\Pi} \int_{-\infty}^{Y} f(y_1 - \beta_1' X_1, y_2 - \beta_2' X_2)\, dy_2$$

which, when $\sigma_{12} = 0$, factors to yield the probit likelihood function for equation (4),

(15) $$L(\beta_2, \sigma_2 | I, y_1, X_2) = \overset{\Psi_1}{\Pi} P\left(\frac{\beta_2' X_2 - y_1}{\sigma_2}\right) \cdot \overset{\Psi_2}{\Pi}\left[1 - P\left(\frac{\beta_2' X_2 - y_1}{\sigma_2}\right)\right].$$

Knowledge of both I and y_1 for all observations plus the assumption of zero covariance are sufficient for the identification of all parameters in (14). Contrasting equations (15) and (13), it is the natural normalization of the coefficient of (-1) for y_1 in equation (15) which allows the identification. It can be shown, however, that when the covariance is also to be estimated, as in equation (14), identification is not guaranteed.

To see the identification problem consider the model given by equations (3) and (4) written now in matrix form

$$(16)\qquad \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \beta_1' \\ \beta_2' \end{pmatrix} Z + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

where the subscript t has been deleted and Z is a k element vector including all variables in X_1 and X_2. Variables excluded from an equation are now represented by zero restrictions on elements of β. We can multiply the system of equations (16) by any arbitrary 2×2 nonsingular matrix A and obtain an observationally equivalent system. Consider the following choice for A.

$$(17)\qquad A = \begin{bmatrix} 1 & 0 \\ \dfrac{-\sigma_{12}}{\sigma_1^2} & 1 \end{bmatrix}$$

On multiplication of (16) by A, the first equation is unchanged while the second becomes

$$(18)\qquad Y_2 = \frac{\sigma_{12}}{\sigma_1^2} Y_1 + \left(\beta_2' - \frac{\sigma_{12}}{\sigma_1} \beta_1'\right) Z + \left(u_2 - \frac{\sigma_{12}}{\sigma_1^2} u_1\right)$$

$$= \theta_1 Y_1 + \theta_2' Z + V, \text{ say.}$$

Note that in (18) Y_1 is independent of V and that $Var(v) = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}$.

(the transformed model is recursive.) We could, therefore, estimate the two equations of the transformed model separately. Reimposing the probit structure on the model we note that Y_1 is always observed while Y_2 is never observed - we know only for which observations Y_2 exceeds Y_1. Thus, we would estimate (18) using probit analysis by deriving

$$(19) \qquad Pr(Y_2 < Y_1) = Pr(Y_1(1-\Theta_1) - \Theta_2' Z > v) = P\left(\frac{Y(1-\Theta_1) - \Theta_2' Z}{\sigma_v}\right).$$

But as in the usual probit model we have no information on the scale of Y_2 and cannot therefore directly estimate σ_v. We estimate instead

$$\frac{1-\Theta_1}{\sigma_v} = \frac{1-{}^{\sigma_{12}}/\sigma_1{}^2}{(\sigma_2^2-\sigma_{12}^2/\sigma_1^2)^{1/2}}$$

and

$$\frac{\Theta_2}{\sigma_v} = \frac{1}{(\sigma_2^2-\sigma_{12}^2/\sigma_1^2)^{1/2}} \cdot \left(\beta_2 - \frac{\sigma_{12}}{\sigma_1} \beta_1\right)$$

Clearly not all parameters are identified without further restrictions. That is, since the transformed system is observationally equivalent to the original system we cannot identify the parameters β_2, σ_2 and σ_{12} in that original system. To achieve identification we need at least one linear restriction among this set of parameters, such as a zero restriction on σ_{12} or one element of β_2.

The situation is nearly analogous to simultaneous equation models. The original system in our model "looks like" a reduced form while the transformed system "looks like" a structural form and the identifiablility conditions "look like" the same. The not so subtle difference is that in this probit structure the approach to estimation and identification are backwards. In SE we could estimate the reduced form directly since each equation involves only one endogenous variable. But in our probit formulation the second equation uses Y_1 from the first as its threshold, preventing its direct estimation unless Y_1 happens to be independent of u_2. Thus we must go the other direction and generate a "structural model" with a recursive form to use for estimation.

Looking at our transformed system as if it were a structural form we can count the number of restrictions among the endogenous variable coefficients (our matrix A), noting one restriction for equation one (the 0 in the top right corner) and one for the second (the element in the lower right hand corner which is a linear function of variance terms). Thus we would say that the model is identified. However since the second equation must be estimated with probit analysis rather than OLS we sacrifice one degree of identification and must therefore have one more restriction in equation two. So the identifiability conditions seem to be the same. The difference here is that in simultaneous equations we ask whether the restrictions on the structural coefficients impose sufficient restrictions on the reduced form to permit identification. In this probit model we ask the reverse - do the restrictions on the "reduced form" coefficients impose sufficient conditions on the "structural" coefficients to permit identification.

As in the usual simultaneous equations estimation, too many restrictions result in over identification. In a just identified model we could estimate the probit equation only, provided the condition arises from a zero covariance restriction. Otherwise we need estimates for both equations since σ_1 and β_1 from the first are used in identifying the second. In an over identified model we have the problem of multiple solutions when estimating the equations seperately which is easily solved by the obvious 2SLS analog or FIML estimation of the entire model.

We can now restore equation (5) and re-examine the properties of the censored regression model in light of its probit and tobit analogs. The model is like a tobit model except that it does not admit observations on y_2. It is like a probit model except that y_1 is observed for only some of the observations. We could thus regard it as a hybrid which, unforturatly, exhibits all the unattractive features of its parent strains. Specifically the identifiability conditions are the same as for the last probit model discussed above. Identification, even when the conditions are met, is however in some sense only marginal. The identifiability argument with respect to the subset of non limit observations is identical to that presented above for the last probit model while the under identified result of the first probit model applies to the subset of limit observations. Thus the entire burden of identifiability falls on just the subset of non limit observations.

A second unattractive feature of the censored model from the standpoint of computational difficulty lies in the inseparability, with respect to estimation, of the two equations. This feature is shared with the first probit model examined above and arises because the probability measure for limit observations (see equation (6)) involves all parameters of both equations in an inseparable form.

Consider again the iterative maximization of likelihood functions (7) or (8). Experience with the probit and tobit models suggests that the Newton-Raphson iterative maximization algorithm performs quite well on functions of this sort with rapid convergence rates even when starting from poor initial values. But the author's use of this algorithm on artificial data for the censored model gave mixed and discouraging results. Two factors in particular had to be accounted for. First the log likelihood is not concave over a wide range of the parameter space so that the matrix of second derivatives may not be negative definite, as is required for convergence of the Newton algorithm, at any arbitrary set of initial values for the coefficients. A modification to that Hession matrix such as the one proposed by Greenstadt [2] thus proved necessary. Second, a pattern often observed in the iterative maximization was that the coefficients appeared to be moving in the right direction but the steps taken were so large that eventually the maximum was oversteped with the variance terms driven out of the parameter space, resulting in a failure of the procedure. An algorithm which proved a bit more stable was a "Dogleg" algorithm developed by Rick Becker [1]. That algorithm was derived along the lines of Powell's [5] MINFA routine but uses analytic first and second derivatives. It uses a combination of Newton and steepest ascent iterations, explicitly controlling the length of steps taken.

Obtaining starting values for the iterative maximization procedure proved to be a troublesome task. The procedure adopted for the work presented here was: (a) apply OLS to equation (3) over the subset of observations Ψ_2; (b) obtain $\hat{y}_1$ for the subset Ψ_1 using the OLS estimates; and (c) apply the probit model with observed threshold ($\hat{y}_1$ in the set Ψ_1 and Y in the set Ψ_2) to equation (4). For purposes of obtaining initial estimates σ_{12} was assumed to be zero so that the more simple likelihood function (15) could be applied in step (c).

To test the feasibility of and provide (admitedly weakly) evidence for the performance of maximum likelihood estimation on the censored model some limited simulation experiments were conducted. The model used was

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_1 \\ Y_2 &= \delta_0 + \delta_1 X_3 + \delta_2 X_4 + u_2 \\ Y &= Y_1 \quad \text{if } Y_1 \geq Y_2 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Independent variables were drawn from independent normal distributions with zero mean and unit variance and were held fixed in repeated samples. Parameter values were chosen so that the true coefficient of determination in both regression equations was around.6. Sample size used was 100.

Results of the experiment are reported in table 1 below. Estimates of the parameters of equation (3) are notably better than those for equation (4) as would be expected. Note that the model above is identified by the absence of X_3 and X_4 in the first equation. Simulations on models with differing degrees of identification give similar results with some indication that estimates of equation two and the covariance improve as degrees of identification increase.

Table I

Simulation Results

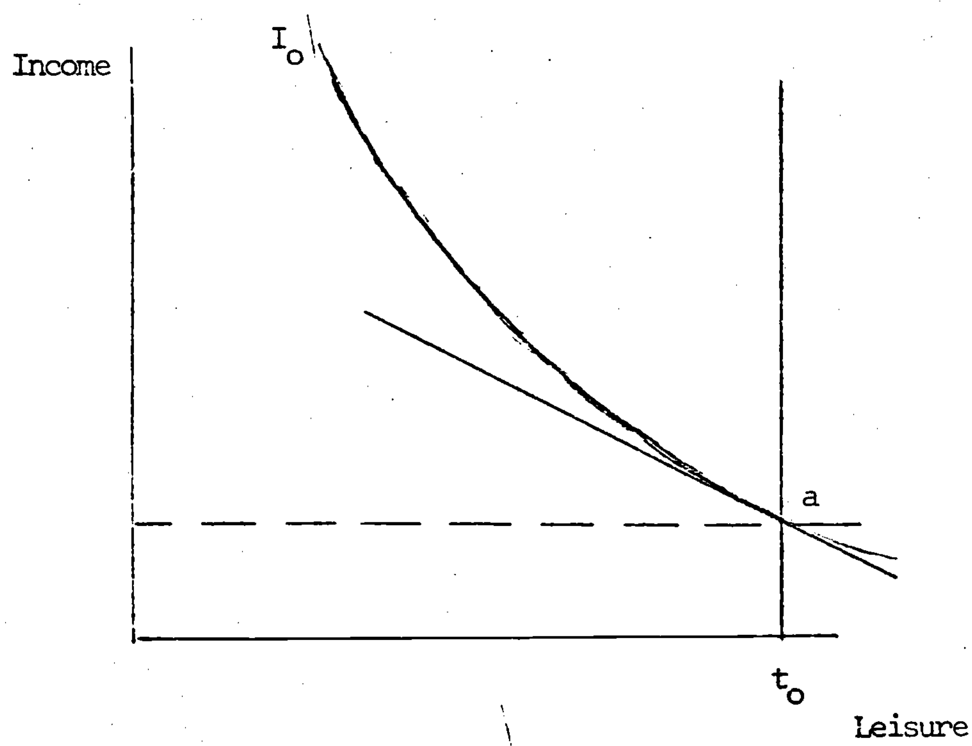
(Summary results for 10 samples)

parameter	true value	mean estimate	minimum estimate	maximum estimate
β_o	0.	-.0674	-.3358	.3417
β_1	-1.	-.9988	-1.2551	-.7079
β_2	1.	.9844	.8163	1.1949
δ_o	0.	-.1111	-.4919	.3337
δ_1	-1.	-.9860	-1.3306	-.7853
δ_2	1.	.9859	.6158	1.4117
σ_1^2	1.	.9914	.7131	1.3362
σ_2^2	1.	.7783	.2917	1.3159
σ_{12}	.64	.5405	.3189	.7963

parameter	mean bias	st. dev.	root mean sq. error	t ratio
β_o	.0675	.1911	.2026	1.117
β_1	-.0012	.1618	.1618	-.023
β_2	.0156	.0981	.0993	.503
δ_o	.1111	.2373	.2620	1.481
δ_1	-.0140	.1654	.1660	-.268
δ_2	.0141	.2156	.2161	.207
σ_1^2	.0086	.1839	.1841	.147
σ_2^2	.2217	.3639	.4261	1.926
σ_{12}	.0995	.1595	.1880	1.972

II. An Application to the Estimation of Value of Time

Estimation of labor supply relationships at the micro level is often frustrated by the abscence of potential wage data for non-participants in the labor force. If the decision to work was made independently of potential wage rates, wage determination relationships could be estimated directly from samples drawn from the labor force. it is more reasonable to assume however that such decisions are directly affected by wage offers. Other things equal the higher the offered or potential wage the more likely a potential worker will accept the offer and enter the labor force. Thus such samples would tend to overestimate potential wages for nonworkers. Such a mechanism is captured in the familiar diagram illustrating indifference curves in the income-leisure plane.



Gronau was concerned with estimating the value of a housewife's time and, more specifically, on the effect of children on the value of her time. The model he used can be formulated as

(20) $W^p = f(E)$

(21) $V = g(C)$

(22) $W^m = W^p \quad \text{if } W^p > V$

$= 0 \quad \text{otherwise}$

where W^p is a housewife's potential wage which depends on her markatable characteristics (E) such as training and work experience, V is the value of her time at home with zero hours of work which is a function of such characteristics as family income and number of children, and W^m is the wage she recieves if she does in fact enter the labor force. The reader is refered to Gronau's paper for a derivation of the relationship from household utility maximization and a discussion of assumptions underlying the model and the possible bias they introduce when violated. One particularly troublesome assumption which was neglected in his paper is flexibility in hours worked for working women. Since the same problem arises in Heckman's analysis a discussion of it will be delayed until later.

Gronau applied probit analysis to obtain estimates for equation (21). As he discussed and as explained in section I of this paper, neglecting any observed wage rates and analyzing the labor force participation decision with straight forward application of probit methods provides estimates of coefficients only up to a scale factor and even then does not permit separate estimates of coefficients for variables common to both equations. On the other hand if potential wages were known for all women this variable, he argued, could be included as a variable in the probit model, its coefficient providing an estimate of the variance and thereby permitting identification of the coefficients in equation (21). Since potential wages are not always observed he devoted considerable attention to obtaining proxy measures for it. His efforts in this direction were admirable and promising but their success hinges crucially on the assumption of zero correlation between W^p and the disturbance in the value of time equation. Other authors, Heckman [4] for example, have provided evidence that the assumption does not hold. If the threshold in a probit model is not independent of the disturbance, consistent estimates will not be obtained. The censored variable estimation procedure directly overcomes the problem of missing potential wage data. Furthermore it relies on the zero correlation assumption only as one means of achieving identification. (Unfortunately the data source used by Gronau and his specification of the model invokes this reliance as will be explained below.)

To illustrate the method we returned to the data source used by Gronau, the 1960 census 1/1000 sample and collected a random sample of 750 observations for urban white married women, spouse present, who belonged to primary families in households with no nonrelatives. The variables obtained were:

W^m = hourly wage rate (in dollars) (1959 earnings/(1959 weeks worked X hours worked last week))

$E_1 = C_1$ = Dummy variable (0,1) for age less than 30

$E_2 = C_2$ = Dummy variable (0,1) for age greater than 49

$E_3 = C_3$ = dummy variable (0,1) for education less than high school

$E_4 = C_4$ = dummy variable (0,1) for education greater than HS

C_5 = family income (in \$10,000) net of wife's earnings

C_6 = husbands age (in years)

C_7 = dummy variable (0,1) for husbands education less than HS

C_8 = dummy variable (0,1) for husbands education greater than HS

C_9 = number of children less than 3 years of age

C_{10} = number of children 3 to 5 years of age

C_{11} = number of children 6 to 12 years of age

C_{12} = number of children greater than 12 years of age

It is important to note that for this specification, as indicated by the variable list above, of factors determining the potential wage and the value of time, the parameters of equation (21) are identified only if there is zero covariance between the disturbances in the two equations. This is unfortunate since, as already noted, the validity of the zero covaraince assumption is doubtful. However since the primary purpose here is illustration we proceeded under this assumption in order to compare as closely as possible the results of the censored and probit approaches to Gronau's model. The identification problem arises here because of the limitations imposed by the data source. Potential wages ought to depend on education, special training and work experience. Since only the first of these is available from the 1960 census, age was used as a proxy for experience and this variable also appears as a factor in value of time. Had a proper measure of experience been available for use in equation (20), exclusion of it in (21) would have been sufficient for identification without the zero covariance assumption.

The choice of variables follows Gronau and the reader is refered to his paper for a justification for that choice. We deviate from his choice only in that he included other measures for the effect of children to account for possible nonlinearities or returns to scale. Gronau experimented with both additive and multiplicative functional forms for the two equations and ultimately adopted the later for more appealing theoretical rational and greater explanatory power. Our experience was the same. Thus the functional form used for the results appearing below was $Y=b_0 b_1^{X1} b_2^{X2} \ldots b_k^{Xk} u$ for both equations where the disturbance u was assumed to follow a log normal distribution. (Estimates presented are for parameters of the form $\ln(b_i)$.)

The model was estimated using both the censored and probit procedures. The details of the later require more detailed explanation. One of the procedures used by Gronau was to estimate, via probit analysis, the model

$$L = 1 \quad \text{if} \quad b'C + u > \ln(\overline{W}^P)$$

$$= 0 \quad \text{if} \quad b'C + u \leq \ln(\overline{W}^P)$$

where L is the labor force participation indicator and $\overline{W}^P$ was taken to be the geometric average of wages recieved by working women with characteristics C_1-C_4. This was the procedure adopted for use here. Results for the two methods are presented in table II below. As can be seen the differences in the coefficient estimates are not striking but there is a sizeable difference in the estimate of the mean value of a housewife's time.

Table II

Estimates of the Value of a Housewife's Time

Variable	censored model coefficient	censored model t ratio	probit model coefficient	probit model t ratio
constant	-.4057	-1.443	-.1803	-.211
C_1	.1518	.982	.1083	.582
C_2	.1815	1.275	.1373	1.395
C_3	-.0235	-.204	-.0175	-.068
C_4	.2166	1.731	.2916	.457
C_5	.6817	5.939	.3635	5.685
C_6	.1141	1.878	.1006	2.964
C_7	-.0276	-.282	.0098	.1615
C_8	.0616	.596	.0215	.335
C_9	.3681	3.397	.2614	4.554
C_{10}	.2004	2.690	.1088	2.321
C_{11}	.1479	2.330	.1417	4.011
C_{12}	-.0903	-1.488	-.0123	-.327
st.error	.4278		.4243	
mean value of time	$2.61		$2.27	
constant	.2689	2.084		
E_1	-.0772	-.704		
E_2	-.0656	-.551		
E_3	-.2400	-2.119		
E_4	.2796	2.247		
st. error	.7287			
mean potential wage	$1.26			

As noted earlier Heckman [4] looked at the same basic problem but used a different estimation procedure. His model formulation is

(23) $W^p = f(E)$;

(24) $V = g(H,C)$

where H represents hours worked and other variables are as previously defined. If hours worked are perfectly flexible then working women will adjust H so as to equate W^p and V. When a corner solution is reached (H = 0) W^p exceeds V, both are unobserved and the individual drops out of the labor force. The interpretation placed on V by the two authors is somewhat different. In Heckman's formulation V is the shadow price of time or the slope of a tangent to the indifference curve, which of course varies as hours of work change. Gronau on the other hand specifically chose V to prepresent the value of time for a nonparticipant, or alternatively the asking wage, and this value of time will be equal to the slope of an indifference curve only at zero working hours.

A crucial assumption in both models is flexibility in hours worked. It might be argued however that Heckman's analysis relies more heavily on that assumption. Any rigidity here would mean that only by chance would the shadow price of time equal the market wage at any institutionally fixed hours of work. In Gronau's analysis on the other hand the only observations violating the conditions of his model are those for which the potential wage exceeds the value of time but, at the rigid hours, places the individual on a lower indifference curve than would nonparticipation. In both cases rigid hours lead to a bias in the estimates obtained but the conjecture is that the bias would be greater using Heckman's approach. Verification of this conjecture and, more important, a method for estimation accounting for such rigidity await further research. In fairness it should be noted that Heckman's procedure is more powerful in terms of the uses to which it may be put since it does permit estimation of indifference curves which the censored model does not.

To estimate his model Heckman used maximum likelihood, deriving, as in the censored model, $Pr(g(0,C)>f(E))$ for nonworking women and for working women using the pdf representing the joint distribution of H and $W^P(=V)$.

References

[1] Becker, R., "Dogleg Minimization Algorithm", NBER-CRC memorandum, April, 1974.

[2] Greenstadt, J., "On the Relative Efficiencies of Gradient Methods", Math. of Comp., July, 1967.

[3] Gronau, R., "The Effect of Children on the Housewife's Value of Time", Journal of Political Economy, March, 1973.

[4] Heckman, J., "Shadow Prices, Market Wages, and Labor Supply", July, 1973, (forthcoming in Econometrica).

[5] Powell, M.J.D., "A New Algorithm for Unconstrained Optimization", Nonlinear Programming, (ed. by Rosen et al.) Academic Press, 1970.

[6] Tobin, J., "Estimation of Relationships for Limited Dependent Variables", Econometrica, 1958.