If you have issues viewing or accessing this file contact us at NCJRS.gov.

National Criminal Justice Reference Service

ncjrs

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.

1.0 2.8 2.5 3.2 2.2 3.6 2.0 4.0 1.1 1.8 1.25 1.4 1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D.C. 20531

9/28/83

88429

Rhodes Associates

Criminal Justice Research Series

TOBIT MODELS: A SURVEY

by

Takeshi Amemiya

July, 1982

CJRS 4

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this ~~copyrighted~~ material has been granted by

Public Domain/LEAA/NIJ
U.S. Department of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the ~~copyright~~ owner.

Prepared under Grant Number 81-IJ-CX-0055 from the National Institute of Justice, Law Enforcement Assistance Administration, U. S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official positions or policies of the U. S. Department of Justice.

Rhodes Associates, 706 Cowper Street, 3rd flr, Palo Alto, California 94301 (415) 326 6246

TOBIT MODELS: A SURVEY

by

Takeshi Amemiya*

I. Introduction

Tobit models refer to regression models in which the range of the dependent variable is constrained in some way. In economics, such a model was first suggested in a pioneering work by Tobin [1958]. He analyzed household expenditure on durable goods using a regression model which specifically took account of the fact that the expenditure (the dependent variable of his regression model) cannot be negative. Tobin called his model the model of limited dependent variables. It and its various generalizations are known popularly among economists as Tobit models, a phrase coined by Goldberger [1964], because of similarities to probit models. These models are also known as censored or truncated regression models.[1/]

Censored and truncated regression models have been developed in other disciplines (notably biometrics and engineering) more or less independently of their development in econometrics. Biometricians use the model to analyze the survival time of a patient. Censoring or truncation occurs either if a patient is still alive at the last observation date or if he or she cannot be located. Similarly, engineers use the

*This research was supported by a National Institute of Justice Grant No. 81-IJ-CX-0055 to Rhodes Associates. The author is grateful to Jeanne E. Anderson, A. Colin Cameron, Yoonbai Kim, and Frederick C. Nold for helpful suggestions.

model to analyze the time to failure of material or of a machine or a system. These models are called survival models.[2/] Sociologists and economists have also used survival models to analyze the duration of such phenomena as unemployment, welfare receipt, employment in a particular job, residing in a particular region, marriage, and the period of time between births.[3/] Mathematically, survival models belong to the same general class of models as Tobit models and share certain characteristics. However, I will not discuss survival models in this survey because they possess special features of their own. They should best be discussed as a topic within the large, separate research area of continuous-time Markov chain models. The interested reader should consult the references I have cited in footnotes 2 and 3 above.

Between 1958, when Tobin's article appeared, and 1970, the Tobit model was used infrequently in econometric applications, but since the early 1970's numerous applications ranging over a wide area of economics have appeared and continue to appear. This phenomenon is clearly due to a recent increase in the availability of micro sample survey data which the Tobit model analyzes well and to a recent advance in computer technology which has made estimation of large-scale Tobit models feasible. At the same time, many generalizations of the Tobit model and various estimation methods for these models have been proposed. In fact, models and estimation methods are now so numerous and diverse that it is difficult for econometricians to keep track of all the existing models and estimation methods and maintain a clear notion as to their relative merits. Thus, it is now particularly useful to examine the current situation to prepare a unified summary and critical assessment of existing results.

I will try to accomplish this objective by means of classifying the diverse Tobit models into five basic types. (A review of the empirical literature has suggested that roughly 95% of the econometric applications of Tobit models fall into one of these five types.) While there are many ways to classify Tobit models, I have chosen to classify them according to the form of the likelihood function. This way seems to me to be the statistically most useful classification because a similarity of the likelihood function implies a similarity of the appropriate estimation and computation methods. It is interesting to note that two models which superficially seem to be very different from each other can be shown to belong to the same type when they are classified according to my scheme.

The remainder of the paper consists of two sections; Section II deals with the Standard Tobit model (or Type 1 Tobit) and Section III deals with the remaining four types of models. Basic estimation methods, which with a slight modification can be applied to any of the five types, are discussed at great length in Section II. More specialized estimation methods are discussed in relevant passages throughout the paper. Each model is illustrated with a few empirical examples.

I should note the topics, in addition to the survival models mentioned above, which I do not discuss. I do not discuss disequilibrium models except for a few basic models which are examined in Section III.E.5. Some general references on disequilibrium models are cited above. Nor do I discuss the related topic of switching regression models. For a survey on these topics, the reader should consult Maddala [1980]. I do not discuss Tobit models for panel data (individuals observed through time), except to mention a few papers in relevant passages, since they can be best discussed with survival models.

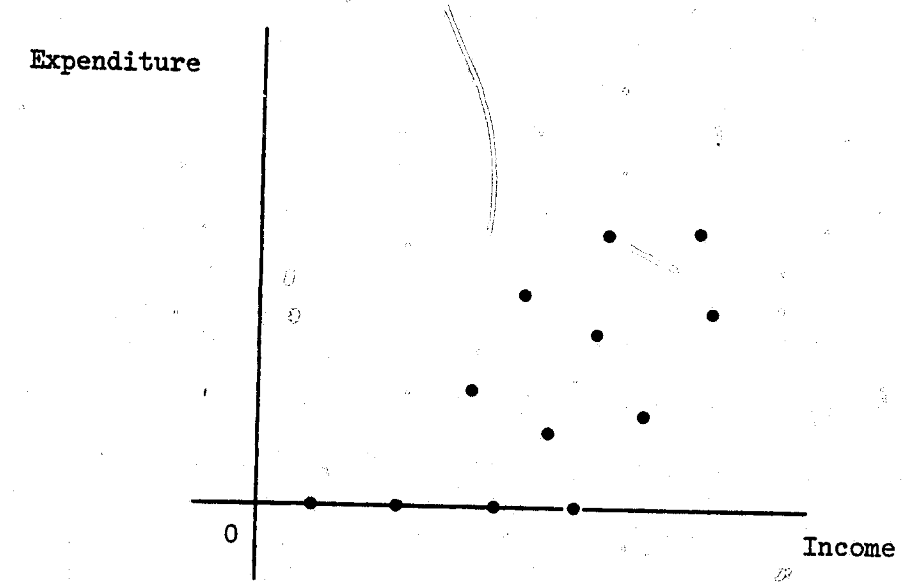
The econometrics text books which discuss Tobit models (with the relevant page numbers) are Goldberger [1964], pp. 251-255; Maddala [1977c], pp. 162-171; and Judge, Griffiths, Hill, and Lee [1980], pp. 609-616. Maddala's survey paper [1980] mentioned above also contains some discussion of Tobit models. There are two more short surveys by Maddala [1977a and b]. However, none of these references offer a comprehensive discussion of Tobit models.

II. Standard Tobit Model (Type 1 Tobit)

A. Definition of the Model

Tobin [1958] noted that the observed relationship between household expenditures on a durable good and household incomes looks like Figure 2.1, where each dot represents an observation of a particular household. An important characteristic of the data is that there are several observations where the expenditure is zero. This feature destroys the linearity assumption so that the least squares method is clearly inappropriate. Should one fit a nonlinear relationship? First, one must determine a statistical model which can generate the kind of data depicted in Figure 2.1. In doing so the first fact one should recognize is that one cannot use any continuous density to explain the conditional distribution of expenditure given income because a continuous density is

Figure 2.1



inconsistent with the fact that there are several observations at zero. Below I develop a crude utility maximization model to explain the phenomenon in question.

Define the symbols needed for the utility maximization model as follows:

y ... a household's expenditure on a durable good,

y_0 ... the price of the cheapest available durable good,

z ... all the other expenditure,

x ... income.

A household is assumed to maximize utility $U(y,z)$ subject to the budget constraint $y + z \leq x$ and the boundary constraint $y \geq y_0$ or $y = 0$. Suppose y^* is the solution of the maximization subject to $y + z \leq x$ but ignoring the other constraint, and assume $y^* = \beta_1 + \beta_2 x + u$, where u may be interpreted as the collection of all the unobservable variables which affect the utility function. Then, the solution to the original problem, denoted by y, can be defined by

$$(2.1) \qquad \begin{cases} y = y^* & \text{if} \quad y^* > y_0 \\ = 0 \quad \text{or} \quad y_0 & \text{if} \quad y^* \leq y_0 \ . \end{cases}$$

If we assume that u is a random variable and that y_0 varies with households but is assumed known, this model will generate data like Figure 2.1. We can write the likelihood function for n independent observations from the model (2.1) as:

$$(2.2) \qquad L = \prod_0 F_i(y_{0i}) \prod_1 f_i(y_i) \quad ,$$

where F_i and f_i are the distribution and density function respectively of y_i^*, $\prod_0$ means the product over those i for which $y_i^* \leq y_{0i}$, and $\prod_1$ means the product over those i for which $y_i^* > y_{0i}$. Note that the actual value of y when $y^* \leq y_0$ has no effect on the likelihood function. Therefore, the second line of equation (2.1) may be changed to the statement "if $y^* \leq y_0$, one merely observes that fact."

The model originally proposed by Tobin [1958] is essentially the same as the above except that he specifically assumed y^* to be normally distributed and assumed y_0 to be the same for all households. We will define the Standard Tobit model (or Type 1 Tobit) as follows:

$$(2.3) \qquad y_i^* = x_i'\beta + u_i \quad , \quad i = 1,2,\ldots,n \ ,$$

$$(2.4) \qquad y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* > 0 \\ 0 & \text{if} \quad y_i^* \leq 0 \ , \end{cases}$$

where $\{u_i\}$ are assumed to be i.i.d. drawings from $N(0,\sigma^2)$. It is assumed that $\{y_i\}$ and $\{x_i\}$ are observed for $i = 1,2,\ldots,n$ but $\{y_i^*\}$ are unobserved if $y_i^* \leq 0$. Defining $\underset{\sim}{X}$ to be the $n \times K$ matrix whose i-th row is x_i', we assume that $\lim_{n\to\infty} n^{-1}\underset{\sim}{X}'\underset{\sim}{X}$ is positive definite.[4/]
As I stated in the previous paragraph, observing $y_i = 0$ is equivalent to observing $y_i^* \leq 0$.

Note that $y_i^* > 0$ and $y_i^* \leqq 0$ in (2.4) may be changed to $y_i^* > y_0$ and $y_i^* \leq y_0$ without essentially changing the model, whether y_0 is known or unknown, since y_0 can be absorbed into the constant term of the regression. If, however, y_{0i} changes with i and is known for every i, the model is slightly changed because the resulting model would be essentially equivalent to the model defined by (2.3) and (2.4) where one of the elements of β other than the constant term is known. The model where y_{0i} changes with i and is unknown is not generally estimable.

Though not needed immediately, it will be later useful to define the binary variable w_i by

$$(2.5) \qquad w_i = \begin{cases} 1 & \text{if} \quad y_i^* > 0 \\ 0 & \text{if} \quad y_i^* \leqq 0 \end{cases} .$$

The likelihood function of the Standard Tobit model is given by

$$(2.6) \qquad L = \prod_0 [1 - \Phi(x_i'\beta/\sigma)] \prod_1 \sigma^{-1}\phi[(y_i - x_i'\beta)/\sigma] \, ,$$

where Φ and ϕ are the distribution and density function respectively of the standard normal variable.

The Tobit model belongs to what is sometimes known as the <u>censored</u> regression model. In contrast, if one observes neither y_i nor x_i when $y_i^* \leqq 0$, the model is known as a <u>truncated</u> regression model. The likelihood function of the truncated version of the Tobit model can be written as

$$(2.7) \qquad L = \prod_1 \Phi(x_i'\beta/\sigma)^{-1} \sigma^{-1}\phi[(y_i - x_i'\beta)/\sigma] \; .$$

Henceforth, the Standard Tobit model refers to the model defined by (2.3) and (2.4), namely a censored regression model, and the model whose likelihood function is given by (2.7) will be called the truncated Standard Tobit model.

B. <u>Empirical Examples</u>

Tobin [1958] obtained the maximum likelihood estimates of his model applied to data on 735 nonfarm households obtained from Surveys of Consumer Finances. The dependent variable of his estimated model was actually the ratio of total durable goods expenditure to disposable income and the independent variables were the age of the head of the household and the ratio of liquid assets to disposable income.

Since then, and especially since the early 1970's, numerous applications of the Standard Tobit model have appeared in economic journals, encompassing a wide range of fields in economics. I will present below a brief list of recent representative papers, with a description of the dependent variable and the main independent variables. In all the papers except Kotlikoff, who uses a two-step estimation method which I will discuss later, the method of estimation is maximum likelihood.

<u>Adams [1980]</u>

y: Inheritance.

x: Income, marital status, number of children.

Ashenfelter and Ham [1979]

y: Ratio of unemployed hours to employed hours.

x: Years of schooling, working experience.

Fair [1978]

y: Number of extra-marital affairs.

x: Sex, age, number of years married, number of children, education, occupation, degree of religiousness.

Keeley, Robins, Spiegelman, and West [1978]

y: Hours worked after a Negative Income Tax program.

x: Pre-program hours worked, change in the wage rate, family characteristics.

Kotlikoff [1979]

y: Expected age of retirement.

x: Ratio of social security benefits lost at full time work to full time earnings.

Reece [1979]

y: Charitable contributions.

x: Price of contributions, income.

Rosenzweig [1980]

y: Annual days worked.

x: Wages of husbands and wives, education of husbands and wives, income.

Stephenson and McDonald [1979]

y: Family earnings after a Negative Income Tax program.

x: Earnings before the program, husband's and wife's

education, other family characteristics, unemployment rate, seasonal dummies.

Wiggins [1981]

y: Annual marketing of new chemical entities.

x: Research expenditure of the pharmaceutical industry, stringency of government regulatory standards.

Witte [1980]

y: Number of arrests (or convictions) per month after release from prison.

x: Accumulated work release funds, number of months after release until first job, wage rate after release, age, race, drug use.

C. Properties of Estimators Under Standard Assumptions

In this section I will discuss the properties of various estimators of the Tobit model under the assumptions of the model. The estimators I consider are probit maximum likelihood (ML), least squares (LS), Heckman's two-step, nonlinear least squares (NLLS), nonlinear weighted least squares (NLWLS), and the Tobit ML.

1. Probit MLE: The Tobit likelihood function (2.6) can be trivially rewritten as follows:

(2.8)
$$L = \prod_0 [1 - \Phi(x_i'\beta/\sigma)] \prod_1 \Phi(x_i'\beta/\sigma) \cdot \prod_1 \Phi(x_i'\beta/\sigma)^{-1}\sigma^{-1}\phi[(y_i - x_i'\beta)/\sigma] .$$

Then, the first two products of the right-hand side of (2.8) constitute the likelihood function of a probit model, and the last product is the likelihood function of the truncated Tobit model as given in (2.7). The probit ML estimator of $\alpha \equiv \beta/\sigma$, denoted $\hat{\alpha}$, is obtained by maximizing the logarithm of the first two products. The maximization must be done by an interation scheme such as Newton-Raphson or the method of scoring (see Amemiya [1981b, p. 1496]), with convergence always assured by the global concavity of the logarithmic likelihood function.[5/]

The probit MLE is consistent and one can show by a standard method (see, for example, Amemiya [1978, p. 1196])

$$(2.9) \qquad \hat{\alpha} - \alpha \stackrel{A}{=} (X' D_1 X)^{-1} X' D_1 D_0^{-1} (w - Ew) \quad ,$$

where D_0 is the $n \times n$ diagonal matrix whose i-th element is $\phi(x_i'\alpha)$, D_1 is the $n \times n$ diagonal matrix whose i-th element is $\Phi(x_i'\alpha)^{-1}[1 - \Phi(x_i'\alpha)]^{-1}\phi(x_i'\alpha)^2$ and w is the vector whose i-th element is the w_i defined in (2.5). (See footnote 4 for usage of the symbol ~.) Note that the i-th element of Ew is equal to $\Phi(x_i'\alpha)$. The symbol $\stackrel{A}{=}$ means that both sides have the same asymptotic distribution.[6/] Therefore, $\hat{\alpha}$ is asymptotically normal with mean α and asymptotic variance-covariance matrix given by

$$(2.10) \qquad V\hat{\alpha} = (X' D_1 X)^{-1} \quad .$$

Note that one can only estimate the ratio β/σ by this method and not β or σ separately. Since the estimator ignores a part of

the likelihood function that involves β and σ, it is not fully efficient. This loss of efficiency is not surprising when one realizes that the estimator uses only the sign of y_i^*, ignoring its numerical value even when it is observed. The main usefulness of the estimator is for providing the first step of Heckman's two-step estimator as I will show later.

2. LS: From Figure 2.1 it is clear that the least squares regression of expenditure on income using all the observations including zero expenditures yields biased estimates. Though it is not so clear from the figure, the least squares regression using only the positive expenditures also yields biased estimates. I will demonstrate these facts mathematically.

First, I will consider the regression using only positive observations of y_i. We get from (2.3) and (2.4)

$$(2.11) \qquad E(y_i | y_i > 0) = x_i'\beta + E(u_i | u_i > -x_i'\beta) \quad .$$

The last term of the right-hand side of (2.11) is generally nonzero (even without assuming u_i is normal). This implies the biasedness of the LS estimator using positive observations on y_i under more general models than the Standard Tobit model. When we assume normality of u_i as in the Tobit model, (2.11) can be shown by straightforward integration to be

$$(2.12) \qquad E(y_i | y_i > 0) = x_i'\beta + \sigma\lambda(x_i'\beta/\sigma) \quad ,$$

where $\lambda(z) = \phi(z)/\Phi(z)$.[7/] As I will show below, this equation plays a key role in the derivation of Heckman's two-step, NLLS, and NLWLS estimators.

Equation (2.12) clearly indicates that the LS estimator of β is biased and inconsistent, but the direction and magnitude of the bias or inconsistency cannot be shown without further assumptions. Goldberger [1981] evaluated the asymptotic bias (the probability limit minus the true value) assuming that the elements of x_i, except the first element which is assumed to be a constant, are normally distributed. More specifically, Goldberger rewrites (2.3) as

(2.13) $$y_i^* = \beta_0 + \bar{x}_i'\beta_1 + u_i$$

and assumes $\bar{x}_i \sim N(0,\Sigma)$ and is distributed independently of u_i. (Here, the assumption of zero mean involves no loss of generality since a nonzero mean can be absorbed into β_0.) Under this assumption he obtains

(2.14) $$\text{plim } \hat{\beta}_1 = \frac{1-\gamma}{1-\rho^2\gamma}\,\beta_1 \quad ,$$

where $\gamma = \sigma_y^{-1}\lambda(\beta_0/\sigma_y)[\beta_0 + \sigma_y\lambda(\beta_0/\sigma_y)]$ and $\rho^2 = \sigma_y^{-2}\beta_1'\Sigma\beta_1$, where $\sigma_y^2 = \sigma^2 + \beta_1'\Sigma\beta_1$. It can be shown that $0 < \gamma < 1$ and $0 < \rho^2 < 1$; therefore, (2.14) shows that $\hat{\beta}_1$ shrinks β_1 toward zero. It is remarkable that the degree of shrinkage is uniform in all the elements of β_1. However, it is not known whether a similar result will hold if $\bar{x}_i$ is not normal. Goldberger gives a nonnormal example where $\beta_1 = (1,1)'$ and $\text{plim } \hat{\beta}_1 = (1.111, 0.887)'$.

Next, I will consider the regression using all the observations of y_i, both positive and zero. To see that the least squares estimator is also biased in this case, one should look at the unconditional mean of y_i:

(2.15) $$Ey_i = \Phi(x_i'\beta/\sigma)\cdot x_i'\beta + \sigma\phi(x_i'\beta/\sigma) \quad .$$

Writing (2.3) again as (2.13) and using the same assumptions as Goldberger, Greene [1981] showed

(2.16) $$\text{plim } \tilde{\beta}_1 = \Phi(\beta_0/\sigma_y)\cdot\beta_1 \quad ,$$

where $\tilde{\beta}_1$ is the LS estimator of β_1 in the regression of y_i on x_i using all the observations. This result is even more remarkable than (2.14) because it implies that $(n/n_1)\cdot\tilde{\beta}_1$ is a consistent estimator of β_1, where n_1 is the number of positive observations of y_i. Unfortunately, however, one cannot confidently use this estimator without knowing its properties when the true distribution of $\bar{x}_i$ is not normal.

3. Heckman's Two-Step Estimator: Heckman [1976], following a suggestion of Gronau [1974], proposed a two-step estimator in a two-equation generalization of the Tobit model. I classify this model as the Type 3 Tobit model and discuss it later. But his estimator can also be used in the Standard Tobit model, as well as in more complex

Tobit models, with only a minor adjustment. I will discuss the estimator in the context of the Standard Tobit model because all the basic features of the method can be revealed in this model. However, one should keep in mind that since the method requires the computation of the probit MLE, which itself requires an iterative method, the computational advantage of the method over the Tobit MLE (which is more efficient) is not as great in the Standard Tobit model as it is in more complex Tobit models.

To explain this estimator, it is useful to rewrite (2.12) as

(2.17) $$y_i = x_i'\beta + \sigma\lambda(x_i'\alpha) + \varepsilon_i \quad , \quad \text{for } i \text{ such that } y_i > 0 \quad ,$$

where I have written $\alpha \equiv \beta/\sigma$ as before and $\varepsilon_i = y_i - E(y_i|y_i > 0)$ so that $E\varepsilon_i = 0$. The variance of ε_i is given by

(2.18) $$V\varepsilon_i = \sigma^2 - \sigma^2 x_i'\alpha\lambda(x_i'\alpha) - \sigma^2\lambda(x_i'\alpha)^2 \quad .$$

Thus, (2.17) is a heteroscedastic nonlinear regression model with n_1 observations. The estimation method Heckman proposed consists of the following two steps: (1) Estimate α by the probit MLE (denoted $\hat{\alpha}$) defined earlier. (2) Regress y_i on x_i and $\lambda(x_i'\hat{\alpha})$ by least squares using only the positive observations on y_i.

To facilitate further the discussion of Heckman's estimator, rewrite (2.17) again as

(2.19) $$y_i = x_i'\beta + \sigma\lambda(x_i'\hat{\alpha}) + \varepsilon_i + \eta_i \quad , \quad \text{for } i \text{ such that } y_i > 0 \quad ,$$

where $\eta_i = \sigma[\lambda(x_i'\alpha) - \lambda(x_i'\hat{\alpha})]$. I will write (2.19) in vector notation as

(2.20) $$y = X\beta + \sigma\hat{\lambda} + \varepsilon + \eta \quad ,$$

where the vectors y, $\hat{\lambda}$, ε, and η have n_1 elements and the matrix X has n_1 rows, corresponding to the positive observations of y_i.[8/] I will further rewrite (2.20) as

(2.21) $$y = \hat{Z}\gamma + \varepsilon + \eta \quad ,$$

where I have defined $\hat{Z} = (X, \hat{\lambda})$ and $\gamma = (\beta', \sigma)'$. Then, Heckman's two-step estimator of γ is defined as

(2.22) $$\hat{\gamma} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'y \quad .$$

The consistency of $\hat{\gamma}$ follows easily from (2.21) and (2.22). I will derive its asymptotic distribution for the sake of completeness, though the result is a special case of Heckman's result [1979]. From (2.21) and (2.22) we have

(2.23) $$\sqrt{n_1}(\hat{\gamma} - \gamma) = (n_1^{-1}\hat{Z}'\hat{Z})^{-1}(n_1^{-\frac{1}{2}}\hat{Z}'\varepsilon + n_1^{-\frac{1}{2}}\hat{Z}'\eta) \quad .$$

Since the probit MLE $\hat{\alpha}$ is consistent, we have

(2.24) $$\underset{n_1\to\infty}{\text{plim}}\; n^{-1}\hat{Z}'\hat{Z} = \lim_{n_1\to\infty} n_1^{-1}Z'Z \quad ,$$

where $Z = (X, \lambda)$. It is easy to prove

$$(2.25) \qquad n_1^{-\frac{1}{2}} \hat{Z}'\varepsilon \to N(0, \lim n_1^{-1} Z'D_2 Z) \quad ,$$

where $D_2 \equiv E\varepsilon\varepsilon'$ is the $n_1 \times n_1$ diagonal matrix whose diagonal elements are $V\varepsilon_i$ given in (2.18). We have by Taylor expansion of $\lambda(x'\hat{\alpha})$ around $\lambda(x'\alpha)$

$$(2.26) \qquad \eta \cong -\sigma \frac{\partial\lambda}{\partial\alpha'} (\hat{\alpha} - \alpha) \quad .$$

Using (2.26) and (2.9) we can prove

$$(2.27) \qquad n_1^{-\frac{1}{2}} \hat{Z}'\eta \to N[0, \sigma^2 Z'D_3 \underset{\sim}{X} (\underset{\sim}{X}'\underset{\sim}{D}_1 \underset{\sim}{X})^{-1} \underset{\sim}{X}'D_3 Z]$$

where $\underset{\sim}{D}_1$ was defined after (2.9) and D_3 is the $n_1 \times n_1$ diagonal matrix whose diagonal elements are $x_i'\alpha\lambda(x_i'\alpha) + \lambda(x_i'\alpha)^2$. Next, note that ε and η are uncorrelated because η is asymptotically a linear function of $\underset{\sim}{w}$ on account of (2.9) and (2.26) and ε and $\underset{\sim}{w}$ are uncorrelated. Therefore, from (2.23), (2.24), (2.25), and (2.27) we finally conclude that $\hat{\gamma}$ is asymptotically normal with mean γ and the asymptotic variance-covariance matrix given by

$$(2.28) \qquad V\hat{\gamma} = (Z'Z)^{-1} Z'[D_2 + \sigma^2 D_3 \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{D}_1\underset{\sim}{X})^{-1}\underset{\sim}{X}'D_3] Z (Z'Z)^{-1} \quad .$$

It is interesting to note that the second matrix within the square bracket above arises because λ had to be estimated. If λ were known, one could apply least squares directly to (2.17) and the exact variance-covariance matrix would be $(Z'Z)^{-1}Z'D_2Z(Z'Z)^{-1}$.

Heckman's two-step estimator uses the conditional mean of y_i given in (2.12). A similar procedure can also be applied to the unconditional mean of y_i given by (2.15).[9/] That is to say, one can regress all the observations of y_i including zeros on Φx_i and ϕ after replacing the α that appears in the argument of Φ and ϕ by the probit MLE $\hat{\alpha}$. In the same way as we derived (2.17) and (2.19) from (2.12), we can derive the following two equations from (2.15):

$$(2.29) \qquad y_i = \Phi(x_i'\alpha)[x_i'\beta + \sigma\lambda(x_i'\alpha)] + \delta_i$$

and

$$(2.30) \qquad y_i = \Phi(x_i'\hat{\alpha})[x_i'\beta + \sigma\lambda(x_i'\hat{\alpha})] + \delta_i + \xi_i \quad ,$$

where $\delta_i = y_i - Ey_i$ and $\xi_i = [\Phi(x_i'\alpha) - \Phi(x_i'\hat{\alpha})]x_i\beta + \sigma[\phi(x_i'\alpha) - \phi(x_i'\hat{\alpha})]$. A vector equation comparable to (2.21) is

$$(2.31) \qquad \underset{\sim}{y} = \underset{\sim}{\hat{D}}\underset{\sim}{\hat{Z}}\gamma + \underset{\sim}{\delta} + \underset{\sim}{\xi} \quad ,$$

where $\underset{\sim}{\hat{D}}$ is the $n \times n$ diagonal matrix whose i-th element is $\Phi(x_i'\hat{\alpha})$. Note that the vectors and matrices are underlined with a "~" because they consist of n elements or rows. The two-step estimator of γ based on all the observations, denoted $\tilde{\gamma}$, is defined as

$$(2.32)\qquad \tilde{\gamma} = (\hat{Z}'\hat{D}^2\hat{Z})^{-1}\hat{Z}'\hat{D}y \ .$$

The estimator can easily be shown to be consistent. To derive its asymptotic distribution, we obtain from (2.31) and (2.32)

$$(2.33)\qquad \sqrt{n}(\tilde{\gamma} - \gamma) = (n^{-1}\hat{Z}'\hat{D}^2\hat{Z})^{-1}(n^{-\frac{1}{2}}\hat{Z}'\hat{D}\delta + n^{-\frac{1}{2}}\hat{Z}'\hat{D}\xi) \ .$$

Here, unlike the previous case, an interesting fact emerges: by expanding $\Phi(x_i'\hat{\alpha})$ and $\phi(x_i'\hat{\alpha})$ in Taylor series around $x_i'\alpha$ one can show $\xi_i = O(n^{-1})$. Therefore,

$$(2.34)\qquad \text{plim}\ n^{-\frac{1}{2}}\hat{Z}'\hat{D}\xi = 0 \ .$$

Corresponding to (2.24), we have

$$(2.35)\qquad \text{plim}\ n^{-1}\hat{Z}'\hat{D}^2\hat{Z} = \lim n^{-1}Z'D^2Z \ ,$$

where D is obtained from $\hat{D}$ by replacing $\hat{\alpha}$ with α. Corresponding to (2.25), we have

$$(2.36)\qquad n^{-\frac{1}{2}}\hat{Z}'\hat{D}\delta \rightarrow N(0, \lim n^{-1}Z'D^2D_4Z) \ ,$$

where $D_4 \equiv E\delta\delta'$ is the $n \times n$ diagonal matrix whose i-th element is $\Phi(x_i'\alpha)[(x_i'\beta)^2 + \sigma^2 x_i'\alpha\lambda(x_i'\alpha) + \sigma^2] - [\Phi(x_i'\alpha)x_i'\beta + \sigma\phi(x_i'\alpha)]^2$. Therefore, from (2.33)∿(2.36), we conclude that $\tilde{\gamma}$ is asymptotically normal with mean γ and the asymptotic variance-covariance matrix given by

$$(2.37)\qquad V\tilde{\gamma} = (Z'D^2Z)^{-1}Z'D^2D_4Z(Z'D^2Z)^{-1} \ .$$

Which of the two estimators $\hat{\gamma}$ and $\tilde{\gamma}$ is preferred? Unfortunately, the difference of the two matrices given by (2.28) and (2.37) is neither positive definite nor negative definite for all the parameter values. Further study is needed on the comparison of the estimators.

Both (2.21) and (2.31) represent heteroscedastic regression models. Therefore, one can obtain asymptotically more efficient estimators by using weighted least squares (WLS) in the second step of the procedure for obtaining $\hat{\gamma}$ and $\tilde{\gamma}$. In doing so, one must use a consistent estimate of the asymptotic variance-covariance matrix of $\varepsilon + \eta$ for the case of (2.21) and of $\delta + \xi$ for the case of (2.31). Since these matrices depend on γ, an initial consistent estimate of γ (say, $\hat{\gamma}$ or $\tilde{\gamma}$) is needed to obtain the WLS estimators. I call these WLS estimators $\hat{\gamma}_W$ and $\tilde{\gamma}_W$ respectively. It can be shown that they are consistent and asymptotically normal with the asymptotic variance-covariance matrices given by

$$(2.38)\qquad V\hat{\gamma}_W = \{Z'[D_2 + \sigma^2 D_3X(X'D_1X)^{-1}X'D_3]^{-1}Z\}^{-1}$$

and

$$(2.39)\qquad V\tilde{\gamma}_W = (Z'D^2D_4^{-1}Z)^{-1} \ .$$

Again, one cannot make a definite comparison between these two matrices.

4. <u>NLLS and NLWLS Estimators</u>: In this subsection I will consider four estimators: the NLLS and NLWLS estimators applied to (2.17),

denoted $\hat{\gamma}_N$ and $\hat{\gamma}_{NW}$ respectively, and the NLLS and NLWLS estimators applied to (2.29), denoted $\tilde{\gamma}_N$ and $\tilde{\gamma}_{NW}$.

All these estimators are consistent and their asymptotic distributions can be obtained straightforwardly by noting that all the results of a linear regression model hold asymptotically for a nonlinear regression model if we treat the derivative of the nonlinear regression function with respect to the parameter vector as the regression matrix.[11] In this way one can show the interesting fact that $\tilde{\gamma}_N$ and $\tilde{\gamma}_{NW}$ have the same asymptotic distributions as $\tilde{\gamma}$ and $\tilde{\gamma}_W$ respectively.[12] One can also show that $\hat{\gamma}_N$ and $\hat{\gamma}_{NW}$ are asymptotically normal with mean γ and with their respective asymptotic variance-covariance matrices given by

$$(2.40) \qquad V\hat{\gamma}_N = (S'S)^{-1}S'D_2S(S'S)^{-1}$$

and

$$(2.41) \qquad V\hat{\gamma}_{NW} = (S'D_2^{-1}S)^{-1} \quad ,$$

where $S = (\sigma^{-2}D_2X, D_5\lambda)$, where D_5 is the $n_1 \times n_1$ diagonal matrix whose i-th element is $1 + (x_i'\alpha)^2 + x_i'\alpha\lambda(x_i'\alpha)$. It seems that one cannot make a definite comparison either between (2.28) and (2.40) or between (2.38) and (2.41).

In the two-step methods defining $\hat{\gamma}$ and $\tilde{\gamma}$ and their generalizations $\tilde{\gamma}_W$ and $\tilde{\gamma}_W$, one can natrually define an iteration procedure by repeating

the two-steps. For example, having obtained $\hat{\gamma}$, one can obtain a new estimate of α, insert it into the argument of λ, and apply least squares again to equation (2.17). The procedure is to be repeated until a sequence of estimates of α thus obtained converges. In the iteration starting from $\hat{\gamma}_W$, one uses the m-th round estimate of γ not only to evaluate λ but also to estimate the variance-covariance matrix of the error term for the purpose of obtaining the $(m + 1)$-st round estimate. Iterations starting from $\tilde{\gamma}$ and $\tilde{\gamma}_W$ can be similarly defined but are probably not worthwhile because $\tilde{\gamma}$ and $\tilde{\gamma}_W$ are asymptotically equivalent to $\tilde{\gamma}_N$ and $\tilde{\gamma}_{NW}$ as I have indicated above. The estimators $(\hat{\gamma}_N, \hat{\gamma}_{NW}, \tilde{\gamma}_N, \tilde{\gamma}_{NW})$ are clearly stationary values of the iterations starting from $(\hat{\gamma}, \hat{\gamma}_W, \tilde{\gamma}, \tilde{\gamma}_W)$. However, they may not necessarily be the converging values.

A simulation study by Wales and Woodland [1980] based on only one replication with sample sizes of 1000 and 5000 showed that $\hat{\gamma}_N$ is distinctly inferior to the MLE and is rather unsatisfactory.

5. <u>The Tobit MLE</u>: The likelihood function of the Tobit model was given in (2.6), from which we obtain the logarithmic likelihood function

$$(2.42) \qquad \log L = \sum_0 \log [1 - \Phi(x_i'\beta/\sigma)] - \frac{n_1}{2} \log \sigma^2 - \frac{1}{2\sigma^2}\sum_1 (y_i - x_i'\beta)^2 \quad .$$

The derivatives are given by

$$(2.43) \qquad \frac{\partial \log L}{\partial \beta} = -\frac{1}{\sigma}\sum_0 \frac{\phi(x_i'\beta/\sigma)x_i}{1 - \Phi(x_i'\beta/\sigma)} + \frac{1}{\sigma^2}\sum_1 (y_i - x_i'\beta)x_i$$

and

$$(2.44) \qquad \frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^3}\sum_0 \frac{x_i'\beta\,\phi(x_i'\beta/\sigma)}{1 - \Phi(x_i'\beta/\sigma)} - \frac{n_1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_1 (y_i - x_i'\beta)^2 \ .$$

Amemiya [1973] proved that the Tobit MLE is strongly consistent and asymptotically normal with the asymptotic variance-covariance matrix equal to $-(\partial^2 \log L/\partial\theta\partial\theta')^{-1}$, where $\theta = (\beta', \sigma^2)'$. The formulae for the second derivatives are given on p. 1000 of Amemiya [1973]. The asymptotic variance-covariance matrix may also be estimated by $-(E\ \partial^2 \log L/\partial\theta\partial\theta')^{-1}$, which is given on p. 1007 of the same reference.

The Tobit MLE is defined as a solution of the equations obtained by equating the partial derivatives (2.43) and (2.44) to zero. The equations are nonlinear in the parameters and hence must be solved iteratively. However, Olsen [1978a] proved the global concavity of $\log L$ in the Tobit model, which implies that a standard iterative method such as Newton-Raphson or the method of scoring always converges to the global maximum of $\log L$. Olsen proved this result by transforming the original parameters of the model to $\alpha = \beta/\sigma$ and $h = \sigma^{-1}$. The $\log L$ in terms of the new parameters can be written as

$$(2.45) \qquad \log L = \sum_0 \log\,[1 - \Phi(x_i'\alpha)] + n_1 \log h - \frac{1}{2}\sum_1 (h y_i - x_i'\alpha)^2 \ ,$$

from which Olsen obtains

$$(2.46) \qquad \begin{bmatrix} \frac{\partial^2 \log L}{\partial\alpha\partial\alpha'} & \frac{\partial^2 \log L}{\partial\alpha\partial h} \\ \frac{\partial^2 \log L}{\partial h\partial\alpha'} & \frac{\partial^2 \log L}{\partial h^2} \end{bmatrix} = \begin{bmatrix} \sum_0 \frac{\phi_i}{1-\Phi_i}\left(x_i'\alpha - \frac{\phi_i}{1-\Phi_i}\right)x_i x_i' & 0 \\ 0 & -\frac{n_1}{h^2} \end{bmatrix} - \begin{bmatrix} \sum_1 x_i x_i' & -\sum_1 x_i y_i \\ -\sum_1 y_i x_i' & \sum_1 y_i^2 \end{bmatrix} ,$$

where $\phi_i = \phi(x_i'\alpha)$ and $\Phi_i = \Phi(x_i'\alpha)$. But, $x_i'\alpha - [1 - \Phi(x_i'\alpha)]^{-1}\phi(x_i'\alpha) < 0$ as shown in Amemiya [1973, p. 1007]. Therefore, the right-hand side of (2.46) is the sum of two negative-definite matrices and hence is negative definite.

Even though convergence is assured by global concavity, it is a good idea to start an iteration with a good estimator because it will improve the speed of convergence. Tobin [1958] used the following simple estimator based on a linear approximation of the reciprocal of Mills' ratio to start his iteration for obtaining the MLE: By equating the right-hand side of (2.43) to zero, we obtain

$$(2.47) \qquad -\sigma\sum_0 \frac{\phi_i}{1-\Phi_i}x_i + \sum_1 (y_i - x_i'\beta)x_i = 0 \ .$$

If we premultiply (2.47) by $\beta'/(2\sigma^4)$ and add it to the equation obtained by setting (2.44) equal to zero, we get

$$\sigma^2 = n_1^{-1} \sum_1 (y_i - x_i'\beta)y_i \quad . \tag{2.48}$$

Approximate $(1 - \Phi_i)^{-1}\phi_i$ by the linear function $a + b \cdot (x_i'\beta/\sigma)$ and substitute it into the left-hand side of (2.47) to obtain

$$-\sigma \sum_0 [a + b \cdot (x_i'\beta/\sigma)]x_i + \sum_1 (y_i - x_i'\beta)x_i = 0 \quad . \tag{2.49}$$

Solve (2.49) for β and insert it into (2.48) to obtain a quadratic equation in σ. If the roots are imaginary, Tobin's method does not work. If the roots are real, one of them can be chosen arbitrarily. Once an estimate of σ is determined, an estimate of β can be determined linearly from (2.49). Amemiya [1973] showed that Tobin's initial estimator is inconsistent. However, empirical researchers have found it to be a good starting value for iteration.

Amemiya [1973] proposed the following simple consistent estimator: We have

$$E(y_i^2|y_i > 0) = (x_i'\beta)^2 + \sigma x_i'\beta\lambda(x_i'\alpha) + \sigma^2 \quad . \tag{2.50}$$

Combining (2.12) and (2.50) yields

$$E(y_i^2|y_i > 0) = x_i'\beta\, E(y_i|y_i > 0) + \sigma^2 \quad , \tag{2.51}$$

which can be alternatively written as

$$y_i^2 = y_i x_i'\beta + \sigma^2 + \zeta_i \quad , \quad \text{for } i \text{ such that } y_i > 0 \quad , \tag{2.52}$$

where $E(\zeta_i|y_i > 0) = 0$. Then, consistent estimates of β and σ^2 are obtained by applying an instrumental variables method to (2.52) using $(\hat{y}_i x_i', 1)$ as the instrumental variables where $\hat{y}_i$ is the predictor of y_i obtained by regressing positive y_i on x_i and, perhaps, powers of x_i. The asymptotic distribution of the estimator is given in Amemiya [1973].

A consistent initial estimator is useful not only because it improves the speed of convergence but also because the second round estimate obtained either by the Newton-Raphson or the method of scoring iteration starting at a regular consistent estimator has the same asymptotic distribution as the MLE, as shown by Amemiya [1973]. Unfortunately, however, simulation studies such as the one by Wales and Woodland [1980] have shown this particular consistent estimator to be rather inefficient.

6. The EM Algorithm: The EM algorithm is a general iterative method for obtaining the MLE, first proposed by Hartley [1958] and generalized by Dempster, Laird, and Rubin [1977], that is especially suited for censored regression models such as Tobit models. It has good convergence properties making it especially useful for handling the more complex Tobit models, which I will discuss later, where global concavity may not hold. However, I will discuss it in the context of the Standard Tobit model because all the essential features of the algorithm can be explained for that model. I will first present the definition and the properties of the EM algorithm under a general setting and then apply it to the Standard Tobit model.

I will explain the EM algorithm in a general model where a vector of observable variables z are related to a vector of unobservable variables y^* in such a way that the value of y^* uniquely determines the value of z but not vice versa. In the Tobit model, $\{y_i^*\}$ defined in (2.3) constitute the elements of y^*, and $\{y_i\}$ and $\{w_i\}$ defined in (2.4) and (2.5) respectively constitute the elements of z. Let the joint density or probability of y^* be $f(y^*)$ and let the joint density or probability of z be $g(z)$. Also, define $k(y^*|z) = f(y^*)/g(z)$. We implicitly assume that f, g, and k depend on a vector of parameters θ. The purpose is to maximize

(2.53) $$L(\theta) \equiv n^{-1} \log g(z) = n^{-1} \log f(y^*) - n^{-1} \log k(y^*|z)$$

with respect to θ. Define

(2.54) $$Q(\theta|\theta_1) = E[n^{-1} \log f(y^*|\theta)|z, \theta_1] ,$$

where we are taking expectation assuming θ_1 is the true parameter value, and doing this conditional on z. Then, the EM algorithm purports to maximize $L(\theta)$ by maximizing $Q(\theta|\theta_1)$ with respect to θ when θ_1 is given at each step of the iteration. The "E" of the name "EM" refers to the expectation taken in (2.54) and the "M" refers to the maximization of (2.54).

I will consider the convergence properties. Define

(2.55) $$H(\theta|\theta_1) = E[n^{-1} \log k(y^*|z, \theta)|z, \theta_1] .$$

Then we have from (2.53), (2.54), and (2.55) and the fact that $L(\theta|\theta_1) = L(\theta)$

(2.56) $$L(\theta) = Q(\theta|\theta_1) - H(\theta|\theta_1) .$$

But we have by Jensen's inequality[13/]

(2.57) $$H(\theta|\theta_1) \leqq H(\theta_1|\theta_1) .$$

Now, given θ_1, let $M(\theta_1)$ maximize $Q(\theta|\theta_1)$ with respect to θ. Then, we have

(2.58) $$L(M) = Q(M|\theta_1) - H(M|\theta_1) .$$

But, since $Q(M|\theta_1) \geqq Q(\theta|\theta_1)$ by definition and $H(M|\theta_1) \leqq H(\theta_1|\theta_1)$ by (2.57), we have from (2.56) and (2.58)

(2.59) $$L(M) \geqq L(\theta_1) .$$

Thus, we have proved the desirable property that L always increases or stays constant at each step of the EM algorithm. Next, let $\hat{\theta}$ be the MLE. Then, $L(\hat{\theta}) \geqq L[M(\hat{\theta})]$ by definition. But $L(\hat{\theta}) \leqq L[M(\hat{\theta})]$ by (2.59). Therefore we have

(2.60) $$L(\hat{\theta}) = L[M(\hat{\theta})] ,$$

which implies that if $L(\theta)$ has a unique maximum and if the EM algorithm converges, it converges to $\hat{\theta}$.

We still need to prove that the EM algorithm converges to the MLE. Unfortunately, it is never easy to find reasonable and easily verifiable conditions for the convergence of any iterative algorithm. Dempster, et. al. do not succeed in this attempt. I will merely give a sufficient set of conditions below.

The conditions I impose are (A) L is bounded and (B) the smallest characteristic root of $-\partial^2 Q(\theta|\theta_1)/\partial\theta\,\partial\theta'$ is bounded away from 0 for all θ_1 and θ. Consider

(2.61) $$L(\theta_r) = Q(\theta_r|\theta_r) - H(\theta_r|\theta_r)$$

and

(2.62) $$L(\theta_{r+1}) = Q(\theta_{r+1}|\theta_r) - H(\theta_{r+1}|\theta_r) \quad .$$

Since we previously established $L(\theta_{r+1}) \geqq L(\theta_r)$, assumption (A) implies $\lim_{r\to\infty} [L(\theta_{r+1}) - L(\theta_r)] = 0$. Therefore, from (2.61) and (2.62) and using (2.57) and $Q(\theta_{r+1}|\theta_r) \geqq Q(\theta_r|\theta_r)$ by definition we have

(2.63) $$\lim_{r\to\infty} [Q(\theta_{r+1}|\theta_r) - Q(\theta_r|\theta_r)] = 0 \quad .$$

Now, denoting only the first argument of Q and suppressing its second argument, we have by a Taylor expansion of $Q(\theta_r)$ about $Q(\theta_{r+1})$

(2.64) $$\begin{aligned} Q(\theta_{r+1}) - Q(\theta_r) &= \frac{1}{2}(\theta_r - \theta_{r+1})'[-\frac{\partial^2 Q}{\partial\theta\partial\theta'}](\theta_r - \theta_{r+1}) \\ &\geqq \frac{1}{2}\lambda_s \cdot (\theta_r - \theta_{r+1})'(\theta_r - \theta_{r+1}) \quad , \end{aligned}$$

where the matrix of the second derivatives is evaluated at a point between θ_r and θ_{r+1} and λ_s denotes its smallest characteristic root. Note that in obtaining the equality above I have noted $\partial Q(\theta_{r+1}|\theta_r)/\partial\theta_{r+1} = 0$ by definition. Thus, (2.63), (2.64), and assumption (B) imply

(2.65) $$\lim_{r\to\infty} (\theta_{r+1} - \theta_r) = 0 \quad ,$$

meaning that the EM algorithm converges.

Now, consider an application of the algorithm to the Tobit model.[14] Define $\theta = (\beta', \sigma^2)'$. Then, in the Tobit model we have

(2.66) $$\log f(y^*|\theta) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^* - x_i'\beta)^2 \quad ,$$

and, for a given estimate $\theta_1 = (\beta_1', \sigma_1^2)'$, the EM algorithm maximizes with respect to β and σ^2

(2.67) $$\begin{aligned} E[\log f(y^*|\theta)|y, w, \theta_1] &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_1 (y_i - x_i'\beta)^2 \\ &\quad - \frac{1}{2\sigma^2}\sum_0 E[(y_i^* - x_i'\beta)^2|w_i = 0, \theta_1] \\ &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_1 (y_i - x_i'\beta)^2 \\ &\quad - \frac{1}{2\sigma^2}\sum_0 [E(y_i^*|w_i = 0, \theta_1) - x_i'\beta]^2 \\ &\quad - \frac{1}{2\sigma^2}\sum_0 V(y_i^*|w_i = 0, \theta_1) \quad , \end{aligned}$$

where

(2.68) $$E(y_i^*|w_i = 0, \theta_1) = x_i'\beta_1 - \frac{\sigma_1\phi_1}{1-\Phi_1} \equiv y_i^0$$

and

(2.69) $$V(y_i^*|w_i = 0, \theta_1) = \sigma_1^2 + x_i'\beta_1 \frac{\sigma_1\phi_1}{1-\Phi_1} - \left[\frac{\sigma_1\phi_1}{1-\Phi_1}\right]^2$$

where $\phi_1 = \phi(x_i'\beta_1/\sigma_1)$ and $\Phi_1 = \Phi(x_i'\beta_1/\sigma_1)$.

From (2.67) it is clear that the second-round estimate of β in the EM algorithm, denoted β_2, is obtained as follows: Assume without loss of generality that the first n_1 observations of y_i are positive and call the vector of those observations y as I did in (2.20). Next, define an $(n - n_1)$-vector y^* whose elements are the y_i^0 defined in (2.68). Then, we have

(2.70) $$\beta_2 = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' \begin{bmatrix} y \\ y^0 \end{bmatrix} ,$$

where X was defined after (2.4). In other words, the EM algorithm amounts to predicting all the unobservable values of y_i^* by their conditional expectations and treating the predicted values as if they were the observed values. The second-round estimate of σ^2, denoted σ_2^2, is given by

(2.71) $$\sigma_2^2 = n^{-1}\Big[\sum_1 (y_i - x_i'\beta_2)^2 + \sum_0 (y_i^* - x_i'\beta_2)^2 + \sum_0 V(y_i^*|w_i = 0, \theta_1)\Big] .$$

Although this follows from the general theory of the algorithm given earlier, we can also directly show that the MLE $\hat\theta$ is the equilibrium solution of the iteration defined by (2.70) and (2.71). Partition $\underset{\sim}{X} = (X', X^{0\prime})'$ so that X is multiplied by y and X^0 by y^0. Then, inserting $\hat\theta$ into both sides of (2.70) yields, after collecting terms

(2.72) $$X'X\hat\beta = X'y - X^{0\prime}\left[\frac{\hat\sigma\phi(x_i'\hat\beta/\hat\sigma)}{1 - \Phi(x_i'\hat\beta/\hat\sigma)}\right] ,$$

where the last bracket denotes an $(n - n_1)$-dimensional vector whose typical element is given inside. But, clearly, (2.72) is equivalent to (2.47). Similarly, the MLE $\hat\theta$ can be shown to be an equilibrium solution of (2.71).

Unfortunately, conditions (A) and (B) do not generally hold for the Tobit model. However, they do hold if the sample size is sufficiently large and if the iteration is started from a point sufficiently close to the MLE. Schmee and Hahn [1979] performed a simulation study of the EM algorithm applied to a censored regression model (a survival model) defined by

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* \le c \\ c & \text{if } y_i^* > c \end{cases} ,$$

where $y_i^* \sim N(\alpha + \beta x_i, \sigma^2)$. They obtained good convergence.

D. Properties of the Tobit MLE Under Nonstandard Assumptions

In this section I will discuss the properties of the Tobit MLE --the estimator which maximizes (2.42)--under various types of nonstandard assumptions: heteroscedasticity, serial correlation, and nonnormality. It will be shown that the Tobit MLE remains consistent under serial correlation but not under heteroscedasticity or nonnormality. This result contrasts with the classical regression model in which the least squares estimator (the MLE under the normality assumption) is generally consistent under all of the three types of nonstandard assumptions mentioned above.

Before proceeding with rigorous argument, I will give an intuitive explanation of the above-mentioned result. By considering (2.17) we see that serial correlation of y_i should not affect the consistency of the NLLS estimator, whereas heteroscedasticity changes σ to σ_i and hence invalidates the estimation of the equation by least squares. If y_i^* is not normal, equation (2.17) itself is generally invalid, which leads to the inconsistency of the NLLS estimator. Though the NLLS estimator is different from the ML estimator, one can expect a certain correspondence between the consistency properties of the two estimators.

It should be noted that the MLE derived under certain assumptions generally loses it desirable properties of consistency and asymptotic efficiency when one or more of the assumptions are removed. This result can be explained as follows: The consistency of the MLE is essentially equivalent to the condition $E\,\partial \log L/\partial\theta = 0$. The equality follows from $E\,\partial \log L/\partial\theta = E\,L^{-1}\cdot(\partial L/\partial\theta) = \int L^{-1}\cdot(\partial L/\partial\theta)\,L\,dy = \int(\partial L/\partial\theta)dy = \partial\int L\,dy/\partial\theta = 0$, if the expectation is taken using the same L as that

which is maximized. If the expectation is taken using a different L, say L_1, the second equality above generally does not hold, leading to the inconsistency of the MLE. Thus, it is best to remember that the classical normal regression model is an exceptional case.

1. Heteroscedasticity: Hurd [1979] evaluated the probability limit of the truncated Tobit MLE when a certain type of heteroscedasticity is present in two simple truncated Tobit models: (1) the i.i.d. case (that is, the case of the regressor consisting only of a constant term) and (2) the case of a constant term plus one independent variable. Recall that the truncated Tobit model is the one in which no information is available for those observation for which $y_i^* < 0$ and therefore the MLE maximizes (2.7) rather than (2.6).

In the i.i.d. case, Hurd created heteroscedasticity by generating rn observations from $N(\mu, \sigma_1^2)$ and $(1 - r)n$ observations from $N(\mu, \sigma_2^2)$. In each case, he recorded only positive observations. Let y_i, $i = 1,2,\ldots,n_1$, be the recorded observations. (Note $n_1 \leq n$.) One can show that the truncated Tobit MLE of μ and σ^2, denoted $\hat{\mu}$ and $\hat{\sigma}^2$, are defined by equating the first two population moments of y_i to their respective sample moments:

$$\hat{\mu} + \hat{\sigma}\lambda(\hat{\mu}/\hat{\sigma}) = n_1^{-1}\sum_{i=1}^{n_1} y_i \tag{2.73}$$

and

$$\hat{\mu}^2 + \hat{\sigma}\hat{\mu}\lambda(\hat{\mu}/\hat{\sigma}) + \hat{\sigma}^2 = n_1^{-1}\sum_{i=1}^{n_1} y_i^2 \ . \tag{2.74}$$

Taking the probability limit of both sides of (2.73) and (2.74) and expressing $\text{plim}\ n_1^{-1}\sum y_i$ and $\text{plim}\ n_1^{-1}\sum y_i^2$ as certain functions of the parameters μ, σ_1^2, σ_2^2, and r, one can define $\text{plim}\ \hat{\mu}$ and $\text{plim}\ \hat{\sigma}^2$ implicitly as functions of these parameters. Hurd evaluated the probability limits for various values of μ and σ_1 after having fixed $r = 0.5$ and $\sigma_2 = 1$. The worst result occured when $\mu = -1$ and $\sigma_1 = 0.5$, leading to $\text{plim}\ \hat{\mu} = -121.02$!

In the case of one independent variable, Hurd generated observations from $N(\alpha + \beta x_i, \sigma_i^2)$ after having generated x_i and $\log|\sigma_i|$ from Bivariate $N(0, 0, V_1^2, V_2^2, \rho)$. For given values of α, β, V_1, V_2, and ρ, Hurd found the values of α, β, and σ^2 that maximize $E \log L$, where L is as given in (2.7). Those values are the probability limits of the MLE of α, β, and σ^2 under Hurd's model if the expectation of $\log L$ is taken using the same model. Again, Hurd found extremely large asymptotic biases in certain cases.

Hurd's results indicate that one should treat Tobit ML estimates cautiously if one suspects heteroscedasticity. In such a case, one should perhaps use Powell's least absolute deviations estimator [1981] (to be discussed in subsection 5 below), which remains consistent under general heteroscedastic as well as nonnormal distributions.

2. Serial Correlation: Robinson [1982] proved the strong consistency and the asymptotic normality of the Tobit MLE under very general assumptions about u_i (normality is presupposed) and obtained its asymptotic variance-covariance matrix. His assumptions are slightly

stronger than the stationarity assumption but are weaker than the assumption that u_i possesses a continuous spectral density. His results are especially useful since the full MLE that takes account of even a simple type of serial correlation seems computationally intractable.

3. Nonnormality: Goldberger [1980] considered an i.i.d. truncated sample model in which data are generated by a certain nonnormal distribution with mean μ and variance 1 and are recorded only when the value is smaller than a constant c. Let y represent the recorded random variable and let $\bar{y}$ be the sample mean. The researcher is to estimate μ by the MLE assuming that the data are generated by $N(\mu,1)$. As in Hurd's i.i.d. model, the MLE $\hat{\mu}$ is defined by equating the population mean of y to its sample mean:

$$\hat{\mu} - \lambda(c - \hat{\mu}) = \bar{y} \quad . \tag{2.75}$$

Taking the probability limit of both sides of (2.75) under the true model and putting $\text{plim}\ \hat{\mu} = \mu^*$ yields

$$\mu^* - \lambda(c - \mu^*) = \mu - h(c - \mu) \quad , \tag{2.76}$$

where $h(c - \mu) = E(\mu - y|y < c)$, the expectation being taken using the true model. Defining $m = \mu^* - \mu$ and $\theta = c - \mu$, we rewrite (2.76) as

$$m = \lambda(\theta - m) - h(\theta) \quad . \tag{2.77}$$

Goldberger calculated m as a function of θ when the data are generated by Student's t with various degrees of freedom, Laplace, and logistic distributions. The asymptotic bias was found to be especially great when the true distribution is Laplace. Goldberger also extended the analysis to the regression model with a constant term and one discrete independent variable. Arabmazar and Schmidt [1981] extended Goldberger's analysis to the case of an unknown variance and found that the asymptotic bias was further accentuated.

4. Tests for Normality: The fact that the Tobit MLE is generally inconsistent when the true distribution is nonnormal makes it important for a researcher to test whether his data are generated by a normal distribution. Nelson [1981] devised tests for normality in the i.i.d. censored sample model and the Tobit model. His tests are applications of the specification test of Hausman [1978].

In Hausman's test, one uses the MLE $\hat{\theta}$ obtained under the null hypothesis, which is asymptotically efficient under the null hypothesis but loses consistency under an alternative hypothesis, and a consistent estimator $\tilde{\theta}$, which is asymptotically less efficient than the MLE under the null hypothesis but remains consistent under an alternative hypothesis. Hausman [1978] noted that $(\hat{\theta} - \tilde{\theta})'V^{-1}(\hat{\theta} - \tilde{\theta})$ is asymptotically distributed under the null hypothesis as chi-square with K degrees of freedom (K being the number of elements in θ), where $V = V(\tilde{\theta}) - V(\hat{\theta})$, the difference of the asymptotic variance-covariance matrices evaluated under the null hypothesis. An advantage of Hausman's test is that one need not know the covariance between $\hat{\theta}$ and $\tilde{\theta}$ to perform the test.

Nelson's i.i.d. censored sample model is defined by

$$y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* > 0 \\ 0 & \text{if} \quad y_i^* \leq 0 \end{cases}, \quad i = 1,2,\ldots,n ,$$

where $y_i^* \sim N(\mu, \sigma^2)$ under the null hypothesis. Nelson considers the estimation of $P(y_i^* > 0)$. Its MLE is $\Phi(\hat{\mu}/\hat{\sigma})$ where $\hat{\mu}$ and $\hat{\sigma}$ are the MLE of the respective parameters. A consistent estimator is provided by n_1/n where, as before, n_1 is the number of positive observations of y_i. Clearly, n_1/n is a consistent estimator of $P(y_i^* > 0)$ under any distribution provided that it is i.i.d. Nelson derived the asymptotic variances under normality of the two estimators.

If we interpret what one is estimating by the two estimators as $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} P(y_i^* > 0)$, Nelson's test can be interpreted as a test of the null hypothesis against a more general misspecification than just nonnormality. In fact, Nelson conducted a simulation study to evaluate the power of the test against a heteroscedastic alternative. The performance of the test was satisfactory but not especially encouraging.

In the Tobit model, Nelson considers the estimation of

$$n^{-1} E \underset{\sim}{X}'\underset{\sim}{y} = n^{-1} \sum_{i=1}^{n} x_i[\Phi(x_i'\alpha)x_i'\beta + \sigma\phi(x_i'\alpha)].$$ Its MLE is given by the right-hand side of this equation evaluated at the Tobit MLE, and its consistent estimator is provided by $n^{-1}\underset{\sim}{X}'\underset{\sim}{y}$. Hausman's test based on these two estimators will work because this consistent estimator is consistent under general distributional assumptions on $\underset{\sim}{y}$. Nelson

derived the asymptotic variance-covariance matrices of the two estimators.

Nelson was quite ingenious in that he considered certain functions of the original parameters for which one can easily obtain estimators which are consistent under very general assumptions. However, it would be better if one could find a general consistent estimator for the original parameters themselves. Therefore, I would suggest a Hausman's test using the Tobit MLE and Powell's least absolute deviations estimator of β as an alternative to Nelson's suggestion. The test will be computationally more burdensome than Nelson's test but seems theoretically preferable.

As still another alternative, Ruud [1982] suggests contrasting the Tobit MLE with the probit MLE for Hausman's test. He argues that though the probit MLE is not consistent under either nonnormality or heteroscedasticity, Hausman's test works as long as the discrepancy between the two estimators is more pronounced under an alternative hypothesis than under the null hypothesis.

5. Nonnormal Tobit: If u_i in the Tobit model (2.3) is not normal, one of two things can be done: (1) Specify a nonnormal distribution and use the true MLE or some other estimator tailormade for the distribution. (2) Use an estimator which is consistent under general distributions, both normal and nonnormal. I will mention an example for each of the two approaches.

Amemiya and Boskin [1974] studied the effect of wage and other independent variables on the number of months during a five-year period in which a household received welfare payments. Since the dependent

variable is naturally bounded between 0 and 60, one must impose both an upper and lower truncation point if one uses a normal Tobit model. Instead, the authors assumed the dependent variable to be lognormal and hence positive, so that only an upper truncation needs to be imposed. The MLE was used.

The majority of models I will discuss in Section III assume a normal distribution. Exceptions are some of the models proposed by Cragg [1971] discussed in Section III.B.5 and the model of Dubin and McFadden [1980] discussed in Section III.E.7.

Powelll [1981] proposed the least absolute deviations (LAD) estimator for censored and truncated regression models, proved its consistency under general unimodal symmetric distributions, and derived its asymptotic distribution. As I mentioned above, the estimator is also consistent under heteroscedastic errors. The intuitive appeal for the LAD estimator in a censored regression model arises from the simple fact that in the i.i.d. sample case, the median (of which the LAD estimator is a generalization) is not affected by censoring (more strictly, left censoring below the mean), whereas the mean is. In a censored regression model, the LAD estimator is defined as that which minimizes $\sum_{i=1}^{n} |y_i - \max(0, x_i'\beta)|$. The motivation for the LAD estimator in a truncated regression model is less obvious. Powell defines the LAD estimator in the truncated case as that which minimizes $\sum_{i=1}^{n} |y_i - \max(2^{-1}y_i, x_i'\beta)|$.

E. Minor Variations of the Standard Tobit Model

In this section I discuss a few models that are minor variations on the Tobit model. More significant generalizations of the Tobit model are discussed in Section III.

Rosett [1959] proposed a model in which the observable random variables $\{y_i\}$ are defined by

$$(2.78) \qquad y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* \leqq 0 \\ 0 & \text{if} \quad 0 < y_i^* < \alpha \\ y_i^* - \alpha & \text{if} \quad \alpha \leqq y_i^* \end{cases}, \quad i = 1,2,\ldots,n \ ,$$

where $y_i^* \sim N(x_i'\beta, \sigma^2)$. One can estimate α as well as β and σ^2. Rosett called it a model of friction because the model implies that the dependent variable assumes a certain value (in this case 0) until a change in an independent variable overcomes the friction. At this point the dependent variable either increases or decreases depending upon the type of the stimulus. Maddala [1977] remarks that this model is useful in analyzing dividend policies, changes in wage offers by firms, and similar examples where firms respond by jumps after a certain cumulative effort.

Rosett and Nelson [1975] considered the following simple generalization of the Tobit model:

$$(2.79) \qquad y_i = \begin{cases} \alpha_1 & \text{if} \quad y_i^* \leqq \alpha_1 \\ y_i^* & \text{if} \quad \alpha_1 < y_i^* < \alpha_2 \\ \alpha_2 & \text{if} \quad \alpha_2 \leqq y_i^* \end{cases},$$

where $y_i^* \sim N(x_i'\beta, \sigma^2)$. If x_i contains a constant term, one can assume $\alpha_1 = 0$ without loss of generality. Then, the Standard Tobit model is obtained as a special case by putting $\alpha_2 = \infty$. According to Maddala [1977a], an example of a problem to which this model has been applied is the demand for health insurance by people on medicare, where both a minimum coverage and a maximum amount are imposed.

Dagenais [1969] proposed a model which is obtained by making the boundary points of Rosett's model stochastic as follows:

$$(2.80) \qquad y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* \leqq v_i \\ 0 & \text{if} \quad v_i < y_i^* < x_i'\gamma + w_i \\ y_i^* + x_i'\gamma & \text{if} \quad x_i'\gamma + w_i \leqq y_i^* \end{cases},$$

where $y_i^* \sim N(x_i'\beta, \sigma^2)$ and v_i and w_i are also normal. Unfortunately, there is a logical inconsistency in the model because $v_i < x_i'\gamma + w_i$ cannot always be guaranteed. Perhaps for this reason, this model does not seem to have been applied to real data. Dagenais [1975] begins to discuss this model but the model he actually estimated is of Type 2 Tobit, which I will discuss later.

III. Generalizations (Type 2 through Type 5 Tobit)

A. Introduction

As I stated in Section I, I will classify the majority of Tobit models into five common types according to similarities in the likelihood function. Type 1 is the Standard Tobit model which I have discussed in Section II. In Section III I will define and discuss the remaining four types of Tobit models.

It is useful to characterize the likelihood function of each type of model schematically as follows:

Table 3.1

Type 1	$P(y_1 < 0) \cdot P(y_1)$
2	$P(y_1 < 0) \cdot P(y_1 > 0, y_2)$
3	$P(y_1 < 0) \cdot P(y_1, y_2)$
4	$P(y_1 < 0, y_3) \cdot P(y_1, y_2)$
5	$P(y_1 < 0, y_3) \cdot P(y_1 > 0, y_2)$.

In the above, each y_j, $j = 1,2$, and 3, is assumed to be distributed as $N(x_j'\beta_j, \sigma_j^2)$, and P denotes a probability or a density or a combination thereof. One is to take the product of each P over the observations that belong to a particular category determined by the sign of y_1. Thus, in Type 1 (Standard Tobit model), $P(y_1 < 0) \cdot P(y_1)$ is an abbreviated notation for $\prod_0 P(y_{1i} < 0) \cdot \prod_1 f_{1i}(y_{1i})$, where f_{1i} is the density of $N(x_{1i}'\beta_1, \sigma_1^2)$. This expression can be rewritten as (2.6) after dropping the unnecessary subscript 1.

Another way to characterize the five types is by the following classification of the three dependent variables which appear in the models:

Table 3.2

	y_1	y_2	y_3
Type 1	C		
2	B	C	
3	C	C	
4	C	C	C
5	B	C	C

In Table 2 above, B is an abbreviation for Binary and C for Censored. In each type of model, the sign of y_1 determines one of the two possible categories for the observations, and a censored variable is observed in one category and unobserved in the other. Note that when y_1 is labelled C, it plays two roles: the role of the variable whose sign determines categories and the role of a censored variable.

We allow for the possibility that there are constraints among the parameters of the model (β_j, σ_j^2), $j = 1,2$, or 3. For example, constraints will occur if the original model is specified as a simultaneous equations model in terms of y_1, y_2, and y_3. For, then, the β's denote the reduced-form parameters.

I will not discuss here models in which there are more than one binary variable and, hence, models whose likelihood function consists of more than two components. Such models are computationally more burdensome because they involve double or higher-order integration of joint normal densities. The only exception occurs in Section III.E.7, which includes models that are obvious generalizations of the Type 5 Tobit model. One notable (at least in my mind) model of the sort I do not discuss is a simultaneous-equation Tobit model of Amemiya [1974b]. The simplest two-equation case of this model is defined by $y_{1i} = \text{MAX}\,(\gamma_1 y_{2i} + x_{1i}'\beta_1 + u_{1i},\ 0)$ and $y_{2i} = \text{MAX}\,(\gamma_2 y_{1i} + x_{2i}'\beta_2 + u_{2i},\ 0)$, where (u_{1i}, u_{2i}) is bivariate normal and $\gamma_1\gamma_2 < 1$ must be assumed for the model to be logically consistent. A schematic representation of the likelihood function of this two equation model is $P(y_1, y_2) \cdot P(y_1 < 0, y_3) \cdot P(y_2 < 0, y_4) \cdot P(y_3 < 0, y_4 < 0)$.

B. Type 2: $P(y_1 < 0) \cdot P(y_1 > 0, y_2)$

1. Definition and Estimation: The Type 2 Tobit model is defined as follows:

$$\text{(3.1)}\qquad \begin{cases} y_{1i}^* = x_{1i}'\beta_1 + u_{1i} \\ y_{2i}^* = x_{2i}'\beta_2 + u_{2i} \\ y_{2i} = \begin{cases} y_{2i}^* & \text{if } y_{1i}^* > 0 \\ 0 & \text{if } y_{1i}^* \leqq 0 \end{cases}, \quad i = 1,2,\dots,n, \end{cases}$$

where $\{u_{1i}, u_{2i}\}$ are i.i.d. drawings from a bivariate normal distribution with mean zero, variances σ_1^2 and σ_2^2, and covariance σ_{12}. It is assumed that only the sign of y_{1i}^* is observed and that y_{2i}^* is observed only when $y_{1i}^* > 0$. It is assumed that x_{1i} are observed for all i but x_{2i} need not be observed for i such that $y_{1i}^* \leqq 0$. One may also define, as in (2.5),

$$\text{(3.2)}\qquad w_{1i} = \begin{cases} 1 & \text{if } y_{1i}^* > 0 \\ 0 & \text{if } y_{1i}^* \leqq 0 . \end{cases}$$

Then, $\{w_{1i}, y_{2i}\}$ constitute the observed sample of the model. It should be noted that, unlike the Type 1 Tobit, y_{2i} may take negative values. See the discussion of Cragg [1971] in Section B.5 below for models that prevent this. As in (2.4), $y_{2i} = 0$ merely signifies the event $y_{1i}^* \leqq 0$.

The likelihood function of the model is given by

$$\text{(3.3)}\qquad L = \prod_0 P(y_{1i}^* \leqq 0) \prod_1 f(y_{2i} \mid y_{1i}^* > 0)\, P(y_{1i}^* > 0) ,$$

where $\prod_0$ and $\prod_1$ stand for the product over those i for which $y_{2i} = 0$ and $y_{2i} > 0$ respectively and $f(\cdot \mid y_{1i}^* > 0)$ stands for the conditional density of y_{2i}^* given $y_{1i}^* > 0$. Note the similarity between (2.8) and (3.3). As in Type 1 Tobit, one can obtain a consistent estimate of β_1/σ_1 by maximizing the probit part of (3.3),

$$\text{(3.4)}\qquad \text{Probit } L = \prod_0 P(y_{1i}^* \leqq 0) \prod_1 P(y_{1i}^* > 0) .$$

Also, (3.4) is a part of the likelihood function for every one of the five types of models; therefore, a consistent estimate of β_1/σ_1 can be obtained by the probit MLE in each of these types of model.

One can rewrite (3.3) as

$$L = \prod_0 P(y_{1i}^* \leq 0) \prod_1 \int_0^\infty f(y_{1i}^*, y_{2i}) dy_{1i}^* \quad , \tag{3.5}$$

where $f(\cdot,\cdot)$ denotes the joint density of y_{1i}^* and y_{2i}^*. One can write the joint density as the product of a conditional density and a marginal density, i.e. $f(y_{1i}^*, y_{2i}) = f(y_{1i}^*|y_{2i})f(y_{2i})$, and determine a specific form of $f(y_{1i}^*|y_{2i})$ from the well-known fact that the conditional distribution of y_{1i}^* given $y_{2i}^* = y_{2i}$ is normal with mean $x_{1i}'\beta_1 + \sigma_{12}\sigma_2^{-2}(y_{2i} - x_{2i}'\beta_2)$ and variance $\sigma_1^2 - \sigma_{12}^2\sigma_2^{-2}$. Thus, one can further rewrite (3.5) as

$$\begin{aligned} L = & \prod_0 [1 - \Phi(x_{1i}'\beta_1\sigma_1^{-1})] \\ & \cdot \prod_1 \Phi\{[x_{1i}'\beta_1\sigma_1^{-1} + \sigma_{12}\sigma_1^{-1}\sigma_2^{-2}(y_{2i} - x_{2i}'\beta_2)] \\ & [1 - \sigma_{12}^2\,\sigma_1^{-2}\,\sigma_2^{-2}]^{-\frac{1}{2}}\}\ \sigma_2^{-1}\phi(y_{2i} - x_{2i}'\beta_2) \quad . \end{aligned} \tag{3.6}$$

Note that L depends on σ_1 only through $\beta_1\sigma_1^{-1}$ and $\sigma_{12}\sigma_1^{-1}$; therefore, if there is no constraint on the parameters, one can put $\sigma_1 = 1$ without any loss of generality. Then, the remaining parameters can be identified. If, however, there is at least one common element in β_1 and β_2, σ_1 can be also identified.

I will show how Heckman's two-step estimator can be used in this model. To obtain an equation comparable to (2.17), we need to evaluate $E(y_{2i}^*|y_{1i}^* > 0)$. For this purpose we use the well-known result

$$y_{2i}^* = x_{2i}'\beta_2 + \sigma_{12}\sigma_1^{-2}(y_{1i}^* - x_{1i}'\beta_1) + \zeta_{2i} \quad , \tag{3.7}$$

where ζ_{2i} is normally distributed independently of y_{1i}^* with mean zero and variance $\sigma_2^2 - \sigma_{12}^2\sigma_1^{-2}$. Using (3.7), one can express $E(y_{2i}^*|y_{1i}^* > 0)$ as a simple linear function of $E(y_{1i}^*|y_{1i}^* > 0)$, which was already obtained in Section II. Using (3.7), one can also derive $V(y_{2i}^*|y_{1i}^* > 0)$ easily. Thus, we obtain

$$y_{2i} = x_{2i}'\beta_2 + \sigma_{12}\sigma_1^{-1}\lambda(x_{1i}'\alpha_1) + \epsilon_{2i} \quad , \quad \text{for } i \text{ such that } y_{2i} > 0 \quad , \tag{3.8}$$

where $\alpha_1 = \beta_1\sigma_1^{-1}$, $E\,\epsilon_{2i} = 0$, and

$$V\,\epsilon_{2i} = \sigma_2^2 - \sigma_{12}^2\sigma_1^{-2}[x_{1i}'\alpha_1\lambda(x_{1i}'\alpha_1) + \lambda(x_{1i}'\alpha_1)^2] \quad . \tag{3.9}$$

As in the case of the Type 1 Tobit, Heckman's two-step estimator is the LS estimator applied to (3.8) after replacing α_1 with the probit MLE. The asymptotic distribution of the estimator can be similarly obtained as in Section II.C.3 by defining η_{2i} in the same way as before. It was first derived by Heckman [1979].

The Standard Tobit (Type 1) is a special case of Type 2, in which $y_{1i}^* = y_{2i}^*$. Therefore, (3.8) and (3.9) will be reduced to (2.17) and (2.18) by putting $x_{1i}'\beta_1 = x_{2i}'\beta_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma_{12}$.

A generalization of the two-step method applied to (2.29) can be easily defined for this model but will not be discussed.

2. A Special Case of Independence: Dudley and Montmarquette [1976] analyzed whether or not the United States gives foreign aid to a particular country and, if it does, how much foreign aid it gives using a special case of the model (3.1) where the independence of u_{1i} and u_{2i} is assumed. In their model, the sign of y^*_{1i} determines whether aid is given to the i-th country, and y^*_{2i} determines the actual amount of aid. They used the probit MLE to estimate β_1 (assuming $\sigma_1 = 1$) and the least squares regression of y_{2i} on x_{2i} to estimate β_2. The LS estimator of β_2 is consistent in their model because of the assumed independence between u_{1i} and u_{2i}. This is the main advantage of their model. However, it is unrealistic to assume that the actual amount of aid, y^*_2, is independent of the variable which determines whether or not aid is given, y^*_1. This model is the opposite extreme of the Tobit model, which can be regarded as a special case of Type 2 model where there is total dependence between y^*_1 and y^*_2, in the whole spectrume of models (with varying correlation between y^*_1 and y^*_2) contained in Type 2.

Because of the computational advantage mentioned above, this "independence" model and its variations were frequently used in econometric applications in 1960's and early 70's. In many of these studies, authors made the additional linear probability assumption: $P(y^*_{1i} > 0) = x'_{1i}\beta_1$, which enabled them to estimate β_1 (as well as β_2) consistently by the least squares method. For examples of these studies, see Huang [1964] and Wu [1965].

3. Gronau [1973]: I take up Gronau's model as the first example of the Type 2 Tobit model because he seems to be the first person to suggest an empirical model of this type, even though he did not use all the information provided by the model and sometimes used incorrect estimation procedures, as I will show below. His model of labor supply, based on the idea of a reservation wage, has since been used and extended by many authors.

First, I will breifly sketch Gronau's theory of how a housewife decides whether or not to work and how much to work. Gronau assumes that the offered wage W^0 is given to each housewife independently of hours worked H, rather than as a schedule $W^0(H)$. Given W^0, a housewife maximizes her utility function $U(C,X)$ subject to $X = W^0 H + V$ and $C + H = T$, where C is time spent at home for child care, X represents all other goods, T is total available time, and V is other income. Thus, a housewife does not work if

$$\left[\frac{\partial U}{\partial C}\Big/\frac{\partial U}{\partial X}\right]_{H=0} > W^0 \qquad (3.10)$$

and works if the inequality in (3.10) is reversed. If she works, the hours of work H and the actual wage rate W must be such that

$$\frac{\partial U}{\partial C}\Big/\frac{\partial U}{\partial X} = W .$$

Gronau calls the left-hand side of (3.10) the housewife's value of time, or, more commonly, the reservation wage, denoted W^r.[15]

Assuming that both W^0 and W^r can be written as linear combinations of independent variables plus error terms, his model may be statistically described as follows:

$$(3.11)\qquad \begin{cases} W_i^0 = x_{2i}'\beta_2 + u_{2i} \\ W_i^r = z_i'\alpha + v_i \\ W_i = \begin{cases} W_i^0 & \text{if} \quad W_i^0 > W_i^r \ , \\ 0 & \text{if} \quad W_i^0 \leq W_i^r \ , \quad i = 1,2,\ldots,n \ , \end{cases} \end{cases}$$

where (u_{2i}, v_i) is an i.i.d. drawing from a bivariate normal distribution with mean zero, variances σ_u^2 and σ_v^2, and covariance σ_{uv}. Thus, the model can be written in the form of (3.1) by putting $W_i^0 - W_i^r = y_{1i}^*$ and $W_i^0 = y_{2i}^*$. Note that H (hours worked) is not explained by this statistical model though it is determined by Gronau's theoretical model. A statistical model explaining H as well as W was later developed by Heckman [1974]. I will discuss this in the section on Type 3 models.

Since the model (3.11) can be transformed into the form (3.1) in such a way that the parameters of (3.11) can be determined from the parameters of (3.1), all the parameters of the model are identifiable except $V(W_i^0 - W_i^r)$, which can be set equal to 1 without loss of generality. If, however, at least one element of x_{2i} is not included in z_i, all the parameters are identifiable.16/ They can be estimated by the MLE or Heckman's two-step estimator by procedures described in Section B.1 above. One can also use the probit MLE (the first step of

Heckman's two-step) to estimate a certain subset of the parameters. However, the two estimation methods used by Gronau are not among the above. I will described his methods and explain in what way they are inappropriate.

The full likelihood function of Gronau's model (3.11) can be written as

$$(3.12)\qquad L = \prod_0 P(W_i^0 \leq W_i^r) \prod_1 \int_{-\infty}^{W_i} f(W_i, W_i^r)\, dW_i^r \ ,$$

where $\prod_0$ and $\prod_1$ are the products over those observations for which $W_i^0 \leq W_i^r$ and $W_i^0 > W_i^r$ respectively and $f(\cdot,\cdot)$ is the joint density of W_i^0 and W_i^r. Gronau assumes that u_{2i} and v_i are independent.17/ Under this assumption, (3.12) can be written as

$$(3.13)\qquad L = L^* \cdot \prod_1 \sigma_u^{-1} \phi[\sigma_u^{-1}(W_i - x_{2i}'\beta_2)] \ ,$$

where

$$(3.14)\qquad L^* = \prod_0 \{1 - \Phi\, [(\sigma_u^2 + \sigma_v^2)^{-\frac{1}{2}} (x_{2i}'\beta_2 - z_i'\alpha)]\} \cdot \prod_1 \Phi\, [\sigma_v^{-1}(W_i - z_i'\alpha)] \ .$$

Maximizing (3.13) yields the MLE of α, β_2, σ_u, and σ_v, which are consistent and asymptotically efficient under Gronau's independence assumption. Maximizing (3.14) yields estimates of α, β_2, and σ_v which are consistent but asymptotically not fully efficient. Gronau's two methods of estimation can be both regarded as attempts to maximize an approximation to (3.13) as I will show below.

In Gronau's first method, W_i is regressed on x_{2i} for those observations where $W_i^0 > W_i^r$ to yield the LS estimates $\hat{\beta}_2$ and then

$$(3.15) \qquad L^+ = \prod_0 \{1 - \Phi[\sigma_v^{-1}(x_{2i}'\hat{\beta}_2 - z_i'\alpha)]\} \prod_1 \Phi[\sigma_v^{-1}(W_i - z_i'\alpha)]$$

is maximized with respect to α and σ_v. There are two problems with this method: (1) $\hat{\beta}_2$ is not consistent, as Gronau notes, and (2) L^+ differs from the correct L^* in that σ_u^2 appears in (3.14) but not in (3.15). Note that this method would be MLE if $\prod_1$ appearing in (3.13) were the product over all the observations and if L^* were used instead of L^+.

In Gronau's second method, the first problem is solved as indicated below, but the second problem remains. We have under Gronau's independence assumption

$$(3.16) \qquad E(W_i|W_i^r < W_i^0) = x_{2i}'\beta_2 + (\sigma_u^2 + \sigma_v^2)^{-\frac{1}{2}} \sigma_u^2 \Phi_i^{-1}\phi_i \quad ,$$

where ϕ_i and Φ_i are ϕ and Φ evaluated at $(\sigma_u^2 + \sigma_v^2)^{-1/2}(x_{2i}'\beta_2 - z_i'\alpha)$. Since Gronau's data are such that there are many individuals with the same value of the independent variables, one can estimate Φ_i directly by the ratio of the number of working wives to the number of wives with the characteristics x_i. Given this estimate, denoted $\hat{\Phi}_i$, one can estimate

ϕ_i by $\hat{\phi}_i = \phi[\Phi^{-1}(\hat{\Phi}_i)]$. Next, one regresses positive W_i on x_{2i} and $\hat{\Phi}_i^{-1}\hat{\phi}_i$ to estimate β_2. This estimate, denoted $\tilde{\beta}_2$, is consistent (provided that the above estimates of ϕ_i and Φ_i are consistent), and, therefore, the first problem of the first estimation method is solved. Gronau, then, maximizez L^+ after replacing $\hat{\beta}_2$ by $\tilde{\beta}_2$, but, had he maximized L^* after replacing β_2 by $\tilde{\beta}_2$, he would have obtained consistent estimates of the remaining parameters.

Despite the minor error in the estimation method, Gronau's article made a significant econometric contribution (besides a substantive empirical contribution which I have ignored) by suggesting a two-step method based on the conditional expectation equation, which became a precursor of Heckman's two-step estimator.

For a panel-data generalization of Gronau's model, see Kiefer and Neumann [1979 and 1981]. The likelihood function of the Kiefer-Neumann model is obtained by taking the product of (3.12) over the time periods in the sample. They used the MLE.

4. Other Applications: Nelson [1977] noted that a Type 2 Tobit model arises if y_0 in (2.1) is assumed to be a random variable with its mean equal to a linear combination of independent variables. He reestimated Gronau's model by maximizing the correct likelihood function (3.13).

Dagenais [1975] used a Type 2 Tobit model to analyze household purchase of automobiles. In this model, y_2^* in (3.1) represents the desired expenditure on a car and x_2 includes permanent income, education, and the number of children. He assumes that a household purchases a car if y_2^* exceeds a stochastic threshold $S = \theta_1 + \theta_2 A + v$, where A is the dummy variable taking unity if the household anticipated buying a car at the time of a prior questionaire and the actual value of purchase $y_2 = y_2^*$ if $y_2^* > S$. Thus, $y_2^* - S$ plays the role of y_1^* in (3.1). Like Gronau, Dagenais assumes independence between y_2^* and S, and, in addition, he assumes equality of the variances of y_2^* and S. These assumptions are not necessary for identification. Dagenais' model, like Gronau's, has a weakness in that an arbitrary separation of the independent variables into some which go into the y_2^* equation and some which go into the S equation (W^r equation and W^0 equation in Gronau's model) is maintained.

In the study of Westin and Gillen [1978], y_2^* represents the parking cost with x_2 including zonal dummies, wage rate (as a proxy for value of walking time), and the square of wage rate. A researcher observes $y_2^* = y_2$ if $y_2^* < C$ where C represents transit cost, which itself is a function of independent variables plus an error term.

5. Cragg [1971]: As I mentioned in the beginning of Section B.1, y_{2i} can be negative in the Type 2 model. Cragg [1971] proposed three models that ensure the nonnegativity of y_{2i}. It will be readily seen that, strictly speaking, Cragg's models 2 and 3 defined below can be classified as nonnormal models and his model 1, though based on the normal distribution, does not belong to any of the five types.

Nevertheless, I discuss these models here as they can be regarded as modifications of the Type 2 model.

Model 1: $(y_1^*, y_2^*) \sim \text{Bivariate } N(x_1'\beta_1, x_2'\beta_2, 1, \sigma_2^2, \sigma_{12})$

$$y_2 = \begin{cases} y_2^* & \text{if} \quad y_1^* > 0 \text{ and } y_2^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Model 2: $(y_1^*, y_2^*) \sim \text{Bivariate } N(x_1'\beta_1, x_2'\beta_2, 1, \sigma_2^2, \sigma_{12})$ with y_2^* truncated so that $y_2^* > 0$

$$y_2 = \begin{cases} y_2^* & \text{if} \quad y_1^* > 0 \\ 0 & \text{if} \quad y_1^* \leq 0 \end{cases}$$

Model 3: Same as Model 2 except $\log y_2^* \sim N(x_2'\beta_2, \sigma_2^2)$.

Cragg compared the above three models and the Standard Tobit model by a simulation study. One purpose of his investigation was to see how close the MLE is to its asymptotic normal distribution in each model. His results were rather inconclusive. Another purpose of the study was to see how often a true model is selected against the other competing models by Bayes' posterior odds ratio. Cragg found it hard to distinguish between Models 1 and 2.

Cragg fitted a simplified version (fewer independent variables) of Wu's model [1965] by the four models and the ranking according to the posterior odds ratio ranked Model 3 best, followed in order by Model 2, Model 1, and finally Tobit. The estimates obtained by Models 2 and 3 were found to be similar to Wu's estimates.

C. Type 3: $P(y_1 < 0) \cdot P(y_1, y_2)$

1. Definition and Estimation: The Type 3 Tobit model is defined as follows:

$$
(3.17)\quad \begin{cases} y_{1i}^* = x_{1i}'\beta_1 + u_{1i} \\ y_{2i}^* = x_{2i}'\beta_2 + u_{2i} \\ y_{1i} = \begin{cases} y_{1i}^* & \text{if} \quad y_{1i}^* > 0 \\ 0 & \text{if} \quad y_{1i}^* \leqq 0 \end{cases} \\ y_{2i} = \begin{cases} y_{2i}^* & \text{if} \quad y_{1i}^* > 0 \\ 0 & \text{if} \quad y_{1i}^* \leqq 0 \end{cases}, \quad i = 1,2,\ldots,n , \end{cases}
$$

where $\{u_{1i}, u_{2i}\}$ are i.i.d. drawings from a bivariate normal distribution with mean zero, variances σ_1^2 and σ_2^2, and covariance σ_{12}. Note that this model differs from Type 2 only in that y_{1i}^* is also observed when it is positive in this model.

Since the estimation of this model can be handled similarly to that of Type 2, I will discuss it only briefly. Instead, in the following I will give a detailed discussion of the estimation of Heckman's model [1974], which constitutes the structural-equations version of the model (3.17).

The likelihood function of the model (3.17) can be written as

$$
(3.18)\qquad L = \prod_0 P(y_{1i}^* \leqq 0) \prod_1 f(y_{1i}, y_{2i}) ,
$$

where $f(\cdot,\cdot)$ is the joint density of y_{1i}^* and y_{2i}^*. Since y_{1i}^* is observed when it is positive, all the parameters of the model are identifiable, including σ_1^2.

Heckman's two-step estimator was originally proposed by Heckman [1976] for this model. Here we obtain two conditional-expectation equations (2.17) and (3.8) for y_1 and y_2 respectively. (Add subscript 1 to all the variables and the parameters in (2.17) to conform to the notation of the present section.) In the first step of the method, $\alpha_1 = \beta_1\sigma_1^{-1}$ is estimated by the probit MLE $\hat{\alpha}_1$. In the second step, least squares is applied separately to (2.17) and (3.8) after replacing α_1 by $\hat{\alpha}_1$. The asymptotic variance-covariance matrix of the resulting estimates of (β_1, σ_1) is given in (2.28) and that for $(\beta_2, \sigma_{12}\sigma_1^{-1})$ can be similarly obtained. The latter is given by Heckman [1979]. A consistent estimate of σ_2 can be obtained using the residuals of equation (3.8). As Heckman [1976] suggested and as I noted in Section II.C.3, a more efficient WLS can be used for each equation in the second-step of the method. An even more efficient GLS can be applied simultaneously to the two equations. However, even GLS is not fully efficient compared to MLE, and the added computational burden may not be sufficiently compensated for by the gain in efficiency. A two-step method based on unconditional means of y_1 and y_2, which is generalization of the method discussed in Section II.B.3, can be also used for this model.

Wales and Woodland [1980] compared the LS estimator, Heckman's two-step estimator, probit MLE, conditional MLE (using only those who worked), MLE, and another inconsistent estimator in a Type 3 Tobit model in a simulation study with one replication (sample size 1000 and 5000). The particular model they used is the labor supply model of Heckman [1974], which I will discuss in the next subsection.[18] The LS estimator was found to be poor, and all three ML estimators were found to perform well. Heckman's two-step estimator was ranked somewhere between LS and MLE.

2. Heckman [1974]: Heckman's model differs from Gronau's model (3.11) in that Heckman includes the determination of hours worked H in his model. Thus, Heckman's model is a natural consequence of Gronau's theory of labor supply. Like Gronau, Heckman assumes that the offered wage W^0 is given independently of H; therefore, Heckman's W^0 equation is the same as Gronau's:

$$W_i^0 = x_{2i}'\beta_2 + u_{2i} \quad . \tag{3.19}$$

Heckman defines $W^r = (\partial U/\partial C)/(\partial U/\partial X)$ and specifies[19]

$$W_i^r = \gamma H_i + z_i'\alpha + v_i \quad . \tag{3.20}$$

It is assumed that the i-th individual works if

$$W_i^r (H_i = 0) \equiv z_i'\alpha + v_i < W_i^0 \tag{3.21}$$

and then, the wage W_i and hours worked H_i are determined by solving (3.19) and (3.20) simultaneously after putting $W_i^0 = W_i^r = W_i$. Thus, we

can define Heckman's model as

$$W_i = x_{2i}'\beta_2 + u_{2i} \tag{3.22}$$

and

$$W_i = \gamma H_i + z_i'\alpha + v_i \tag{3.23}$$

for those i for which desired hours of work

$$H_i^* \equiv x_{1i}'\beta_1 + u_{1i} > 0 \quad , \tag{3.24}$$

where $x_{1i}'\beta_1 = \gamma^{-1}(x_{2i}'\beta_2 - z_i'\alpha)$ and $u_{1i} = \gamma^{-1}(u_{2i} - v_i)$. Note that (3.21) and (3.24) are equivalent because $\gamma > 0$.

I will call (3.22) and (3.23) the structural equations; then, (3.22) and the identity part of (3.24) constitute the reduced form equations. The reduced form equations of Heckman's model can be shown to correspond to the Type 3 Tobit model (3.17) if we put $H^* = y_1^*$, $H = y_1$, $W^0 = y_2^*$, and $W = y_2$. Since I have already discussed the estimation of the reduced-form parameters in the context of the model (3.17), I will now discuss the estimation of the structural parameters.

Heckman [1974] estimated the structural parameters by MLE. In the next two subsections I will discuss three alternative methods of estimating the structural parameters.

For a panel-data generalization of Heckman's model, see Heckman and MaCurdy [1980].

3. Heckman [1976]: This article proposes the Heckman two-step estimator of the reduced-form parameters, which I have discussed in subsection 1 above, but also reestimates the labor supply model of Heckman [1974] using the structural equation version. Since (3.22) is a reduced-form as well as a structural equation, the estimation of β_2 is done in the same way as I have discussed in subsection 1: namely, by applying least squares to the regression equation for $E(W_i|H_i^* > 0)$ after estimating the argument of λ (the hazard rate) by probit MLE. So I will only discuss the estimation of (3.23) here. Rewrite (3.23) as

$$(3.25) \qquad H_i = \gamma^{-1}W_i - z_i'\alpha\gamma^{-1} - \gamma^{-1}v_i \quad .$$

By subtracting $E(v_i|H_i^* > 0)$ from v_i and adding the same, we rewrite (3.25) further as

$$(3.26) \qquad H_i = \gamma^{-1}W_i - z_i'\alpha\gamma^{-1} - \sigma_{1v}\sigma_1^{-1}\gamma^{-1}\lambda(x_{1i}'\beta_1/\sigma_1) - \gamma^{-1}\varepsilon_i \quad ,$$

where $\sigma_{1v} = \text{Cov}(u_{1i}, v_i)$, $\sigma_1^2 = Vu_{1i}$, and $\varepsilon_i = v_i - E(v_i|H_i^* > 0)$. Then, consistent estimates of γ^{-1}, $\alpha\gamma^{-1}$, and $\sigma_{1v}\sigma_1^{-1}\gamma^{-1}$ are obtained by the least squares regression applied to (3.26) after replacing β_1/σ_1 by its probit MLE and W_i by $\hat{W}_i$, the least squares predictor of W_i obtained by applying Heckman's two-step estimator to (3.22). The asymptotic variance-covariance matrix of this estimator can be deduced from the results in Heckman [1978], who considered the estimation of a more general model (which I will discuss in the section on Type 5 Tobit models).

Actually, there is no apparent reason why one must first solve (3.23) for H_i and proceed as I have indicated above. Heckman could just as easily have subtracted and added $E(v_i|H_i^* > 0)$ to (3.23) itself and proceeded similarly. This method would yield alternative consistent estimates. Inferring from a well-known fact that the two-stage least squares estimates of the standard simultaneous equations model yield asymptotically equivalent estimates regardless of which normalization is chosen, I conjecture that the Heckman two-step method applied to (3.23) and (3.25) would also yield asymptotically equivalent estimates of γ and α.

Lee, Maddala, and Trost [1978] extended Heckman's simultaneous-equations two-step estimator and its WLS version (taking account of the heteroscedasticity) to more general simultaneous-equations Tobit models and obtained their asymptotic variance-covariance matrices.

4. Amemiya's LS and GLS: Amemiya [1978 and 1979] proposed a general method of obtaining the estimates of the structural parameters from given reduced-form parameter estimates in general Tobit-type models and derived the asymptotic distribution. The structural parameters γ and β of a particular equation are generally related to the relevant reduced-form parameters π and Π in the following way:

$$(3.27) \qquad \pi = \Pi\gamma + J\beta \quad ,$$

where J is a known matrix consisting of only ones and zeros. It is assumed that π, γ, and β are vectors and Π and J are matrices of conformable sizes. Equation (3.27) holds for Heckman's model and more general simultaneous-equations Tobit models, as well as the standard simultaneous-equations model.

Now, suppose certain estimates $\hat{\pi}$ and $\hat{\Pi}$ of the reduced-form parameters are given. Then, using them, we rewrite (3.27) as

$$\hat{\pi} = \hat{\Pi}\gamma + J\beta + (\hat{\pi} - \pi) - (\hat{\Pi} - \Pi)\gamma \quad . \tag{3.28}$$

Amemiya proposed applying LS and GLS estimation to (3.28). From Amemiya's result [1978], one can infer that Amemiya's GLS applied to Heckman's model yields more efficient estimates than Heckman's simultaneous-equations two-step estimator discussed above. Amemiya [1982] shows the superiority of the Amemiya GLS estimator to the WLS version of the Lee-Maddala-Trost estimator in a general simultaneous-equations Tobit model.

5. Other Examples: Shishko and Rostker [1976] used Heckman's model to explain the wage and hours worked in a second job. They estimated the wage equation (3.22) by least squares (yielding inconsistent estimates) and estimated the hours equation (3.25) by the Tobit MLE after replacing W_i by its least squares predictor. This estimation procedure is not recommended.

Roberts, Maddala, and Enholm [1978] estimated two types of simultaneous-equations Tobit models to explain how utility rates are determined. One of their models has a reduced form which is essentially Type 3 Tobit and the other is a simple extension of Type 3.

The structural equations of their first model are

$$y_{2i}^* = x_{2i}'\beta_2 + u_{2i} \tag{3.29}$$

and

$$y_{3i}^* = \gamma y_{2i}^* + x_{3i}'\beta_3 + u_{3i} \quad , \tag{3.30}$$

where y_{2i}^* is the rate requested by the i-th utility firm, y_{3i}^* is the rate granted for the i-th firm, x_{2i} includes the embedded cost of capital and the last rate granted minus the current rate being earned, and x_{3i} includes only the last variable mentioned. It is assumed that y_{2i}^* and y_{3i}^* are observed only if

$$y_{1i}^* \equiv z_i'\alpha + v_i > 0 \quad , \tag{3.31}$$

where z_i include the earnings characteristics of the i-th firm. (Vv_i is assumed to be unity.) The variable y_1^* may be regarded as an index affecting a firm's decision as to whether or not it requests a rate increase. The above model can be labelled as $P(y_1 < 0) \cdot P(y_1 > 0, y_2, y_3)$ in my short-hand notation and therefore is a simple generalization of Type 3. The authors' estimation method is that of Lee, Maddala, and Trost [1978] and can be described as follows: (1) Estimate α by the probit MLE. (2) Estimate β_2 by Heckman's two-step method. (3) Replace y_{2i}^* in the right-hand side of (3.30) by $\hat{y}_{2i}^*$ obtained in step (2) and estimate γ and β_3 by the least squares applied to (3.30) after adding the hazard rate term $E(u_{3i}|y_{1i}^* > 0)$.

The second model of Roberts, et. al. is the same as the first model except that (3.31) is replaced by

$$y_{2i}^* > R_i \quad , \tag{3.32}$$

where R_i refers to the current rate being earned, an independent variable. Thus, this model is essentially Type 3. (It would be exactly Type 3 if $R_i = 0$.) The estimation method is as follows: (1) Estimate β_2 by the Tobit MLE. (2) Repeat (3) as described in the preceding paragraph.

Nakamura, Nakamura, and Cullen [1979] estimated essentially the same model as Heckman [1974] using Canadian data on married women. They used the WLS version of Heckman's simultaneous-equations two-step estimator; that is, they applied WLS to (3.26). Nakamura and Nakamura [1981] estimated a more elaborate version of the preceding model incorporating income tax, leading to a complex nonlinear hours equation.

Hausman and Wise [1976, 1977, and 1979] used Type 3 and its generalizations to analyze the labor supply of participants in the Negative Income Tax (NIT) experiments. Their models are truncated models since they used observations on only those who participated in the experiments. The first model of Hausman and Wise [1977] is a minor variation of the Standard Tobit model where earnings Y follow

$$(3.33) \qquad Y_i = Y_i^* \quad \text{if} \quad Y_i^* < L_i \quad , \quad Y_i^* \sim N(x_i'\beta, \sigma^2) \quad ,$$

where L_i is a (known) poverty level which qualifies the i-th person to participate in the NIT program. It varies systematically with family size. The model is estimated by LS and MLE. (The LS estimates were always found to be smaller in absolute value, confirming Greene's result given in Section II.C.2.) In the second model of the same article, earnings are split into wage and hours as $Y = W \cdot H$, leading to the same equations

as Heckman's (3.22) and (3.23) except that the conditioning event is

$$(3.34) \qquad \log W_i + \log H_i < \log L_i$$

instead of Heckman's (3.24). Thus, this model is a simple extension of Type 3 and belongs to the same type of models as the first model of Roberts, Maddala, and Enholm [1978], which I discussed earlier, except for the fact that the model of Hausman and Wise is a truncated one. The model of Hausman and Wise [1979] also belongs to this type. The model of their [1976] article is an extension of (3.33), where earnings observations are split into the pre-experiment (subscript 1) and experiment (subscript 2) periods as

$$(3.35) \qquad Y_{1i} = Y_{1i}^* \quad \text{and} \quad Y_{2i} = Y_{2i}^* \quad \text{if} \quad Y_{1i}^* < L_i \quad .$$

Thus, the model is essentially Type 3, except for a minor variation due to the fact that L_i varies with i.

D. Type 4: $P(y_1 < 0, y_3) \cdot P(y_1, y_2)$

1. Definition and Estimation: The Type 4 Tobit model is defined as follows:

$$
(3.36) \quad \begin{cases}
y^*_{1i} = x'_{1i}\beta_1 + u_{1i} \\
y^*_{2i} = x'_{2i}\beta_2 + u_{2i} \\
y^*_{3i} = x'_{3i}\beta_3 + u_{3i} \\
y_{1i} = \begin{cases} y^*_{1i} & \text{if} \quad y^*_{1i} > 0 \\ 0 & \text{if} \quad y^*_{1i} \leqq 0 \end{cases} \\
y_{2i} = \begin{cases} y^*_{2i} & \text{if} \quad y^*_{1i} > 0 \\ 0 & \text{if} \quad y^*_{1i} \leqq 0 \end{cases} \\
y_{3i} = \begin{cases} y^*_{3i} & \text{if} \quad y^*_{1i} \leqq 0 \\ 0 & \text{if} \quad y^*_{1i} > 0 \end{cases} , \quad i = 1,2,\ldots,n ,
\end{cases}
$$

where $\{u_{1i}, u_{2i}, u_{3i}\}$ are i.i.d. drawings from a trivariate normal distribution.

This model differs from Type 3 defined by (3.17) only by the addition of y^*_{3i}, which is observed only if $y^*_{1i} \leqq 0$. The estimation of this model is not significantly different from that of Type 3. The likelihood function can be written as

$$
(3.37) \qquad L = \prod_0 \int_{-\infty}^{0} f_3(y^*_{1i}, y_{3i})dy^*_{1i} \cdot \prod_1 f_2(y_{1i}, y_{2i}) ,
$$

where $f_3(\cdot,\cdot)$ is the joint density of y^*_{1i} and y^*_{3i} and $f_2(\cdot,\cdot)$ is the joint density of y^*_{1i} and y^*_{2i}. Heckman's two-step method for this model is similar to the method for the preceding model. However, one must deal with three conditional expectation equation in the present model. The equation for y_{3i} will be slightly different from the other two because the variable is positive when y^*_{1i} is nonpositive. We obtain

$$
(3.38) \qquad E(y_{3i}|y^*_{1i} \leqq 0) = x'_{3i}\beta_3 - \sigma_{13}\sigma_1^{-1}\lambda(-x'_{1i}\beta_1/\sigma_1) .
$$

I will discuss three examples of the Type 4 Tobit model below: Kenny, Lee, Maddala, and Trost [1979], Nelson and Olson [1978], and Tomes [1981]. In the first two models, the y^* equations are written as simultaneous equations, like Heckman's model [1974], for which the reduced-form equations take the form of (3.36). Tomes' model has a slight twist. The estimation of the structural parameters of such models can be handled in much the same way as the estimation of Heckman's model [1974]: that is, by either Heckman's simultaneous-equations two-step method (and its Lee-Maddala-Trost extension) or by Amemiya's LS and GLS, both of which I discussed in Section C above.

In fact, these two estimation methods can easily accomodate the following very general simultaneous-equations Tobit model:

$$
(3.39) \qquad \Gamma' y^*_i = B' x_i + u_i \quad , \quad i = 1,2,\ldots,n ,
$$

where the elements of the vector y^*_i contain the following three types of variables: (1) always completely observable, (2) sometimes completely observable and sometimes observed to lie in intervals (like $y^*_{1i} > 0$), and (3) always observed to lie in intervals. Note that the variable classified as C in Table 3.2 belongs to Class (2) above, and the variable classified as B belongs to Class (3). The models of Heckman [1974], Kenny, Lee, Maddala, and Trost [1979], and Nelson and Olson [1978], as well as a few more models I will discuss under Type 5 such as Heckman [1978], are all special cases of the model (3.39).

2. Kenney, Lee, Maddala, and Trost [1979]: These authors tried to explain earnings differentials between those who went to college and those who did not. I will explain their model using the variables appearing in (3.36). In their model, y^*_1 refers to the desired years of college education, y^*_2 the earnings of those who go to college, and y^*_3 the earnings of those who do not go to college. A small degree of simultaneity is introduced into the model by letting y^*_1 appear in the right-hand side of the y^*_2 equation. The authors used the MLE. They note that the MLE iterations did not converge when started from the LS estimates, but did converge very fast when started from Heckman's two-step estimates (simultaneous-equations version).

3. Nelson and Olson [1978]: The empirical model actually estimated by these authors is more general than Type 4 and is a general simultaneous-equations Tobit model (3.39). The Nelson-Olson empirical model involves the following four elements of the vector y^*:

y^*_1: Time spent on vocational school training, completely observed if $y^*_1 > 0$, and otherwise, observed to lie in the interval $(-\infty, 0]$.

y^*_2: Time spent on college education, observed to lie in one of the three intervals $(-\infty, 0]$, $(0, 1]$, and $(1, \infty)$.

y^*_3: Wage, always completely observed.

y^*_4: Hours worked, always completely observed.

These variables are related to each other by simultaneous equations. However, they merely estimate each reduced-form equation separately by various appropriate methods and obtain the estimates of the structural parameters from the estimates of the reduced-form parameters in an arbitrary way.

The model which Nelson and Olson analyzes theoretically in more detail is the following two-equation model:

(3.40) $$y^*_{1i} = \gamma_1 y_{2i} + x'_{1i}\alpha_1 + v_{1i}$$

and

(3.41) $$y_{2i} = \gamma_2 y^*_{1i} + x'_{2i}\alpha_2 + v_{2i} \quad ,$$

where y_{2i} is always observed and y^*_{1i} is observed to be y_{1i} if $y^*_{1i} > 0$. This model may be used, for example, if one is interested in explaining only y^*_1 and y^*_3 in the Nelson-Olson empirical model. The likelihood function of this model may be characterized by $P(y_1 < 0, y_2) \cdot P(y_1, y_2)$, and therefore, the model is a special case of Type 4.

Nelson and Olson proposed estimating the structural parameters of this model by the following sequential method: (1) Estimate the parameters of the reduced-form equation for y_1^* by the Tobit MLE and that for y_2^* by LS. (2) Replace y_{2i} in the right-hand side of (3.40) by its LS predictor obtained in step (1) above and estimate the parameters of (3.40) by the Tobit MLE. (3) Replace y_{1i}^* in the right-hand side of (3.41) by its predictor obtained in step (1) and estimate the parameters of (3.41) by LS. Amemiya [1979] obtained the asymptotic variance-covariance matrix of the Nelso-Olson estimator and showed that the Amemiya GLS (see Section III.C.4) based on the same reduced-form estimates is asymptotically more efficient.

4. Tomes [1981]: Though it is not stated explicitly, Tomes' model can be defined by

$$y_{1i}^* = \gamma_1 y_{2i} + x_{1i}'\beta_1 + u_{1i} \quad , \tag{3.42}$$

$$y_{2i} = \gamma_2 y_{1i} + x_{2i}'\beta_2 + u_{2i} \quad , \tag{3.43}$$

and

$$y_{1i} = \begin{cases} y_{1i}^* & \text{if} \quad y_{1i}^* > 0 \\ 0 & \text{if} \quad y_{1i}^* \leqq 0 \quad , \end{cases} \tag{3.44}$$

where y_{1i} is the inheritance and y_{2i} is the recipient's income. Note that this model differs from Nelson's model defined by (3.40) and (3.41) only in that y_{1i}, not y_{1i}^*, appears in the right-hand side of (3.43). Assuming $\gamma_1\gamma_2 < 1$ for the logical consistency of the model (as in Amemiya [1974] and mentioned in Section III.A), we may rewrite (3.42) as

$$y_{1i}^* = (1 - \gamma_1\gamma_2)^{-1}[\gamma_1(x_{2i}'\beta_2 + u_{2i}) + x_{1i}'\beta_1 + u_{1i}] \tag{3.45}$$

and (3.43) as

$$y_{2i} = \begin{cases} y_{2i}^{(1)} \equiv (1 - \gamma_1\gamma_2)^{-1}[\gamma_2(x_{1i}'\beta_1 + u_{1i}) + x_{2i}'\beta_2 + u_{2i}] & \text{if} \quad y_{1i}^* > 0 \\ y_{2i}^{(0)} \equiv x_{2i}'\beta_2 + u_{2i} & \text{if} \quad y_{1i}^* \leqq 0 \quad . \end{cases} \tag{3.46}$$

Thus, the likelihood function of the model is

$$L = \prod_0 \int_{-\infty}^{0} f(y_{1i}^*, y_{2i}^{(0)})\, dy_{1i}^* \prod_1 f(y_{1i}, y_{2i}^{(1)}) \quad , \tag{3.47}$$

which is the same as (3.37).

E. Type 5: $P(y_1 < 0, y_2) \cdot P(y_1 > 0, y_2)$

1. Definition and Estimation: The Type 5 Tobit model is obtained from the Type 4 model (3.36) by omitting the equation for y_{1i}. One merely observes the sign of y_{1i}^*. Thus, the model is defined by

$$
(3.48)\quad \begin{cases}
y^*_{1i} = x'_{1i}\beta_1 + u_{1i} \\
y^*_{2i} = x'_{2i}\beta_2 + u_{2i} \\
y^*_{3i} = x'_{3i}\beta_3 + u_{3i} \\
y_{2i} = \begin{cases} y^*_{2i} & \text{if} \quad y^*_{1i} > 0 \\ 0 & \text{if} \quad y^*_{1i} \leqq 0 \end{cases} \\
y_{3i} = \begin{cases} y^*_{3i} & \text{if} \quad y^*_{1i} \leqq 0 \\ 0 & \text{if} \quad y^*_{1i} > 0 \quad , \quad i = 1,2,\ldots,n \quad , \end{cases}
\end{cases}
$$

where $\{u_{1i}, u_{2i}, u_{3i}\}$ are i.i.d drawings from a trivatiate normal distribution.

The likelihood function of the model is

$$
(3.49)\qquad L = \prod_0 \int_{-\infty}^{0} f_3(y^*_{1i}, y_{3i})\, dy^*_{1i} \cdot \prod_1 \int_0^{\infty} f_2(y^*_{1i}, y_{2i})\, dy^*_{1i} \quad ,
$$

where f_3 and f_2 are as defined in (3.37). Since this model is somewhat simpler than Type 4, the estimation methods I discussed in the preceding section apply to this model *a fortiori*. Hence, I will immediately go into the discussion of applications.

2. Lee [1978] and Lee and Trost [1978]: In Lee's model [1978], y^*_{2i} represents the logarithm of the wage rate of the i-th worker in case he or she joins the union and y^*_{3i} represents the same in case he or she does not join the union. Whether or not the worker joins the union is determined by the sign of the variable

$$
(3.50)\qquad y^*_{1i} = y^*_{2i} - y^*_{3i} + z'_i\alpha + v_i \quad .
$$

Since we observe only y^*_{2i} if the worker joins the union and y^*_{3i} if the worker does not, the logarithm of the observed wage, denoted y_i, is defined by

$$
(3.51)\qquad y_i = \begin{cases} y^*_{2i} & \text{if} \quad y^*_{1i} > 0 \\ y^*_{3i} & \text{if} \quad y^*_{1i} \leqq 0 \quad . \end{cases}
$$

Lee assumes that x_2 and x_3 (the independent variables in the y^*_2 and y^*_3 equations) include the individual characteristics of firms and workers such as regional location, city size, education, experience, race, sex, and health, whereas z includes certain other individual characteristics and variables which represent the monetary and non-monetary costs of becoming a union member. Since y^*_1 is unobserved except for the sign, the variance of y^*_1 can be assumed to be unity without loss of generality.

Lee estimated his model by Heckman's two-step method applied separately to the y^*_2 and the y^*_3 equations. In Lee's model, simultaneity exists only in the y^*_1 equation and hence is ignored in the application of Heckman's two-step method. Amemiya's LS or GLS, which accounts for the simultaneity, will of course work for this model as well and the latter will yield more efficient estimates, though, of course, not as fully efficient as the MLE.

The model of Lee and Trost [1978] is identical to Lee's model above except that y^*_{1i} is defined simply as $z'_i\alpha + v_i$ and does not depend on the difference $y_{2i}^* - y_{3i}^*$ as in Lee's model. Thus, there is no simultaneity in the Lee-Trost model. In their model, y_1^* and y_3^* represent annual expenditure on the housing owned and rented respectively, x_2 and x_3 include the age, race, sex of the family head, family size, income, city size, distance from center of city, and the relative price index of housing, while z includes all the independent variables above except the last. In estimation, Heckman's two-step estimates were obtained and then used to start the Newton-Raphson iteration.

3. Heckman [1978]: Heckman's model is a simultaneous equations model consisting of two equations

(3.52) $$y_{1i}^* = \gamma_1 y_{2i} + x_{1i}'\beta_1 + \delta_1 w_i + u_{1i}$$

and

(3.53) $$y_{2i} = \gamma_2 y_{1i}^* + x_{2i}'\beta_2 + \delta_2 w_1 + u_{2i} \quad ,$$

where we observe y_{2i}, x_{1i}, x_{2i}, and w_i defined by

(3.54) $$w_i = \begin{cases} 1 & \text{if} \quad y_{1i}^* > 0 \\ 0 & \text{if} \quad y_{1i}^* \leqq 0 \end{cases} \quad .$$

There are no empirical results in this article, but the same model is estimated by Heckman [1976], in which y_{2i}^* represents the average income of black people in the i-th state, y_{1i}^* the unobservable sentiment toward blacks in the i-th state, and $w_i = 1$ if an antidiscrimination law is instituted in the i-th state.

Another possible application of the model is to the same problem to which Lee's article was addressed (though Lee's model seems more suitable for this problem). Then, y_{2i} would represent the i-th worker's wage (or earnings) for both union and nonunion workers, and y_{1i}^* would represent the i-th worker's propensity to join a union. As I will discuss later in subsection 4, such an application was made by Schmidt and Strauss [1976] using a special case of Heckman's model.

When one solves (3.52) and (3.53) for y_{1i}^*, the solution should not depend upon w_i, for that would clearly lead to logical inconsistencies. Therefore, one must assume

(3.55) $$\gamma_1\delta_2 + \delta_1 = 0$$

in order for Heckman's model to be logically consistent.[20] Using the above constraint, the reduced-form equations (though strictly speaking not a reduced form because of the presence of w_i) of the model can be written as

(3.56) $$y_{1i}^* = x_i'\pi_1 + v_{1i}$$

and

(3.57) $$y_{2i} = \delta_2 w_i + x_i'\pi_2 + v_{2i} \quad ,$$

where one can assume $V v_{1i} = 1$ without loss of generality. Thus Heckman's model is a special case of Type 5 with just a constant shift between y_2^* and y_3^* (i.e., $y_{2i}^* = x_i'\pi_2 + v_{2i}$ and $y_{3i}^* = \delta_2 + x_i'\pi_2 + v_{2i}$). Moreover, if $\delta_2 = 0$, it is a special case of Type 5 where $y_2^* = y_3^*$.

Let us compare Heckman's reduced-form model defined by (3.56) and (3.57) with Lee's model. Heckman's (3.56) is essentially the same as Lee's (3.50). Lee's (3.51) can be rewritten as

$$
\begin{aligned}
(3.58)\qquad y_i &= w_i(x_{2i}'\beta_2 + u_{2i}) + (1 - w_i)(x_{3i}'\beta_3 + u_{3i}) \\
&= x_{3i}'\beta_3 + u_{3i} + w_i(x_{2i}'\beta_2 + u_{2i} - x_{3i}'\beta_3 - u_{3i}) \quad .
\end{aligned}
$$

By comparing (3.57) and (3.58), we readily see that Heckman's reduced-form model is a special case of Lee's model in which the coefficient multiplied by w_i is a constant.

Heckman proposed a sequential method of estimation for the structural parameters, which can be regarded as an extension of Heckman's simultaneous-equations two-step estimation discussed in Section III.C.3. His method consists of the following steps: (1) Estimate π_1 by applying the probit MLE to (3.56). Denote the estimator $\hat{\pi}_1$ and define $\hat{F}_i = F(x_i'\hat{\pi}_1)$. (2) Insert (3.56) into (3.53), replace π_1 with $\hat{\pi}_1$ and w_i with $\hat{F}_i$, and then estimate γ_2, β_2, and δ_2 by least squares applied to (3.53). (3) Solve (3.52) for y_{2i}, eliminate y_{1i}^* by (3.56), and then apply least squares to the resulting equation after replacing π_1 by $\hat{\pi}_1$ and w_i by $\hat{F}_i$ to estimate γ_1^{-1}, $\gamma_1^{-1}\beta_1$, and $\gamma_1^{-1}\delta_1$.

Amemiya [1978] derived the asymptotic variance-covariance matrix of Heckman's estimator defined above and showed that Amemiya's GLS (defined in Section III.C.4) applied to the model yields an asymptotically more efficient estimator in the special case of $\delta_1 = \delta_2 = 0$. As pointed out by Lee [1981], however, Amemiya's GLS can be also applied to the model with nonzero δ's as follows: (1) Estimate π_1 by the probit MLE $\hat{\pi}_1$ applied to (3.56). (2) Estimate δ_2 and π_2 by applying the instrumental variables method to (3.57) using $\hat{F}_i$ as the instrument for w_i. Denote these estimators as $\hat{\delta}_2$ and $\hat{\pi}_2$. (3) Derive the estimates of the structural parameters γ_1, β_1, δ_1, γ_2, β_2, and δ_2 from $\hat{\pi}_1$, $\hat{\pi}_2$, and $\hat{\delta}_2$ using the relationship between the reduced-form parameters and the structural parameters as well as the constraint (3.55) in the manner described in Section III.C.4. The resulting estimator can be shown to be asymptotically more efficient than Heckman's.

4. Schmidt and Strauss [1976] and Related Papers: Schmidt and Strauss [1976] studied the effect of unions on earnings and earnings on unions by the following model:

$$(3.59)\qquad P(w_i = 1|y_{2i}) = \mathcal{L}(x_i'\beta_1 + \gamma_1 y_{2i}) \quad ,$$

where $\mathcal{L}(x) = (1 + e^{-x})^{-1}$, and

$$(3.60)\qquad f(y_{2i}|w_i) = N(x_i'\beta_2 + \gamma_2 w_i, \sigma^2) \quad .$$

In this model, $w_i = 1$ if the i-th worker is a union member, y_{2i} represents the i-th worker's earnings, and x_i includes education, experience, race, sex, and regional dummies.

Equation (3.60) can be written as a regression equation like (3.57), but, unlike (3.57), w_i is independent of the error term of the regression because (3.60) describes a conditional distribution. From (3.59) and (3.60) one can derive the marginal distribution of w_i as

$$(3.61)\qquad P(w_i=1)=\mathcal{L}(x_i'\beta_1+\sigma^{-2}\gamma_2 x_i'\beta_2+2^{-1}\sigma^{-2}\gamma_2^2) \quad .$$

In the process of obtaining the above result, it becomes apparent that one must have

$$(3.62)\qquad \sigma^2\gamma_1=\gamma_2$$

in order for the model to be logically consistent because, unless (3.62) holds, y_{2i} will appear in the argument of $\mathcal{L}$ in the right-hand side of (3.61).[21/] Note that (3.61) can be written in the form of (3.56) with v_{1i} following a logistic distribution. Hence, we conclude that the Schmidt-Strauss model is essentially a special case of Heckman's model in which v_{1i} and v_{2i} are independent. This independence considerably simplifies the estimation: assuming that x_i does not contain a constant term, the MLE of all the parameters can be obtained by applying LS to (3.60) and the logit MLE to (3.61) separately.

Warren and Strauss [1979] used the same model as above to study a related but different problem. In their study, $w_i = 1$ if the i-th state has right-to-work legislation and y_{2i} represents the proportion of nonargicultural employement that is unionized. The constraint (3.62) was also ignored in this study.

Schmidt [1978] considered the same union and earnings problem using a model which is a slight generalization of the Schmidt-Strauss model. It can be interpreted as Heckman's model in which (3.57) is generalized as

$$(3.63)\qquad y_{2i}=w_i\cdot z_i'\alpha+x_i'\pi_2+v_{2i} \quad .$$

Note that this equation is between (3.57) and (3.58) in its degree of generality concerning the term multiplied by w_i. While it is more general than Heckman's model in this sense, it is more restrictive than Heckman's in the more significant sense that Schmidt, like Schmidt and Strauss or Warren and Strauss, assumes independence between v_{1i} and v_{2i}.

Another example of the Schmidt-Strauss model is the model of Ray [1981], in which $w_i = 1$ if nontariff barriers existed in the i-th industry (U.S. four-digit manufacturing industry), y_{2i} represents an average of tariffs within the i-th industry, and x_i includes various industry characteristics.

5. Disequilibrium Models: Disequilibrium models constitute an extensive area of research, in which numerous papers have been written. The lastest bibliography compiled by Quandt [1982] contains 93 economic-theoretic references and 121 econometric references concerning disequilibrium models. Some of the early econometric models are surveyed by Maddala and Nelson [1974]. See, also, Hartley [1976] for a connection between a disequilibrium model and the Standard Tobit model. Here I will only mention two basic models first discussed in the pioneering work of Fair and Jaffee [1972].

The simplest disequilibrium model of Fair and Jaffee is a special case of the Type 5 model (3.48), in which y_{2i}^* is the quantity demanded in the i-th period, y_{3i}^* is the quantity supplied in the i-th period, and $y_{1i}^* = y_{3i}^* - y_{2i}^*$. Thus, the actual quantity sold, which a researcher observes, is the minimum of supply and demand. The fact that the variance-covariance matrix of (y_1^*, y_2^*, y_3^*) is only of rank 2 because of the linear relationship above does not essentially change the nature of the model because the likelihood function (3.49) involves only bivariate densities.

Another model considered by Fair and Jaffee adds a price equation to the above as

$$y_{4i} = \gamma(y_{2i}^* - y_{3i}^*) \quad , \tag{3.64}$$

where y_{4i} denotes a change in the price at the i-th period. The likelihood function of this model can be written as[22/]

$$L = \prod_0 \int_{-\infty}^{0} f_3(y_{1i}^*, y_{3i} \mid y_{4i}) f(y_{4i}) d y_{1i}^* \cdot \prod_1 \int_0^{\infty} f_2(y_{1i}^*, y_{2i} \mid y_{4i}) f(y_{4i}) d y_{1i}^* \quad . \tag{3.65}$$

The form of the likelihood function does not change if one adds a normal error term to the right-hand side of (3.64). In either case, the model may be schematically characterized by

$$P(y_1 < 0, y_3, y_4) \cdot P(y_1 > 0, y_2, y_4) \quad , \tag{3.66}$$

which is a simple generalization of the Type 5 model.

6. Multivariate Generalizations: By a multivariate generalization of Type 5, I mean a model in which y_{2i}^* and y_{3i}^* in (3.48) are vectors, whereas y_{1i}^* is a scalar variable whose sign is observed as before. Therefore, the Fair-Jaffee model with likelihood function characterized by (3.66) is an example of this type of model.

In Lee's model [1977], the y_{2i}^* equation is split into two equations

$$C_{2i}^* = x_{2i}'\beta_2 + u_2 \tag{3.67}$$

and

$$T_{2i}^* = z_{2i}'\alpha_2 + v_2 \quad , \tag{3.68}$$

where C_{2i}^* and T_{2i}^* denote the cost and the time incurred by the i-th person travelling by a private mode of transportation and, similarly, the cost and the time of travelling by a public mode are specified as

$$C_{3i}^* = x_{3i}'\beta_3 + u_3 \tag{3.69}$$

and

$$T_{3i}^* = z_{3i}'\alpha_3 + v_3 \quad . \tag{3.70}$$

Lee assumes that C_{2i}^* and T_{2i}^* are observed if the i-th person uses a private mode and C_{3i}^* and T_{3i}^* are observed if he or she uses a public mode. A private mode is used if $y_{1i}^* > 0$, where y_{1i}^* is given by

$$y^*_{1i} = s'_i\delta_1 + \delta_2 T^*_{2i} + \delta_3 T^*_{3i} + \delta_4(C^*_{3i} - C^*_{2i}) + \varepsilon_i \quad . \tag{3.71}$$

Lee estimated his model by the following sequential procedure: (1) Apply the probit MLE to (3.71). (2) Apply LS to each of the four equations (3.67) through (3.70) after adding to the right-hand side of each the estimated hazard from step (1). (3) Predict the dependent variables of the four equations (3.67) through (3.70) using the estimates obtained in step (2) above, insert the predictors into (3.71) and apply the probit MLE again. (4) Calculate the MLE by iteration starting from the estimates obtained at the end of the step (3).

Willis and Rosen [1979] studied earnings differentials between those who went to college and those who did not using a more elaborate model than that of Kenny, Lee, Maddala, and Trost [1979], which I discussed in Section III.D.2. In the model of Kenny, et al. y^*_{1i} (the desired years of college education, whose sign determines whether one attends college) is specified not to depend directly on y^*_{2i} and y^*_{3i} (the earnings of the college-goer and the non-college-goer respectively). The first inclination of a researcher might be to hypothesize $y^*_{1i} = y^*_{2i} - y^*_{3i}$. However, this would be an oversimplification because the decision to go to college should depend on the difference in expected life time earnings rather than current earnings.

Willis and Rosen solved this problem by developing an ingenious theory of the maximization of discounted, expected life-time earnings, which led to the following model:

$$I^*_{2i} = x'_{2i}\beta_2 + u_2 \tag{3.72}$$

$$G^*_{2i} = z'_{2i}\alpha_2 + v_2 \tag{3.73}$$

$$I^*_{3i} = x'_{3i}\beta_3 + u_3 \tag{3.74}$$

$$G^*_{3i} = z'_{3i}\beta_3 + v_3 \tag{3.75}$$

and

$$R_i = s'_i\gamma + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \quad , \tag{3.76}$$

where I^*_{2i} and G^*_{2i} denote the initial earnings and the growth rate of earnings for the college-goer, I^*_{3i} and G^*_{3i} denote the same for the non-college-goer, and R_i denotes the discount rate. It is assumed that the i-th person goes to college if $y^*_{1i} > 0$ where

$$y^*_{1i} = \log I^*_{2i} - \log I^*_{3i} + \delta_0 + \delta_1 G^*_{2i} + \delta_2 G^*_{3i} + \delta_3 R_i \quad , \tag{3.77}$$

and that the variables with subscript 2 are observed if $y^*_{1i} > 0$, those with subscript 3 are observed if $y^*_{1i} \leq 0$, and R_i is never observed. Thus, the model is formally identical to Lee's model [1977]. The estimation method of Willis and Rosen used only the first two steps of Lee's method given above.

Borjas and Rosen [1980] used the same model as Willis and Rosen to study the earnings differential between those who changed jobs and those who did not within a certain period of observation.

7. <u>Multi-response Generalizations</u>: In all the models we have considered so far in Section E, the sign of y^*_{1i} determined two

basic categories of observations, such as union members versus non-union members, states with an anti-discrimination law versus those without, or college-goers versus non-college govers. By a multi-response generalization of Type 5, I mean a model in which observations are classified into more than two categories. I will devote most of this section to a discussion of Duncan [1980], who seems to be the first person to present a full discussion of estimation methods applicable to this type of model.

Duncan presents a model of joint determination of the location of a firm and its input-output vectors. A firm chooses the location for which profits are maximized, and only the input-output vector for the chosen location is observed. Let $s_i(k)$ be the profit of the i-th firm when it chooses location k, $i = 1,2,\ldots,n$ and $k = 1,2,\ldots,K$, and let $y_i(k)$ be the input-output vector for the i-th firm at the k-th location. To simplify the analysis, I will subsequently assume $y_i(k)$ is a scalar, for a generalization to the vector case is straightforward. It is assumed that

$$(3.78)\qquad s_i(k) = x_{ik}^{(1)\prime}\beta + u_{ik}$$

and

$$(3.79)\qquad y_i(k) = x_{ik}^{(2)\prime}\beta + v_{ik} ,$$

where $x_{ik}^{(1)}$ and $x_{ik}^{(2)}$ are vector functions of the input-output prices and economic theory dictates that the same β appears in both equations.[23/] It is assumed that $(u_{i1}, u_{i2},\ldots, u_{iK}, v_{i1}, v_{i2},\ldots, v_{iK})$ is an i.i.d.

drawing from a 2K-variate normal distribution. Suppose $s_i(k_i) > s_i(j)$ for any $j \neq k_i$. Then, a researcher observes $y_i(k_i)$ but does not observe $y_i(j)$ for $j \neq k_i$.

For the subsequent discussion it is useful to define K binary variables for each i by

$$(3.80)\qquad w_i(k) = \begin{cases} 1 & \text{if i-th firm chooses k-th location} \\ 0 & \text{otherwise} \end{cases}$$

and define the vector $w_i = [w_i(1), w_i(2),\ldots, w_i(K)]'$. Also define $P_{ik} = P(w_i(k) = 1)$ and the vector $P_i = (P_{i1}, P_{i2},\ldots, P_{iK})'$.

There are many ways to write the likelihood function of the model, but perhaps the most illuminating way is to write it as

$$(3.81)\qquad L = \prod_i f\,[y_i(k_i) \mid w_i(k_i) = 1]\, P_{ik_i}$$

where k_i is the actual location the i-th firm was observed to choose.

The estimation method proposed by Duncan can be outlined as follows: (1) Estimate the β that characterize f in (3.81) above by nonlinear WLS. (2) Estimate the β that characterize P in (3.81) above by the multi-response probit MLE using the nonlinear WLS iteration. (3) Choose the optimum linear combination of the two estimates of β obtained in steps (1) and (2). I will explain these steps in more detail below.

In order to describe step (1) explicitly, we must evaluate $\mu_i \equiv E[y_i(k_i)|w_i(k_i) = 1]$ and $\sigma_i^2 = V[y_i(k_i)|w_i(k_i) = 1]$ as functions of β and the variances and covariances of the error terms of equations (3.78) and (3.79). These conditional moments can be obtained as follows.

Define $z_i(j) = s_i(k_i) - s_i(j)$ and the $(K - 1)$-vector $z_i = [z_i(1),\ldots, z_i(k_i - 1), z_i(k_i + 1),\ldots, z_i(K)]'$. To simplify the notation, write z_i as z omitting the subscript. Similarly, write $y_i(k_i)$ as y. Also, define $R = E(y - Ey)(z - Ez)' \cdot [E(z - Ez)(z - Ez)']^{-1}$ and $Q = Vy - RE(y - Ey)(z - Ez)$. Then, we obtain[24]

$$\mu_i = E(y|z > 0) = Ey + RE(z|z > 0) - REz \tag{3.82}$$

and

$$\sigma_i^2 = V(y|z > 0) = RV(z|z > 0)R' + Q \quad . \tag{3.83}$$

The conditional moments of z appearing in the formulae above can be found in Amemiya [1974, p. 1002] as well as in Duncan [1980, p. 850]. Finally, I can describe the nonlinear WLS iteration of step (1) above as follows: Estimate σ_i^2 by inserting the initial estimates (for example, those obtained by minimizing $[y_i(k_i) - \mu_i]^2$) of the parameters into the right-hand side of (3.83)--call it $\hat{\sigma}_i^2$. Minimize

$$\sum_i \hat{\sigma}_i^{-2}[y_i(k_i) - \mu_i]^2 \tag{3.84}$$

with respect to the parameters that appear in the right-hand side of (3.82). Use these estimates to evaluate the right-hand side of (3.83) again to get another estimate of σ_i^2. Repeat the process, to yield new estimates of β.

Now, consider step (2). Define

$$\Sigma_i \equiv E(w_i - P_i)(w_i - P_i)' = D_i - P_iP_i' \quad , \tag{3.85}$$

where D_i is the $K \times K$ diagonal matrix whose k-th diagonal element is P_{ik}. To perform the nonlinear WLS iteration, first, estimate Σ_i by inserting the initial estimates of the parameters into the right-hand side of (3.85) (denote the estimate thus obtained as $\hat{\Sigma}_i$) and, second, minimize

$$\sum_i (w_i - P_i)' \hat{\Sigma}_i^- (w_i - P_i) \quad , \tag{3.86}$$

where the minus sign in the superscript denotes a generalized inverse, with respect to the parameters that characterize P_i, and repeat the process until the estimates converge.

Finally, regarding step (3) above, if we denote the two estimates of β obtained by step (1) and (2) by $\hat{\beta}_1$ and $\hat{\beta}_2$ respectively and their respective asymptotic variance-covariance matrices[25] by V_1 and V_2, the optimal linear combination of the two estimates is given by $(V_1^{-1} + V_2^{-1})^{-1} V_1^{-1}\hat{\beta}_1 + (V_1^{-1} + V_2^{-1})^{-1} V_2^{-1}\hat{\beta}_2$. This final estimator is asymptotically not fully efficient, however. To see this, suppose the regression coefficients of (3.78) and (3.79) differ: call them β_1 and β_2, say. Then, by a result of Amemiya [1976], we know that $\hat{\beta}_1$ is an asymptotically efficient estimator of β_1. However, as I have indicated in Section II.C.4, $\hat{\beta}_2$ is not. So a weighted average of the two could not be asymptotically efficient.

Dubin and McFadden [1980] used a similar model to Duncan's in their study of the joint determination of the choice of electric appliances and the consumption of electricity. In their model, $s_i(k)$ may be interpreted as the utility of the i-th family when they use the k-th portfolio of appliances, and $y_i(k)$ as the consumption of electricity for the i-th person holding the k-th portfolio. The estimation method is essentially similar to Duncan's. The main difference is that Dubin and McFadden assume that the error terms of (3.78) and (3.79) are distributed as Type I extreme value distribution and hence the P part of (3.81) is multinomial logit. (Cf. Amemiya [1981, p. 1516]).

Footnotes

1. The model is called <u>truncated</u> if the observations outside the specified range are totally lost and <u>censored</u> if one can at least observe the exogenous variables. A more precise definition will be given later.
2. See Kalbfleisch and Prentice [1980] and Miller [1981].
3. See Bartholomew [1973], Singer and Spilerman [1976], Tuma, Hannan, and Groeneveld [1979], Lancaster [1979], Tuma and Robins [1980], and Flinn and Heckman [1982].
4. In the Tobit model one needs to distinguish the vectors and matrices of positive observations from the vectors and matrices of all the observations. I will do so by putting the symbol _ under the latter.
5. Let $\log L(\theta)$ be a logarithmic likelihood function of a parameter vector θ in general. Then, global concavity means that $\partial^2 \log L/\partial\theta\partial\theta'$ is negative definite over the whole parameter space. Let $\hat{\theta}$ be the MLE. Then, by a Taylor expansion we have

$$\log L(\theta) = \log L(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})' \frac{\partial^2 \log L}{\partial\theta\partial\theta'}(\theta - \hat{\theta}) ,$$

where we have used the fact that $\partial \log L/\partial\theta$ evaluated at $\hat{\theta}$ is zero by definition of the MLE, and $\partial^2 \log L/\partial\theta\partial\theta'$ is evaluated at a point between θ and $\hat{\theta}$. Therefore, global concavity implies $\log L(\theta) < \log L(\hat{\theta})$ for any $\theta \neq \hat{\theta}$.

6. More precisely, $\stackrel{A}{=}$ means in this particular case that $\sqrt{n}$ times both sides of the equation have the same limit distribution.

7. $\lambda(\cdot)$ is known as the hazard rate and its reciprocal is known as Mills' ratio. Tobin [1958] gives a figure which shows that $\lambda(z)$ can be closely approximated by a linear function of z for $-1 < z < 5$. Johnson and Kotz [1970, p. 278f.] give various expansions of Mills' ratio.

8. See footnote 4.

9. This was suggested by Wales and Woodland [1980].

10. To the best of my knowledge, this result was first obtained by Stapleton and Young [1981].

11. See Amemiya [1981a]. Hartley [1976b] proved the asymptotic normality of $\hat{\gamma}_N$ and $\hat{\gamma}_{NW}$ and that they are asymptotically not as efficient as the MLE.

12. The asymptotic equivalence of $\tilde{\gamma}_N$ and $\tilde{\gamma}$ was proved by Stapleton and Young [1981].

13. We have by (2.55)

$$n[H(\theta|\theta_1) - H(\theta_1|\theta_1)] = E_{\theta_1} \log [k(y^*|\theta)/k(y^*|\theta_1)] \quad ,$$

where I have omitted the conditioning variable z to simplify notation and E_{θ_1} means that the expectation is taken on the assumption that the density of y^* is $k(y^*|\theta_1)$. But, by Jensen's inequality (see Rao [1973, p. 149)]

$$E_{\theta_1} \log [k(y^*|\theta)/k(y^*|\theta_1)] \leq \log E_{\theta_1} [k(y^*|\theta)/k(y^*|\theta_1)] \quad .$$

Thus, (2.57) follows from the above results and by noting

$$\log E_{\theta_1} [k(y^*|\theta)/k(y^*|\theta_1)] = \log \int k(y^*|\theta)\,dy^* = 0 \quad .$$

14. For an alternative account, see Hartley [1976c].

15. For a more elaborate derivation of the reservation wage model based on search theory, see Gronau [1974].

16. Gronau specifies that the independent variables in the W^r equation include woman's age and education, family income, number of children, and husband's age and education whereas the independent variables in the W^0 equation include only woman's age and education. However, Gronau readily admits to the arbitrariliness of the specification and the possibility that all the variables are included in both.

17. This may not be a realistic assumption since common independent variables, which are excluded from the set of regressors, may be included in both u_i and v_i. The assumption is not necessary if one uses either the MLE or Heckman's two-step estimator. It should be noted that the independence of u_i and v_i does not imply the independence of u_{1i} and u_{2i} in (3.1), so that Gronau's model is not as simple as the model considered in Section B.2 above. Also note that this assumption makes all the parameters identifiable even if no element of α is set equal to zero.

18. Though Heckman's model [1974] is a simultaneous-equations model, Heckman's two-step estimator studied by Wales and Woodland is essentially a reduced-form estimator which I have discussed in the

CONTINUED

1 OF 2

present section, rather than the structural equation version I will discuss in the next subsection.

19. Actually, Heckman uses $\log W^r$ and $\log W^0$. The independent variables x_2 include husband's wage, asset income, prices, and individual characteristics and z include housewife's schooling and experience.

20. Constraints like (3.51) are often necessary in simultaneous equations model involving binary or truncated variables, as was first noted by Amemiya [1974b]. For an interesting unified approach to this problem, see Gourieroux, Laffont, and Monfort [1980].

21. This constraint was overlooked by Schmidt and Strauss [1976] and was noted by Olsen [1978b].

22. A more explicit expression for the likelihood function was obtained by Amemiya [1974a], who pointed out the incorrectness of the likelihood function originally given by Fair and Jaffee.

23. (3.78) is the maximized profit function and (3.79) is an input demand or output supply function obtained by differentiating (3.78) with respect to the own input or output price (Hotelling's lemma). For convenience only one input or output has been assumed, so strictly speaking $x_{ik}^{(1)}$ and $x_{ik}^{(2)}$ are scalars.

24. These two equations correspond to the two equations in the proposition on p. 851 of Duncan [1980]. It seems that Duncan inadvertently omitted the last term from (3.82).

25. These matrices can be obtained by a standard procedure. See, for example, Amemiya [1981]. The matrices must be evaluated at some consistent estimates; either $\hat{\beta}_1$ or $\hat{\beta}_2$ will do.

References

Adams, J.D. [1980], "Personal Wealth Transfers," Quarterly Journal of Economics, Vol. 95, 159-179.

Amemiya, T. [1973], "Regression Analysis when the Dependent Variable Is Truncated Normal," Econometrica, Vol. 41, 997-1016.

Amemiya, T. [1974a], "A Note on a Fair and Jaffee Model," Econometrica, Vol. 42, 759-762.

Amemiya, T. [1974b], "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables Are Truncated Normal," Econometrica, Vol. 42, 999-1012.

Amemiya, T. [1976], "The Maximum Likelihood, the Minimum Chi-Square and the Nonlinear Weighted Leasted-Squares in the General Qualitative Response Models," Journal of the American Statistical Association, Vol. 71, 347-351.

Amemiya, T. [1978], "The Estimation of a Simultaneous Equation Generalized Probit Model," Econometrica, Vol. 46, 1193-1205.

Amemiya. T. [1979], "The Estimation of a Simultaneous-Equation Tobit Model," International Economic Review, Vol. 20, 169-181.

Amemiya, T. [1981a], "Nonlinear Regression Models," Technical Report No. 330, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Amemiya, T. [1981b], "Qualitative Response Models: A Survey," Journal of Economic Literature, Vol. 19, 1483-1536.

Amemiya, T. [1982], "A Comparison of the Amemiya GLS and the Lee-Maddala-Trost G2SLS in a Simultaneous-Equations Tobit Model," Mimeo., Rhodes Associates.

Amemiya, T. and M. Boskin [1974], "Regression Analysis when the Dependent Variable Is Truncated Lognormal, with an Application to the Determinants of the Duration of Welfare Dependency," International Economic Review, Vol. 15, 485-496.

Arabmazar, A. and P. Schmidt [1981], "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," Memeo., Michigan State University.

Ashenfelter, O. and J. Ham [1979], "Education, Unemployment, and Earnings," Journal of Political Economy, Vol. 87, S99-S116.

Bartholomew, D.J. [1973], Stochastic Models for Social Processes, 2nd Ed., John Wiley & Sons, Inc., New York.

Borjas, G.J. and S. Rosen [1980], "Income Prospects and Job Mobility of Young Men," Research in Labor Economics, Vol. 3, 159-181.

Cragg, J.G. [1971], "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica, Vol. 39, 829-844.

Dagenais, M.G. [1969], "A Threshold Regression Model," Econometrica, Vol. 37, 193-203.

Dagenais, M.G. [1975], "Application of a Threshold Regression Model to Household Purchases of Automobiles," Review of Economics and Statistics, Vol. 57, 275-285.

Dempster, A.P., N.M. Laird, and D.B. Rubin [1977], "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of Royal Statistical Society, Series B, Vol. 39, 1-38 (with discussion).

Dubin, J.A. and D. McFadden [1980], "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Mimeo., MIT.

Dudley, L. and C. Montmarquette [1976], "A Model of the Supply of Bilateral Foreign Aid," American Economic Review, Vol. 66, 132-142.

Duncan, G.M. [1980], "Formulation and Statistical Analysis of the Mixed, Continuous/Discrete Dependent Variable Model in Classical Production Theory," Econometrica, Vol. 48, 839-852.

Fair, R.C. [1978], "A Theory of Extramarital Affairs," Journal of Political Economy, Vol. 86, 45-61.

Fair, R.C. and D.M. Jaffee [1972], "Methods of Estimation for Markets in Disequilibrium," Econometrica, Vol. 40, 497-514.

Flinn, C.J. and J.J. Heckman [1982], "Models for the Analysis of Labor Force Dynamics," Working Paper No. 857, NBER.

Goldberger, A.S. [1964], Econometric Theory, John Wiley & Sons, Inc., New York.

Goldberger, A.S. [1980], "Abnormal Selection Bias," SSRI Workshop Series No. 8006, University of Wisconsin.

Goldberger, A.S. [1981], "Linear Regression After Selection," Journal of Econometrics, Vol. 15, 357-366.

Gourieroux, C., J.J. Laffont, and A. Monfort [1980], "Coherency Conditions in Simultaneous Linear Equation Models with Endogeneous Switching Regimes," Econometrica, Vol. 48, 675-695.

Greene, W.H. [1981], "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," Econometrica, Vol. 49, 505-513.

Gronau, R. [1973], "The Effects of Children on the Housewife's Value of Time," Journal of Political Economy, Vol. 81, S168-S199.

Gronau, R. [1974], "Wage Comparisons - a Selectivity Bias," Journal of Political Economy, Vol. 82, 1119-1143.

Hartley, H.O. [1958], "Maximum Likelihood Estimation from Incomplete Data," Biometrics, Vol. 14, 174-194.

Hartley, M.J. [1976a], The Estimation of Markets in Disequilibrium: The Fixed Supply Case," International Economic Review, Vol. 17, 687-699.

Hartley, M.J. [1976b], "Estimation of the Tobit Model by Nonlinear Least Squares Methods," Discussion Paper No. 373, SUNY at Buffalo.

Hartley, M.J. [1976c], "The Tobit and Probit Models: Maximum Likelihood Estimation by Ordinary Least Squares," Discussion Paper No. 374, SUNY at Buffalo.

Hausman, J.A. [1978], "Specification Tests in Econometrics," Econometrica, Vol. 46, 1251-1271.

Hausman, J.A. and D.A. Wise [1976], "The Evaluation of Results from Truncated Samples: The New Jersey Income Maintenance Experiment," Annals of Economic and Social Measurement, Vol. 5, 421-445.

Hausman, J.A. and D.A. Wise [1977], "Social Experimentation, Truncated Distributions, and Efficient Estimation," Econometrica, Vol. 45, 919-938.

Hausman, J.A. and D.A. Wise [1979], "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," Econometrica, Vol. 47, 455-473.

Heckman, J.J. [1974], "Shadow Prices, Market Wages, and Labor Supply," Econometrica, Vol. 42, 679-693.

Heckman, J.J. [1976], "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, Vol. 5, 475-492.

Heckman, J.J. [1978], "Dummy Endogeneous Variables in a Simultaneous Equation System," Econometrica, Vol. 46, 931-960.

Heckman, J.J. [1979], "Sample Selection Bias as a Specification Error," Econometrica, Vol. 47, 153-161.

Heckman, J.J. and T.E. MaCurdy [1980], "A Life Cycle Model of Female Labor Supply," Review of Economic Studies, Vol. 47, 47-74.

Huang, D.S. [1964], "Discrete Stock Adjustment: The Case of Demand for Automobiles," International Economic Review, Vol. 5, 46-62.

Hurd, M. [1979], "Estimation in Truncated Samples when There is Heteroscedasticity," Journal of Econometrics, Vol. 11, 247-258.

Jhonson, N.L. and S. Kotz [1970], Continuous Univariate Distributions 1, Houghton Mifflin, Boston.

Judge, N.L., W.E. Griffiths, R.C. Hill, and T.C. Lee [1980], Theory and Practice of Econometrics, John Wiley & Sons, Inc., New York.

Kalbfleish, J.G. and R.L. Prentice [1980], The Statistical Analysis of Failure Time Data, John Wiley & Sons, Inc., New York.

Keeley, M.C., P.K. Robins, R.G. Spiegelman, and R.W. West [1978], "The Estimation of Labor Supply Models Using Experimental Data," American Economic Review, Vol. 68, 873-887.

Kenny, L.W., L.F. Lee, G.S. Maddala, and R.P. Trost [1979], "Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data," International Economic Review, Vol. 20, 775-789.

Kiefer, N.M. and G.R. Neumann [1979], "An Empirical Job-Search Model, with a Test of the Constant Reservation-Wage Hypothesis," Journal of Political Economy, Vol. 87, 89-107.

Kiefer, N.M. and G.R. Neumann [1981], "Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job-Search Model," Econometrica, Vol. 49, 965-979.

Kotlikoff, L.J. [1979], "Testing the Theory of Social Security and Life Cycle Accumulation," American Economic Review, Vol. 69, 396-410.

Lancaster, T. [1979], "Econometric Methods for the Duration of Unemployement," Econometrica, Vol. 47, 939-956.

Lee, L.F. [1977], "Estimation of a Modal Choice Model for the Work Journey with Incomplete Observations," Mimeo., Department of Economics, University of Minnesota.

Lee, L.F. [1978], "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," International Economic Review, Vol. 19, 415-433.

Lee, L.F., G.S. Maddala, and R.P. Trost [1980], "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," Econometrica, Vol. 48, 491-503.

Lee, L.F., and R.P. Trost [1978], "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand," Journal of Econometrics, Vol. 8, 357-382.

Maddala, G.S. [1977a], "Identification and Estimation Problems in Limited Dependent Variable Models" in National Resources, Uncertainty and General Equilibrium Systems, A.S. Blinder and P. Friedman, eds. Academic Press, New York, 219-239.

Maddala, G.S. [1977b], "Self-Selectivity Problems in Econometric Models," in Applications of Statistics, P.R. Krishniah, ed., North Holland Pub. Co., Amsterdam, 351-366.

Maddala, G.S. [1977c], Econometrics, McGraw-Hill, New York.

Maddala, G.S. [1980], "Disequilibrium, Self-Selection and Switching Models," Social Science Working Paper 303, Cal. Institute of Tech. (To appear in Handbook of Econometrics, Z. Griliches and M.D. Intrilligator, eds., North-Holland Pub. Co., Amsterdam.)

Maddala, G.S. and F.D. Nelson [1974], "Maximum Likelihood Methods for Models of Markets in Disequilibrium," Econometrica, Vol. 42, 1013-1030.

Miller, R. [1981], Survival Analysis, John Wiley & Sons, Inc., New York.

Nakamura, A. and M. Nakamura [1981], "A Comparison of the Labor Force Behavior of Married Women in the United States and Canada, with Special Attention to the Impact of Income Taxes," Econometrica, Vol. 49, 451-489.

Nakamura, M., A. Nakamura, and D. Cullen [1979], "Job Opportunities, the Offered Wage, and the Labor Supply of Married Women," American Economic Review, Vol. 69, 787-805.

Nelson, F.D. [1977], "Censored Regression Models with Unobserved, Stochastic Censoring Thresholds," Journal of Econometrics, Vol. 6, 309-327.

Nelson, F.D. [1981], "A Test for Misspecification in the Censored Normal Model," Econometrica, Vol. 49, 1317-1329.

Nelson, F.D. and L. Olson [1978], "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables," International Economic Review, Vol. 19, 695-709.

Olsen, R.J. [1978a], "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," Econometrica, Vol. 46, 1211-1215.

Olsen, R.J. [1978b], "Comment on 'The Effect of Unions on Earnings and Earnings on Unions: A Mixed Logit Approach,'" International Economic Review, Vol. 19, 259-261.

Powell, J.L. [1981], "Least Absolute Deviations Estimation for Censored and Truncated Regression Models," Technical Report No. 356, IMSSS, Stanford University.

Quandt, R.E. [1982], "Bibliography of Quantity Rationing and Disequilibrium Models," Department of Economics, Princeton University.

Rao, C.R. [1973], Linear Statistical Inference and Its Application, 2nd Edition, John Wiley and Sons, Inc., New York.

Ray, E.J. [1981], "The Determinants of Tariff and Nontariff Trade Restrictions in the United States," Journal of Political Economy, Vol. 89, 105-121.

Reece, W.S. [1979], "Charitable Contributions: The New Evidence on Household Behavior," American Economic Review, Vol. 69, 142-151.

Roberts, R.B., G.S. Maddala, and G. Enholm [1978], "Determinants of the Requested Rate of Return and the Rate of Return Granted in a Formal Regulatory Process," Bell Journal of Economics, Vol. 9, 611-621.

Robinson, P.M., [1982], "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables," Econometrica, Vol. 50, 27-41.

Rosenzweig, M.R. [1980], "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labor Supply in a Developing Country," Quarterly Journal of Economics, Vol. 94, 31-55.

Rosett, R.N. [1959], "A Statistical Model of Friction in Economics," Econometrica, Vol. 27, 263-267.

Rosett, R.N. and F.D. Nelson [1975], "Estimation of the Two-Limit Probit Regression Model," Econometrica, Vol. 43, 141-146.

Ruud, P.A. [1982], "A Score Test of Consistency," Mimeo., University of California, Berkeley.

Schmee, J. and G.J. Hahn [1979], "A Simple Method for Regression Analysis with Censored Data," Technometrics, Vol. 21, 417-432.

Schmidt, P. and R.P. Strauss [1976], "The Effect of Unions on Earnings and Earnings on Union: A Mixed Approach," International Economic Review, Vol. 17, 204-212.

Shishko, R. and B. Rostker [1976], "The Economics of Multiple Job Holding," American Economic Review, Vol. 66, 298-308.

Singer, B and S. Spilerman [1976], "Some Methodological Issues in the Analysis of Longitudinal Surveys," Annals of Economic and Social Measurement, Vol. 5, 447-474.

Stapleton, D.C. and D.J. Young [1981], "Censored Normal Regression with Measurement Error on the Dependent Variable," Discussion Paper No. 81-30, Department of Economics, University of British Columbia.

Stephenson, S.P. and J.F. McDonald [1979], "Disaggregation of Income Maintenance Impacts on Family Earnings," Review of Economics and Statistics, Vol. 61, 354-360.

Tobin, J. [1958], "Estimation of Relationships for Limited Dependent Variables," Econometrica, Vol. 26, 24-36.

Tomes, N. [1981], "The Family, Inheritance, and the Intergenerational Transmission of Inequality," Journal of Political Economy, Vol. 89, 928-958.

Tuma, N.B., M.T. Hannan, and L.P. Groeneveld [1979], "Dynamic Analysis of Event Histories," American Journal of Sociology, Vol. 84, 820-854.

Tuma, N.B. and P.K. Robins [1980], "A Dynamic Model of Employment Behavior: An Application to the Seattle and Denver Income Maintenance Experiments," Econometrica, Vol. 48, 1031-1052.

Wales, T.J. and A.D. Woodland [1980], "Sample Selectivity and the Estimation of Labor Supply Functions," International Economic Review, Vol. 21, 437-468.

Warren, R.S., and R.P. Strauss [1979], "A Mixed Logit Model of the Relationship between Unionization and Right-to-Work Legislation," Journal of Political Economy, Vol. 87, 648-655.

Westin, R.B. and P.W. Gillen [1978], "Parking Location and Transit Demand," Journal of Econometrics, Vol. 8, 75-101.

Wiggins, S.N. [1981], "Product Quality Regulation and New Drug Introductions: Some New Evidence from the 1970's," Review of Economics and Statistics, Vol. 63, 615-619.

Willis, R.J. and S. Rosen [1979], "Education and Self-Selection," Journal of Political Economy," Vol. 87, S7-S36.

Witte, A.D. [1980], "Estimating the Economic Model of Crime with Individual Data," Quarterly Journal of Economics, Vol. 94, 57-84.

Wu, D.M. [1965], "An Empirical Analysis of Household Durable Goods Expenditure," Econometrica, Vol. 33, 761-780.

Addendum

Lee, L.F. []981], "Simultaneous Equations Models with Discrete and Censored Variables," in Manski, C.F. and D. McFadden, eds., Structured Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass.

END