Journal of Econometrics 6 (1977) 309–327. 

# CENSORED REGRESSION MODELS WITH UNOBSERVED, STOCHASTIC CENSORING THRESHOLDS*

Forrest D. NELSON

*California Institute of Technology, Pasadena, CA 91125, USA*

Received May 1975, final version received April 1977

The 'Tobit' model is a useful tool for estimation of regression models with truncated or limited dependent variables, 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.

## 1. Introduction

Of concern in this paper is a method of estimation for a regression model with a 'censored' dependent variable. That model may be written as

$$Y_i = \beta' X_i + u_i, \quad \text{if} \quad \text{RHS} \geqq t_i, \tag{1}$$

$$Y_i = \text{n.a.}, \qquad \text{otherwise}, \tag{2}$$

where $X_i$ is an always observable vector of independent variables, $\beta$ is a vector of unknown parameters, $u_i$ is the disturbance, $t_i$ is an *unobserved* censoring threshold and $Y_i$ is the censored dependent variable.

The similarity with the Tobit or limited dependent variable model introduced by Tobin (1958) is obvious. The distinction is that Tobin treated the threshold $t_i$ as either a known constant or an observable variable whereas here we take it to be unobserved. To motivate this modification consider the durable goods expenditure problem for which Tobin suggested the Tobit model and assigned the constant 0 to $t_i$, thereby ruling out the possibility of negative expenditures. With such a model we would expect a smooth distribution of expenditures from

*The author is indebted to Professors G.S. Maddala and Sherwin Rosen for their helpful comments. The research reported here was supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research.

0 to large amounts, but observed behavior contradicts this hypothesis. Households spend either nothing or sizable amounts and the Tobit model is incapable of explaining this bias away from small purchases. If there is some threshold below which purchases are not made it must not be zero. This then leads us to the model described by eq. (1) in which we take the threshold to be unobserved.

Attempts at estimation of the censored model would be futile without some assumed structure on $t_i$. In section 2 we complete the specification of the model by treating $t_i$ as the dependent variable of a second regression equation. Here too the likelihood function is derived, indentifiability of parameters is examined and some problems associated with maximum likelihood estimation are discussed.

In section 3 the similarity between the censored variable model and several multivariate probit and Tobit models is explored more fully. These models are of some interest in their own right but in addition they provide a means of assessing the 'value of information'. The results of some limited Monte Carlo experiments are reported.

Finally, in section 4 a number of interesting economic relationships to which the censored model might apply are noted. One of these, the value of a housewife's time, is considered in detail and used to demonstrate the feasibility of the model with a worked out example.

## 2. The censored dependent variable model

The model to be considered here is

$$y_{1t} = \beta_1' X_t + u_{1t}, \tag{3}$$

$$y_{2t} = \beta_2' X_t + u_{2t}, \tag{4}$$

and

$$\begin{aligned} Y_t &= y_{1t}, \quad \text{if} \quad y_{1t} \geqq y_{2t}, \\ &= 0, \qquad \text{if} \quad y_{2t} > y_{1t}. \end{aligned} \tag{5}$$

$Y_t$ is the censored dependent variable which, for convenience only, is assigned the value zero for censored observations. $y_{1t}$ and $y_{2t}$ are latent (i.e., not directly observable) endogenous variables and $X_t$ is a $K$-element vector of observable exogenous variables which may include the constant one. $u_{1t}$ and $u_{2t}$ are random disturbances assumed here to follow a bivariate normal distribution with a zero mean vector and unknown variances and covariance, $\sigma_1^2$, $\sigma_2^2$, and $\sigma_{12}$. Both disturbances are assumed to be independent across observations and independent of $X_t$. There may be zero restrictions on elements of $\beta_1$ and $\beta_2$.

Note that the values of $Y$ below $y_{2t}$, not zero, are censored so that if $y_{2t}$ can take on negative values then clearly $Y_t$ can also be negative. Non-negativity constraints, if desirable in a particular application, are readily imposed by constraining $y_{2t}$ to be positive, for example by expressing $y_{1t}$, $y_{2t}$ and $Y_t$ in log form.

From a sample of $T$ observations on $Y_t$ and $X_t$ we require estimates of the vectors $\beta_1$ and $\beta_2$ and the scalars $\sigma_1^2$, $\sigma_2^2$ and $\sigma_{12}$. For notational convenience let $\psi_1$ and $\psi_2$ denote the subsets of censored and non-censored observations respectively. That is, if $\psi$ is the set of integers $\{1,\ldots,T\}$ then $\psi_1$ is the subset of $\psi$ corresponding to $y_{1t} < y_{2t}$, and $\psi_2$ is the subset corresponding to $y_{1t} \geqq y_{2t}$. Determination of the subsets $\psi_1$ and $\psi_2$ should be obvious from an inspection of the data. The subscript $t$ will generally be deleted in what follows for ease of notation.

Clearly ordinary least squares is not the appropriate estimation procedure for even $\beta_1$ and $\sigma_1^2$ over the subsample $\psi_2$. The method of censoring implies that observations with algebraically small values for $u_1$ are more likely to be censored than observations with large values. Thus the expected value of $u_1$ over the subsample $\psi_2$ is not zero and OLS will yield biased estimates. Moreover, the censoring induces a correlation between $u_1$ and $X$ within the subset of non-censored observations.

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

$$u_1 - u_2 < \beta_2' X - \beta_1' X.$$

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

$$\Pr(Y = 0) = \Pr(v < \beta_2' X - \beta_1' X) = P\left(\frac{\beta_2' X - \beta_1' X}{\sigma}\right), \tag{6}$$

where $P(A)$ represents the unit normal distribution function, $P(A) = \int_{-\infty}^{A} (2\pi)^{-\frac{1}{2}} e^{-a^2/2}\, da$. The expression in (6) is the appropriate measure of probability for $Y$ for observations in the set $\psi_1$. For observations in the set $\psi_2$ we know that $y_1 = Y$ while $y_2 < Y$. Letting $f(y_1 - \beta_1' X, y_2 - \beta_2' X)$ be the bivariate normal density function for $u_1$ and $u_2$, the appropriate measure of probability for $Y$ for observations in the set $\psi_2$ is

D

$$g(Y) = \int_{-\infty}^{Y-\beta_2'X} f(Y-\beta_1'X, u_2)\mathrm{d}u_2 \tag{7}$$

$$= \frac{1}{\sigma_1} Z\left(\frac{Y-\beta_1'X}{\sigma_1}\right) P\left[\frac{(1-\sigma_{12}/\sigma_1^2)Y-\beta_2'X+(\sigma_{12}/\sigma_1^2)\beta_1'X}{(\sigma_2^2-\sigma_{12}^2/\sigma_1^2)^{\frac{1}{2}}}\right],$$

where $Z(A)$ is the unit normal density evaluated at $A$ and the expression on the right is obtained by writing the joint density $f(u_1, u_2)$ as the product of the marginal density of $u_1$ and the conditional density of $u_2$ given $u_1$. Using (6) and (7) the likelihood function may be written as

$$L(\beta_1,\beta_2,\sigma_1,\sigma_2,\sigma_{12}| Y,X) = \prod^{t\epsilon\psi_1} \Pr(Y_t = 0)\cdot \prod^{t\epsilon\psi_2} g(Y_t). \tag{8}$$

At this point the question of indentifiability of all the parameters in (8) arises. Indeed inspection reveals certain restrictions on the parameters as a necessary condition for indentification. To see this define the scalar $\delta$ and the vector $\theta$ as

$$\delta = (\sigma_2^2-\sigma_{12}^2/\sigma_1^2)^{1/2}/(1-\sigma_{12}/\sigma_1^2), \tag{9}$$

and

$$\theta = [1/(1-\sigma_{12}/\sigma_1^2)]\,[\beta_2-(\sigma_{12}/\sigma_1^2)\beta_1]. \tag{10}$$

Now using (9) and (10) the likelihood given in (8) may be rewritten in the observationally equivalent form

$$L = \prod^{t\epsilon\psi_1} P\left[\frac{(\theta'-\beta_1)X_t}{(\sigma_1^2+\delta^2)^{1/2}}\right] \prod^{t\epsilon\psi_2} \frac{1}{\sigma_1} Z\left(\frac{Y_t-\beta_1'X_t}{\sigma_1}\right)\cdot P\left(\frac{Y_t-\theta'X_t}{\delta}\right). \tag{11}$$

If our interest in the model is confined to estimates of equation (3) we could stop here but identification of $\beta_2$, $\sigma_2$ and $\sigma_{12}$ requires further inspection. Aside from $\beta_1$ and $\sigma_1$, the likelihood given in (11) contains only $K+1$ parameters. Clearly identification of the $K+2$ parameters $\beta_2$, $\sigma_2$ and $\sigma_{12}$ will require certain restrictions. The analogy to simultaneous equation models provides a necessary condition for identification of eq. (4), namely one restriction among the set $\beta_2, \sigma_2, \sigma_{12}$. For example, if some element of $\beta_2$ is restricted to be zero the necessary condition is satisfied. If the corresponding element in $\beta_1$ is non-zero then the sufficient condition is also satisfied. Likewise, restricting $\sigma_{12}$ to be zero is sufficient for identification.

Before leaving this discussion it is perhaps worth noting that the model expressed by (3)–(5) might, in some applications, represent the reduced form of a more general structural model, say,

$$y_1 = \alpha_1 y_2 + \gamma_1' X + v_1, \tag{3'}$$

and

$$y_2 = \alpha_2 y_1 + \gamma_2' X + v_2. \tag{4'}$$

It is important to recognize that in such cases the discussion above refers only to the reduced form coefficients. Derivation of identifiability conditions for structural parameters in (3′) and (4′) would be straightforward but will not be pursued here for lack of interesting applications to motivate that simultaneous equation model.

With the identifiability question resolved we can return to the maximum likelihood estimation problem. Like the Tobit model, the likelihood function given in eq. (11) involves both density and distribution functions and yields nonlinear normal equations. Estimation, then, requires the use of some iterative maximization alogrithm. One such algorithm, Newton–Raphson, has proven to be highly satisfactory for Tobit and Probit models. But in the author's experimentation with artificial data on this censored model it yielded mixed and generally discouraging results. Two problems became apparent. First, the likelihood surface for this model is not concave over the entire parameter space so the matrix of second derivatives may not be negative definite, as is required for convergence of the Newton algorithm, at any arbitrary point in the space. A modification to that Hession matrix such as the one proposed by Greenstadt (1967) 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 occasionally large enough to overshoot the maximum and drive the variance terms out of the parameter space, resulting in a failure of the procedure. Apparently, the quadratic approximation of the log likelihood used by the Newton algorithm is sufficiently poor far away from the maximum to cause serious problems in determining the optimum step size. An algorithm which proved a bit more stable was a 'Dogleg' algorithm developed by Rick Becker (1974). That algorithm was derived along the lines of Powell's (1970) 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 eq. (3) over the subset of observations $\psi_2$; (b) obtain $\hat{y}_1$ for the subset $\psi_1$ using the OLS estimates; and (c) apply the probit model with observed threshold ($\hat{y}_1$ in the set $\psi_1$ and $Y$ in the set $\psi_2$) to eq. (4). $\sigma_{12}$ was set initially to zero to simplify step (c).

## 3. Information in the sample

A cautious reader might question the ability to estimate parameters of the model discussed here, and particularly those of eq. (4), given the minimal amount of information on the endogenous variables provided by that model. To characterize that estimability we consider in this section the appropriate estimation techniques under varying degrees of information and present for comparison the results of some simulation experiments. The levels of information we consider are as follows: (a) $y_1$ and $y_2$ always observed; (b) $y_2$ and $Y$ observed (i.e., $y_1$ observed only when $y_1 > y_2$); (c) $y_1$ and the sample partition, $\psi_1$ and $\psi_2$, observed; (d) $Y$, $\psi_1$ and $\psi_2$ observed (i.e., the censored model), and (e) only $\psi_1$ and $\psi_2$ observed.

### *3.1. $y_1$ and $y_2$ observed*

The indicated estimation technique here is of course OLS or perhaps Zellner's seemingly unrelated regression procedure for more efficient estimates when $\sigma_{12} \neq 0$ and $\beta_1$ and $\beta_2$ contain zero restrictions. OLS will be used as a baseline, for evaluation of the models to be discussed below, in the simulation experiments at the end of this section.

### *3.2. $Y$ and $y_2$ observed*

As was noted in the introduction the Tobit model is appropriate when the threshold variable, $y_2$, is observed. With observations on $y_2$ the likelihood function to be maximized is

$$L = \prod^{\psi_1} \int_{-\infty}^{Y_2} f(y_1 - \beta_1' X_1, y_2 - \beta_2' X_2) \, \mathrm{d}y_1 \cdot \prod^{\psi_2} f(Y - \beta_1' X_1, y_2 - \beta_2' X_2). \tag{12}$$

If we assume $u_1$ and $u_2$ to be independent the likelihood factors,

$$L = \prod^{\psi_1} \int_{-\infty}^{y_2} f(y_1 - \beta_1' X_1) \, \mathrm{d}y_1 \cdot \prod^{\psi_2} f(Y - \beta_1' X_1) \cdot \prod^{\psi} f(y_2 - \beta_2' X_2), \tag{13}$$

allowing estimation of eq. (3) by Tobit analysis and eq. (4) by OLS separately.[1]

[1]Even if $\sigma_{12} \neq 0$ we might proceed to estimate the equations seperately 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.

Note that no identification problem arises and all parameters are readily estimated in either case.

### 3.3. $y_1$, $\psi_1$ and $\psi_2$ observed

Define an indicator variable $I$ by

$$\begin{aligned} I_t &= 1, \quad \text{if} \quad t \in \psi_2 \quad (y_{1t} \geqq y_{2t}), \\ &= 0, \quad \text{if} \quad t \in \psi_1 \quad (y_{1t} > y_{2t}), \end{aligned} \tag{14}$$

and assume for the moment that $\sigma_{12} = 0$. Now eq. (4) may be estimated with a simple probit model, using $I_t$ as the dichotomous dependent variable. The likelihood function is given by[2]

$$L(\beta_2,,X) = \prod^{\psi_1} P\left(\frac{\beta_2' X - y_1}{\sigma_2}\right) \cdot \prod^{\psi_2} \left[1 - P\left(\frac{\beta_2' X - y_1}{\sigma_2}\right)\right]. \tag{15}$$

Again no identification problem arises and all parameters of the model are readily estimable. But when the zero covariance assumption is dropped, identification is no longer guaranteed without other restrictions. The relevant likelihood becomes

$$\begin{aligned} &L(\beta_1, \beta_2, \sigma_1, \sigma_2, \sigma_{12} \mid I, y_1, X) \\ &= \prod^{\psi_1} \int_{y_1}^{\infty} f(y_1 - \beta_1' X, y_2 - \beta_2' X) \mathrm{d}y_2 \cdot \prod^{\psi_2} \int_{-\infty}^{y_1} f(y_1 - \beta_1' X, y_2 - \beta_2' X) \mathrm{d}y_2 \\ &= \prod^{\psi} \frac{1}{\sigma_1} z\left(\frac{y_1 - \beta_1' X}{\sigma_1}\right) \cdot \prod^{\psi_1} \left[1 - P\left(\frac{y_1 - \theta' X}{\delta}\right)\right] \cdot \prod^{\psi_2} P\left(\frac{y_1 - \theta' X}{\delta}\right), \end{aligned} \tag{16}$$

with $\delta$ and $\theta$ defined by eqs. (9) and (10) of section 1. Here again, as in the censored model, we see that identification requires some restriction on $\beta_1$ and $\beta_2$, say $\beta_{2i} = 0$, $\beta_{1i} \neq 0$ for at least one $i$. Given such a restriction, estimation proceeds by OLS estimation of eq. (3) and estimation of $\delta$ and $\theta$ via probit analysis, followed by solution for $\beta_2$, $\sigma_2$ and $\sigma_{12}$ in eqs. (9) and (10).

### 3.4. $Y$, $\psi_1$ and $\psi_2$ observed

The information content here is that provided by the censored model discussed in section 2. To summarize, the relevant likelihood function is given by

[2]To maximize (15) with a canned probit program, treat $y_1$ as an exogenous variable. Its coefficient is then the reciprocal of the $\sigma_2$ and the invariance property of ML estimates is used to obtain estimates of $\beta$.

eq. (8) and identification requires some linear restriction among the set of parameters $\sigma_{12}$, $\sigma_2$ and $\beta_2$.

### 3.5. $\psi_1$ and $\psi_2$ observed

With only the indicator variable $I$ defined as in eq. (14) observed, Probit analysis may be used to estimate certain combinations of the original parameters but those original parameters are not identified. The likelihood function is given by

$$L(\alpha\,|\,I,X) = \prod^{\psi_1} 1-P(\alpha'X)\cdot\prod^{\psi_2} P(\alpha'X), \tag{17}$$

where

$$\alpha = (\beta_1-\beta_2)/(\sigma_1^2+\sigma_2^2-2\sigma_{12})^{\frac{1}{2}}. \tag{18}$$

The models discussed above provide a means of assessing the 'value of information' in estimation of the parameters of the model given by eqs. (3) and (4). To perform this assessment the following simulation experiment was conducted. A sample of 100 observations was generated for $y_1$ and $y_2$ according to the structure

$$y_1 = \beta_0+\beta_1X_1+\beta_2X_2+u_1, \tag{19}$$

and

$$y_2 = \gamma_0+\gamma_1X_2+\gamma_2X_3+u_2. \tag{20}$$

$X_1$ and $X_3$ were chosen from $N(0,1)$ populations and held fixed in repeated samples, $u_1$ and $u_2$ were drawn from independent $N(0,1)$ populations and the parameter vectors were fixed at $(\beta_0\ \beta_1\ \beta_2) = (\gamma_0\ \gamma_1\ \gamma_2) = (0\ -1\ 1)$. Parameters for both equations were first estimated with OLS. Then a variable $Y$ was constructed according to (5), and (19) was reestimated using Tobit analysis with the variables $Y$, $X_1$ and $X_2$ and the observed threshold $y_2$. Next the indicator variable $I$ [eq. (14)] was constructed and used in the estimation of eq. (20) by Probit analysis with variables $I$, $X_2$ and $X_3$ and the observed threshold $y_1$. Finally, both (19) and (20) were estimated by maximizing the likelihood function given by eq. (8) and using only the variables $Y$, $I$, $X_1$, $X_2$ and $X_3$. The procedure was repeated for 25 samples, and the results of the experiment are summarized in table 1. These results suggest no significant bias in any of the estimation methods, but the variance of the estimators does increase substantially as the level of endogenous information in the sample declines. Using OLS as a base the average variance ratio for Tobit and censored model estimates of eq. (19) parameters

Table 1

Summary statistics for 'value of information' experiment (mean, standard deviation and root mean square errors for estimates across 25 samples).

| | Parameter | *Level of information and estimation method* A (OLS) | B (Tobit) | C (Probit) | D (Censored) |
|---|---|---|---|---|---|
| Mean | $\beta_0$ | −0.038 | −0.070 | | −0.053 |
| | $\beta_1$ | −0.973 | −0.943 | | −0.942 |
| | $\beta_2$ | 0.999 | 1.040 | | 1.013 |
| | $\sigma_1$ | 1.009 | 0.982 | | 0.971 |
| | $\gamma_0$ | 0.007 | | 0.022 | −0.005 |
| | $\gamma_1$ | −0.950 | | −1.150 | −1.080 |
| | $\gamma_2$ | 1.011 | | 1.061 | 1.032 |
| | $\sigma_2$ | 1.013 | | 0.968 | 0.887 |
| Std. dev. | $\beta_0$ | 0.106 | 0.174 | | 0.264 |
| | $\beta_1$ | 0.098 | 0.115 | | 0.120 |
| | $\beta_2$ | 0.096 | 0.131 | | 0.200 |
| | $\sigma_1$ | 0.066 | 0.077 | | 0.204 |
| | $\gamma_0$ | 0.098 | | 0.264 | 0.372 |
| | $\gamma_1$ | 0.097 | | 0.440 | 0.593 |
| | $\gamma_2$ | 0.076 | | 0.247 | 0.277 |
| | $\sigma_2$ | 0.069 | | 0.300 | 0.593 |
| RMSE | $\beta_0$ | 0.110 | 0.184 | | 0.264 |
| | $\beta_1$ | 0.100 | 0.126 | | 0.131 |
| | $\beta_2$ | 0.094 | 0.135 | | 0.196 |
| | $\sigma_1$ | 0.065 | 0.078 | | 0.202 |
| | $\gamma_0$ | 0.097 | | 0.260 | 0.365 |
| | $\gamma_1$ | 0.108 | | 0.456 | 0.586 |
| | $\gamma_2$ | 0.076 | | 0.249 | 0.273 |
| | $\sigma_2$ | 0.068 | | 0.296 | 0.591 |

are 1.8 and 5.4, respectively, and those for probit and censored model estimates of eq. (20) parameters are 14.3 and 34.7, respectively.

As a final note, the feasibility of estimating the censored model when the disturbances are not independent was tested with another small simulation experiment. The structure used was similar to that given by eqs. (19) and (20) except that $u_1$ and $u_2$ were drawn from a bivariate normal distribution with a non-zero covariance. Note that exclusion of $X_1$ from eq. (20) identifies the model. Results of the experiment are reported in table 2.

## 4. Applications of the censored model

The objective of this section is to further motivate the censored variable model by examining in more detail the variety of economic relationships to which it might apply. The durable good purchase problem mentioned in the introduction will be examined more critically, two similar relationships will be noted, and finally a fourth example will be worked out in detail including the examination of estimates obtained from fitting the model to census data.

Table 2

Simulation results on censored model with correlated disturbances (summary results for 10 samples).

| Parameter | True value | Mean estimate | Minimum estimate | Maximum estimate |
|---|---|---|---|---|
| $\beta_0$ | 0.0 | −0.0674 | −0.3358 | 0.3417 |
| $\beta_1$ | −1.0 | −0.9988 | −1.2551 | −0.7079 |
| $\beta_2$ | 1.0 | 0.9844 | 0.8163 | 1.1949 |
| $\delta_0$ | 0.0 | −0.1111 | −0.4919 | 0.3337 |
| $\delta_1$ | −1.0 | −0.9860 | −1.3306 | −0.7853 |
| $\delta_2$ | 1.0 | 0.9859 | 0.6158 | 1.4117 |
| $\sigma_1^2$ | 1.0 | 0.9914 | 0.7131 | 1.3362 |
| $\sigma_2^2$ | 1.0 | 0.7783 | 0.2917 | 1.3159 |
| $\sigma_{12}$ | 0.64 | 0.5405 | 0.3189 | 0.7963 |

| Parameter | Mean bias | Std. dev. | RMSE | $T$ ratio |
|---|---|---|---|---|
| $\beta_0$ | 0.0675 | 0.1911 | 0.2026 | 1.117 |
| $\beta_1$ | −0.0012 | 0.1618 | 0.1618 | −0.023 |
| $\beta_2$ | 0.0156 | 0.0981 | 0.0993 | 0.503 |
| $\delta_0$ | 0.1111 | 0.2373 | 0.2620 | 1.481 |
| $\delta_1$ | −0.0140 | 0.1654 | 0.1660 | −0.268 |
| $\delta_2$ | 0.0141 | 0.2156 | 0.2161 | 0.207 |
| $\sigma_1^2$ | 0.0086 | 0.1839 | 0.1841 | 0.147 |
| $\sigma_2^2$ | 0.2217 | 0.3639 | 0.4261 | 1.926 |
| $\sigma_{12}$ | 0.0995 | 0.1595 | 0.1880 | 1.972 |

### 4.1. *Durable good purchases*

As noted in the introduction, the simple Tobit model with a constant threshold of zero fails to adequately describe the durable goods expenditure problem since it cannot account for the bias away from very small purchases. In the case of automobiles, for instance, no new cars are purchased for \$10 or for \$100 or even \$1000. If there is a threshold constraint operating it must certainly be larger than zero and in fact is not any a priori known constant.[3]

One plausible unobserved and variable constraint operates on the supply side. We might argue that a household enters the market for a particular automobile described in terms of say a certain minimum size. Available in the market is a wide variety of cars of different make, model and option configurations which satisfy the size criterion, each of which represents a different 'quantity' of cars in the sense of the bundle of services we call an automobile. But technological

[3]In Cragg's (1971) treatment of the durable good problem he considered a two-stage model in which a consumer decides first whether to buy a car and second how much to spend. Two possible criticisms of his analysis are that his model, like the Tobit model, does not eliminate small purchases and that the assumption of separate decisions of whether to buy and how much to buy may be unsatisfactory. The threshold approach pursued here, on the other hand, makes the decisions simultaneous and explicitly rules out small purchases.

constraints provide a lower bound on this range of quantities, that lower bound being the cheapest car of a given size which can be produced; there is no finer divisibility of 'quantities' below this level. That lower bound then represents the threshold in the censored model and might depend on, say, size of the family and primary purpose of the car. The household makes a purchase if and only if its desired purchase exceeds that threshold.

An alternative constraint on purchases operates on the demand side and arises from lump sum transactions costs.[4] The utility maximizing level of durable goods purchases in the presence of such costs may in fact yield a level of utility lower than the level associated with no purchase and, therefore, no payment of the transactions cost. The problem may perhaps be easiest to see by using indifference curves. Let $g$ be the level of durable goods and $W$ be an aggregation of all other goods including wealth assets and leisure time. Then let $W^0$ be the initial level of $W$ and $W^c = W^0 - C$ be the level of $W$ remaining after payment of the lump sum transactions cost of $C$, and consider the indifference curve maps in the $g$, $W$ plane of figs. 1a and 1b. The budget constraint in both cases consists of a vertical segment at $g = 0$ from $W^0$ to $W^c$ and a downward sloping segment extending from $W^c$ to the right with a slope determined by the price of $g$ relative to the price of $W$. The level $d$ of $g$ represents a point of tangency of this budget constraint with some indifference curve. In fig. 1a, the household will choose to purchase an amount $D = d$ and thereby attain the level of utility indicated by indifference curve $I_3$. But in fig. 1b the indifference curve tangent to the budget constraint is lower than the curve passing through the point $(0, W^0)$; the household will choose to not make a purchase, $D = 0$, but instead keep $W^0$.

The lower threshold on $g$ below which no purchase will be made is revealed in figs. 1c and 1d. In these figures we draw a line parallel to the budget constraint which is just tangent to the indifference curve passing through the point $(0, W^0)$. The difference, $C_{\max}$, between $W^0$ and $W^a$, which is the point where this new line intersects the $W$ axis, represents the maximum transations cost a household will pay in order to make a non-zero purchase, since if he must pay any greater transactions cost he cannot make a purchase yielding a level of utility as great as the level he attains with no purchase. Thus we argue that a purchase will be made if and only if $C_{\max} \geqq C$. Now associated with $C_{\max}$ is the level $t$ of $g$. Provided $g$ is a superior good we can argue that a purchase will be made if and only if $t \leqq d$.[5]

The elements of the censored model are thus established. $d$ represents the desired purchase and depends on income, wealth and household characteristics. $t$ represents the lower threshold on purchases and depends on the same variables.

[4]Such transactions costs might, for example, represent the time required to locate a dealer, choose a car, sign the contract and accept delivery.

[5]By a superior goods it is meant here that the locus of the points of tangency between indifference curves and the budget constraint, does not bend to the left as the budget constraint shifts outward.

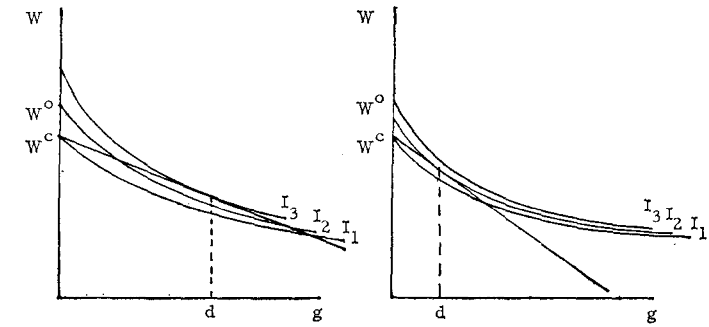


Fig. 1a. A purchase of $d$ yields maximum utility $I_3$.
Fig. 1b. Maximum utility attained with no purchase.

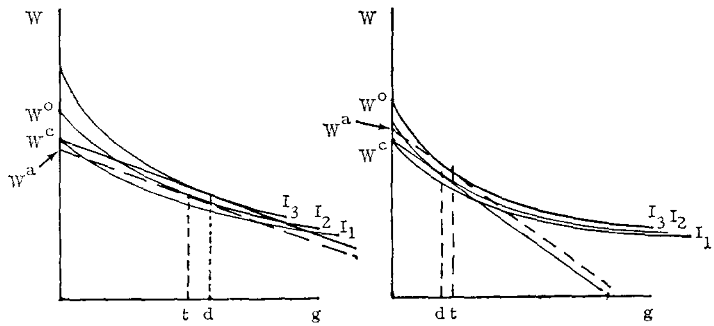


Fig. 1c. The threshold relationship with a purchase of $d$ ($d > t$).
Fig. 1d. The threshold relationship with no purchase ($d < t$).

A purchase equal to $d$ will be made only if $d$ exceeds $t$. Note that with such a specification for $t$ the parameters of its distribution are not identified, but since estimates of $t$ are of little interest this failure is of little consequence. The important parameters, those of the expenditure equation, are identified.[6]

### 4.2. *Advertising expenditures*

As a second example of a censored variable model consider the expenditures by a candidate for political office on a particular advertising medium. The objective of the candidate might be to allocate his campaign funds among alternative

[6]The supply and demand side constraints noted separately above may of course operate simultaneously so that the effective constraint is the maximum of the two. In this case a single regression equation does not adequately describe the threshold, and ML estimates of the censored model of eqs. (3)–(4) are likely to be biased. Derivation of the appropriate likelihood function would be straightforward – it would involve terms similar to those for a switching regressions problem [see Maddala and Nelson (1975)]. But computational difficulties in ML estimation would be severe and certainly not all parameters would be identified.

activities so as to maximize his recognition in the electorate. The problem of interest here is to estimate the demand of candidates for a particular activity such as television advertising. As in the durable goods problem we observe many candidates who purchase no advertising time while others spend large amounts, but no candidates spend trivial amounts, say less than \$1000. The analysis we perform to derive a reasonable econometric model of the expenditure behavior follows the same argument as in the durable goods problem and will thus not be pursued further. The suggestion is again a threshold relationship in which a candidate undertakes a TV campaign only if the desired expenditure exceeds some unobservable and variable threshold.

### *4.3. The needs versus reluctance hypothesis*

A third example arises out of bank borrowing behavior. As a reserve accountability deadline approaches, banks with inadequate reserves must increase those reserves by cashing in assets or taking out short term loans. The two most commonly used loan sources are the federal funds market and the federal reserve's discount window. Observation of bank behavior reveals the apparent paradox that quite often banks borrow from the federal funds market rather than the discount window even though the federal funds rate is higher than the discount rate. One explanation of this paradox is the 'needs versus reluctance' hypothesis which argues that banks are reluctant to too often frequent the discount window for fear of adverse sanctions from the Federal reserve for such behavior.[7]

The empirical problem faced here is a means of testing the relative importance of the two conflicting forces – needs versus reluctance. On the one hand a bank has some demand for funds from the discount window which rises as the ratio of federal funds rate to discount rate rises. That is desired borrowings is an increasing function of the relative gain from using the discount window. On the other hand is the reluctance to use the discount window which rises as the implicit cost of adverse sanctions rises. We might then argue that a bank will borrow from the fed an amount sufficient to meet reserve needs only if the gain from using this source exceeds the implicit cost. If those costs are independent of the amount borrowed, or at least the relationship between desired borrowings and implicit costs is weaker than the one between borrowing and gain, then again there exists a threshold level of desired borrowings below which banks will not use the discount window. The censored model would thus posit desired borrowing as a function of previous trips to the discount window and other activities viewed with disfavor by the federal reserve.

[7]See Polakoff and Sibler (1976) for a more detailed account of the problem.

### 4.4. *Wage and value of time relationships*

As a fourth and final example consider the estimation of wage and labor supply equations.[8] This example is of particular interest because of the useful interpretation of the relevant threshold. This feature and active recent interest in the problem make it an attractive example to use as an illustration of estimation of the censored model. Thus the analysis of the problem and specification of the equations are carried out in more detail than the first three examples and actual estimates based on census data are reported.

Estimation of labor supply relationships at the micro level is often frustrated by the absence 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 non-workers. Such a mechanism is captured in the familiar diagram illustrating indifference curves in the income–leisure plane as in fig. 2.

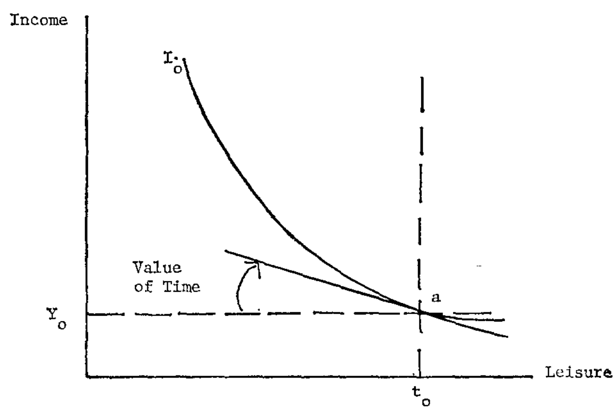


Fig. 2. Value of a non-workers' time.

With a given level of non-labor income ($Y_0$) a non-participant is at point *a* on indifference curve $I_0$ where $t_0$ represents the upper time constraint. The slope of the indifference curve at that point represents the marginal rate at which he would exchange his leisure for additional income. Thus the slope of the tangent to $I_0$ at point *a* represents the value of a non-workers time. Only a wage offer greater than that value will induce him to enter the labor force and thereby reveal to the waiting econometrician his potential wage. If he does not receive

[8]The author is grateful to Sherwin Rosen for suggesting this application.

such an offer he remains out of the labor force, his potential wage is not revealed, and the econometrician remains frustrated.

Clearly this mechanism is a candidate for the application of the censored regression model and estimation procedure discussed in section 2. The potential wage rate is the censored dependent variable and the value of time is the unobserved threshold. Through estimation of the determinants of the latter simultaneously with the wage function the missing data problem is resolved. In fact the value of time relation may be of particular importance in its own right. Such was the case in recent papers by Gronau (1973) and Heckman (1974). Gronau's model corresponds precisely to the censored regression model. Thus his model has been used below for illustrative and comparative purposes. Heckman, on the other hand, viewed the two equations as simultaneous and included hours of work and wage rates as the two endogenous variables. At the end of this section the relation between his and the censored models will be discussed.

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

$$W^p = f(E), \tag{21}$$

$$V = g(C) \tag{22}$$

$$\begin{aligned} W^m &= W^p, \quad \text{if} \quad W^p > V, \\ &= 0, \qquad \text{otherwise,} \end{aligned} \tag{23}$$

where $W^p$ is a housewife's potential wage which depends on her marketable 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 receives if she does in fact enter the labor force. The reader is referred 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 eq. (22). As he discussed and as explained in section 3 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 eq. (22). 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 (1974) 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 HS.
$E_4$ = $C_4$ = dummy variable (0, 1) for education greater than HS.
$C_5$ = family income (in $10,000) net of wife's earnings.
$C_6$ = husband's age (in years).
$C_7$ = dummy variable (0, 1) for husband's education less than HS.
$C_8$ = dummy variable (0, 1) for husband's 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 eq. (22) 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 covariance 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 eq. (21), exclusion of it in (22) would have been sufficient for identification without the zero covariance assumption.

The choice of variables follows Gronau and the reader is referred 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 non-linearities 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 explantory power. Thus the functional form used for the results appearing below was $Y = b_0 b_1^{X1} b_2^{X2} \dots b_k^{Xk} u$ 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 explanation. The procedure used by Gronau was to estimate, via probit analysis, the model

$$
\begin{aligned}
L &= 1, \quad \text{if} \quad b'C+u > \ln(\bar{W}^p), \\
&= 0, \quad \text{if} \quad b'C+u \leqq \ln(\bar{W}^p),
\end{aligned}
$$

where $L$ is the labor force participation indicator and $\bar{W}^p$ was taken to be the geometric average of wages received by working women with characteristics $C_1$ through $C_4$. This was the procedure adopted for use here. Results for the two methods are presented in table 3. 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.

As noted earlier Heckman (1974) looked at the same basic problem but used a different specification of the model. His model formulation is

$$
\begin{aligned}
W^p &= f(E), \\
V &= g(H, C),
\end{aligned}
$$

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 ($H = 0$) $V$ exceeds $W^p$, both are unobserved and the individual drops out of the labor force. To estimate his model Heckman used maximum likelihood, deriving, as in the

Table 3

Estimates of the value of a housewife's time.

| Variable | Censored model | | Probit model | |
|---|---|---|---|---|
| | Coefficient | $t$ ratio | Coefficient | $t$ ratio |
| Constant | −0.4057 | −1.443 | −0.1803 | −0.211 |
| $C_1$ | 0.1518 | 0.982 | 0.1083 | 0.582 |
| $C_2$ | 0.1815 | 1.275 | 0.1373 | 1.395 |
| $C_3$ | −0.0235 | −0.204 | −0.0175 | −0.068 |
| $C_4$ | 0.2166 | 1.731 | 0.2916 | 0.457 |
| $C_5$ | 0.6817 | 5.939 | 0.3635 | 5.685 |
| $C_6$ | 0.1141 | 1.878 | 0.1006 | 2.964 |
| $C_7$ | −0.0276 | −0.282 | 0.0098 | 0.1615 |
| $C_8$ | 0.0616 | 0.596 | 0.0215 | 0.335 |
| $C_9$ | 0.3681 | 3.397 | 0.2614 | 4.554 |
| $C_{10}$ | 0.2004 | 2.690 | 0.1088 | 2.321 |
| $C_{11}$ | 0.1479 | 2.330 | 0.1417 | 4.011 |
| $C_{12}$ | −0.0903 | −1.488 | −0.0123 | −0.327 |
| Std. error | 0.4278 | | 0.4243 | |
| Mean value of time | \$2.61 | | \$2.27 | |
| Constant | 0.2689 | 2.084 | | |
| $E_1$ | −0.0772 | −0.704 | | |
| $E_2$ | −0.0656 | −0.551 | | |
| $E_3$ | −0.2400 | −2.119 | | |
| $E_4$ | 0.2796 | 2.247 | | |
| Std. error | 0.7287 | | | |
| Mean potential wage | \$1.26 | | | |

censored model, $\Pr(g(0,C) > f(E))$ for non-working women and for working women using the pdf representing the joint distribution of $H$ and $W^p(=V)$.

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 represent the value of time for a non-participant, 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. The difference between the models, therefore, is whether $H$ is included in the value of the equation. With $H$ left out of the model, Gronau's interpretation of $V$ and the censored model and its likelihood function apply. With $H$ included, Heckman's interpretation and model are appropriate and a different likelihood function follows.

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 which the rigid hours places the individual on a lower indifference curve than would non-participation. 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.

## References

Becker, R., 1974, Dogleg minimization alogrithm, NBER–CRC memorandum, April.
Cragg, J.G., 1971, Some statistical models for limited dependent variables with applications to the demand for durable goods, Econometrica 39, Sept., 829–844.
Greenstadt, J., 1967, On the relative efficiences of gradient methods, Mathematics of Computation 21, July.
Gronau, R., 1973, The effect of children on the housewife's value of time, Journal of Political Economy, Supp. 81, March, 168–199.
Heckman, J., 1974, Shadow prices, market wages, and labor supply, Econometrica 42, July, 455–462.
Maddala, G.S. and F.D. Nelson, 1975, Switching regression models with exogenous and endogenous switching, American Statistical Association, Proceedings of the Business and Economics Statistics Section.
Polakoff, M.E. and W.L. Sibler, 1967, Reluctance and member bank borrowing: Additional evidence, Journal of Finance 22, March.
Powell, M.J.D., 1970, A new algorithm for unconstrained optimization, in: J.B. Rosen et al., Nonlinear programming (Academic Press, New York).
Tobin, J., 1958, Estimation of relationships for limited dependent variables, Econometrica 26, 24–36.