Chapter 1

Introduction

Economists frequently encounter situations where the observations on the dependent variable in a regression model are limited to a certain range. For example, in labour supply functions where hours worked is explained as a function of wage rate and a number of other independent variables, the dependent variable cannot be negative, and it will be equal to zero for all those individuals who do not work. Thus, in this and other similar situations, the stochastic specification of the function takes into account the non-negativity of the dependent variable as well as the clustering of a number of observations at zero. If traditional least squares techniques are applied to estimate the model, the resulting estimates will generally be biased and inconsistent and therefore are no longer appropriate.

In economics, this model was first suggested by Tobin (1958) who analysed household expenditure on durables by considering the fact that the dependent variable cannot take negative values. Tobin's model and its generalisations are usually known as tobit models because of their relationship to the probit model. Tobit models are also known as censored or truncated regression models. A model is called censored if all the independent variables are observable but the dependent variable cannot be observed outside a specified range. If, on the other hand, both the dependent and the independent variables are not observable outside a certain range, it is referred to as a truncated regression model.

In recent years numerous applications of tobit models have appeared over a wide range of areas in economics. Examples of applications include labour supply models [Heckman (1976, 1979), Keely, Robins, Spiengelman and West (1978) and Wales and Woodland (1980)], demand for housing [Lee and Trost (1978)], modelling for public decision [Foot and Poirier (1980)], household expenditure models [Jarque (1987)], Demand for imports [Wu (1992)]. Theoretical and empirical surveys on tobit models were given by Amemiya (1981, 1984).

The increase in applications of tobit models has been associated with the increase of survey data for which tobit model analysis is well suited and with the availability of computer technology. On the other hand, many types of tobit models have been suggested and various estimation methods proposed. In fact, Amemiya (1984, p.4) stated,

> *'... models and estimation methods are now so numerous and diverse that it is difficult for econometricians to keep track for all the existing models and estimation methods and maintain a clear notion of their relative merits'.*

Examples of papers which are related to the theoretical aspects of various estimators and their properties include those of Amemiya (1973, 1978, 1981), Goldberger (1980, 1981), Greene (1981a, 1981b, 1983, 1990), Heckman (1979), Olsen (1978), White (1980b), Powell (1984, 1986b), Peracchi (1990) and Chib (1992). Furthermore, related models have been studied in the statistical literature of physical sciences [for example, Scheme and Hahn (1979), Aitkin (1981)].

However, almost all the theoretical studies are concerned with the asymptotic or large sample properties and/or computational ease of alternative estimators. In other words little attention is given to the finite (small) sample properties of the various estimators. Thus the purpose of this study is to make a contribution towards this end. Specifically, we use Monte Carlo techniques to assess the relative performance of the various estimators of the model.

Moreover, this study suggests an improved estimator for the tobit model along the lines of the existing Heckman's estimator and provides the asymptotic properties of the estimator. Furthermore, the finite sample properties of the proposed estimator are also investigated and compared along with other existing estimators of the model.

1.1 Objectives of the Study

This study is aimed towards a performance comparison of several estimators of the tobit model through a Monte Carlo experiment. There are many types of tobit models; this study will concentrate on the estimators that are relevant for what is known as the standard tobit model. Along the lines of existing Heckman two-step estimators of the tobit model, the study also proposes an alternative three-step estimator and its weighted version, the weighted three-step estimator, for the standard tobit model.

In general the objectives of this study can be summarized as follows:

1. To investigate the relative performance of the alternative estimators that have been proposed in the literature. That is,

   1.1 To investigate the relative efficiency of the various estimators.

   1.2 To investigate whether the asymptotic distributions of the various estimators are good guides to the finite sample properties of the estimators and to determine the sample sizes where the asymptotic distributions of the estimators can be used as good guides.

   1.3 To examine whether the asymptotic covariance matrices are accurate (or good) estimates of the finite sample covariance matrices and their implications for hypothesis tests and/or confidence intervals of coefficients.

   1.4 To assess the relative effects of relaxing the assumption of normally distributed disturbances which is a standard assumption in almost all cases.

   1.5 To investigate the effects of the degree of censoring on the performance of the estimators of the model.

2. To propose an alternative estimator for the standard tobit model and provide its asymptotic properties. Furthermore, to investigate the finite sample properties of the suggested estimator and compare the results with those of the other estimators.

3. To investigate the use of the alternative, but asymptotically equivalent, variance-covariance matrix estimators in the estimation of variances and

their implications in statistical inference, i.e., for hypothesis testing and/or confidence intervals of the coefficients of the model. This is particularly relevant in the maximum likelihood framework.

4. To examine the effects of correlation between the explanatory variables on the performance of the estimators.

5. To provide specific recommendations that can be used as a guide in applied research.

Note that the sequence of these objectives does not necessarily imply one is more important than the other. Some of these objectives will be elaborated upon later in this study under their respective chapter or topic headings.

1.2 Outline of the Study

The main focus of this research is concerned with the small sample performance of the estimators of the standard tobit model. Thus, although there are many types of tobit models, it is the aim of this research to concentrate on the literature that is most relevant to the standard tobit model.

Chapter 2 starts with the definition and the specification of the standard tobit model, the model whose estimators are to be investigated in this thesis. Then, the various estimators of the model and their properties are discussed in the subsequent sections of the Chapter. These estimators include, among others, the maximum likelihood estimator, Heckman's two-step estimator and its weighted version, the weighted Heckman's two-step estimator, nonlinear least squares estimators and other

Heckman-type estimators of the model. This Chapter also provides a brief review of the non-parametric, bounded influence and Bayesian methods of estimation of the tobit model.

Chapter 3 proposes an alternative estimator for the tobit model along the lines of the existing Heckman's two-step estimator. This estimator is referred to as the three-step estimator. The properties of the proposed estimator, i.e., its consistency and asymptotic distribution, are derived and discussed in this Chapter. Further, this Chapter provides some generalizations of the three-step procedure in estimating the standard tobit model as well as other similar models.

In Chapter 4, studies which are related to the small sample properties (Monte Carlo and/or simulation studies) viz-a-viz the model defined in Chapter 2 are discussed. Although there are many types of tobit models and related small sample studies in the literature, this Chapter concentrates mainly on those studies that are most relevant to the finite sample properties of the estimators of the standard tobit model.

Chapter 5 deals with the design of the Monte Carlo experiment which is employed in this study. This Chapter starts by defining the specific form of the model to be investigated in the Monte Carlo experiment. Once the model is specified it is followed by the various details that are involved in the data generation and estimation processes of the experiment. The main points discussed in this Chapter include the generation of the explanatory variables of the model, the determination of the true values (parameter values), the determination of sample sizes and the degrees of censoring. Further, this Chapter discusses the various distributions considered for the random error term of the model. Other important issues such as the generation mechanism of the random variates associated with each distribution as well as the output statistics to be computed in the Monte Carlo experiment for comparison purposes are included.

The main comparisons and findings of the experimental results begin in Chapter 6. This Chapter presents a detailed analysis and comparison of most of the estimators discussed in Chapters 2 and 3 of the study. A total of 11 estimators are included at the beginning of the Chapter. These estimators are compared using several criteria, as outlined in Chapter 5. A few estimators are then selected on a step by step basis depending on their relative performance. While comparing the various estimators, this Chapter also raises several questions and issues some of which are considered for further analysis in the subsequent chapters of this study.

Chapter 7 is entirely devoted to a further investigation of the small sample properties of the maximum likelihood estimator of the model, which is one of the most frequently used estimators in applied research. This Chapter is motivated by the findings and questions raised in Chapter 6 as well as by taking into consideration the wide use of the maximum likelihood estimator in applied research. The Chapter provides more information with regard to the consistency of the maximum likelihood estimator under a variety of conditions. Furthermore, this Chapter presents a detailed analysis on the performance of the alternative, but asymptotically equivalent, variance-covariance matrix estimators in the estimation of variances of the coefficients of the model as well as their implications for hypothesis testing and/or construction of confidence intervals for the coefficients of the model.

Chapter 8 examines the effects of correlation between the explanatory variables and the estimated inverse of Mill's ratio on the performance of the estimators of the model; in particular on the performance of Heckman's two-step and the three-step estimators. The main purpose of this Chapter is to examine the advantages of using the three-step estimator proposed in this study as compared to the usual Heckman's two-step estimator which often is characterized by a strong and unavoidable correlation problem.

This is followed by Chapter 9 which presents selective discussion on various topics related to this study. These topics cover a range of issues that are related to the design of the experiment as well as the outcomes (findings) of the study. This Chapter is designed to provide more information regarding the flexibility (or restrictiveness) of the experimental design as well as to examine the implications of some of the findings of the experiment for applied research.

Finally, Chapter 10 presents the summary, conclusions and recommendations of the study.

Chapter 2

Review of Literature

2.1 Introduction

Tobit models refer to regression models involving dependent variables for which observations are limited to a certain range. In economics, the tobit model was first suggested in the pioneering work of Tobin (1958). He investigated the relationship between household expenditure on durable goods, income and a number of other explanatory variables by taking into account the non-negativity of the dependent variable in the model. The name 'Tobit', which refers to Tobin's probit, was nicknamed by Goldberger (1964) because of its relevance to probit models. Tobit models are also referred to as censored or truncated models. They are called truncated if the observations outside a specified range are totally unobserved and censored if one can observe at least the exogenous variables.

Although the tobit model was first suggested by Tobin in 1958, there was very slow progress both in theoretical and empirical applications in the economic literature between this time and the 1970's. However, in the last two decades, a vast amount of applied and theoretical papers have appeared in a wide range of areas in economics. As a result many types of tobit models have been suggested and various estimation methods proposed. Amemiya (1981, 1984) provides a survey on the theoretical and empirical applications of the model [see also the books by Manski and McFadden (1981), Maddala (1983), Amemiya (1985), Judge et al. (1985) and Greene (1991)].

It is clear that the literature on tobit models is vast, covering from the simplest single equation tobit model to more complex tobit models involving simultaneous equations. Amemiya (1984) used five classifications for convenience, and defined the standard tobit model suggested by Tobin (1958) as Type-I. Similarly, there are many estimators that have been suggested in the literature to estimate the parameters of these models. Of course, like any other models, most of the estimators of the model are quite similar and are usually generalizations of the simplest cases of the model.

The main focus of this Chapter is to discuss the various estimators of the standard tobit model (Type-I Tobit) and their properties. In other words, although there are many types of generalisations of the tobit model, the review of the literature is limited to those estimators which are most relevant to the standard tobit model. However, as mentioned above, it is also important to note that many of the properties of these estimators may apply to other types of tobit models with some adjustments.

In general, the review of literature in this Chapter is organized as follows: Section 2.1 defines the specification of the standard tobit model. The properties of the traditional least squares estimators and their limitations in estimating the standard tobit model are discussed in Section 2.3. Sections 2.4 through 2.8 present the various theoretically feasible estimators of the model and their properties. These estimators include, among others, the maximum likelihood estimator, the Heckman's two-step estimator and its variations, and some nonlinear least squares estimators. Highlights of semi-parametric, bounded-influence and Bayesian estimators of the models are also provided in Section 2.9. Section 2.10 presents some useful results associated with the tobit model. Finally, a summary appears in Section 2.11.

2.2 The Standard Tobit Model

As stated above, the overall objectives of this study relate to the estimators of the tobit model and their finite sample properties, with particular emphasis on the standard tobit model. The standard tobit model is defined as follows:

$$y_i^* = x_i'\beta + u_i, \qquad i = 1, ..., N. \tag{2.1}$$

$$\begin{aligned} y_i &= y_i^* \quad \text{if} \quad y_i^* > 0, \\ &= 0 \quad \text{if} \quad y_i^* \leq 0. \end{aligned} \tag{2.2}$$

where x_i' is a (1xk) vector of explanatory variables which are assumed to be observed,

β is a (kx1) vector of unknown parameters to be estimated,

u is an (Nx1) vector of random disturbances, u_i, which are assumed to be independently and identically distributed drawings from N(0,σ^2),

y^* is an (Nx1) vector of values on the latent variable, y_i^*,

y is an (Nx1) vector of observations on the dependent variable, y_i, consisting of positive (non-limit) observations corresponding to the positive values of y_i^* and zero (limit) observations corresponding to those observations for which $y_i^* \leq 0$.

Since there are many generalizations of the tobit model, the model specified in (2.1)-(2.2) above is usually referred to as the standard tobit model, or simply the tobit model. It is also sometimes referred to as the Type-I tobit model, according to Amemiya's (1984) classification of five types of tobit models.

The model (2.1)-(2.2) is also known as the censored regression model. It is called censored because all the values on the explanatory variables are observed (known) but the dependent variable, y_i, is observed only when the latent variable, y_i^*, is positive. On the other hand, if all explanatory variables were observed (sampled) only for those observations for which y_i^* is positive, the model would be referred to as a truncated regression model. In other words, in a truncated model we have no information regarding $y_i^* \leq 0$. One of the main distinctions between a censored and a truncated regression model is that in a censored regression model one can use the available data to estimate the probability that an observation yields complete data, whereas in a truncated model one cannot. It is clear that there are many similarities between the two models and many results may apply to both cases with some adjustments. However, this study will concentrate solely on the estimation of the censored regression (standard tobit) model defined in (2.1)-(2.2).

Given this model, the main interest is to estimate β and σ^2 on the basis of N observations on the variables y_i and x_i. Estimation of these parameters using ordinary least squares techniques provides estimates which are generally biased and inconsistent. The properties of these estimators are discussed in the next section. Alternatively, a number of consistent estimators have been suggested for the estimation of the parameters of the tobit model; some of these estimators will be considered subsequently in this Chapter.

2.3 Properties of Least Squares Estimators of the Tobit Model

It is well known that the conventional ordinary least squares estimators for the tobit model defined in (2.1)-(2.2) are biased and inconsistent. However, it is important to understand the properties of the least squares estimators and their limitations. In order to do this we first define the following:

Let N_0 be the number of observations for which $y_i = 0$, and N_1 be the number of observations for which $y_i > 0$ such that $N = N_0 + N_1$. Further, we define f_i and F_i to be the density function and the cumulative distribution function of the standard normal random variable, respectively, evaluated at $z_i = x_i'\beta/\sigma$. That is,

$$f_i = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}{z_i}^2} \tag{2.3}$$

and

$$F_i = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt. \tag{2.4}$$

Suppose we consider the observations in which the dependent variable is positive (i.e. $y_i > 0$). That is

$$\begin{aligned} E[y_i|y_i > 0] &= x_i'\beta + E[u_i|y_i > 0] \\ &= x_i'\beta + E[u_i|u_i > -x_i'\beta] \end{aligned} \tag{2.5}$$

Note that if the conditional expectation of the last term in the right hand side of equation (2.5) is zero, then the least squares estimator of β would be unbiased. However, this is not generally the case, implying that the least squares estimator that uses the N_1 positive observations of y_i is biased.

Further, if we assume normality as in the tobit model, equation (2.5) can be shown by straightforward integration to be

$$E[y_i|y_i > 0] = x_i'\beta + \sigma\lambda(z_i), \tag{2.6}$$

where $\lambda(z_i) = f(z_i)/F(z_i)$ which is known as the hazard rate in reliability theory and its inverse is known as Mill's ratio.

As can be seen from equation (2.6), applying least squares omits the term $\sigma\lambda(z_i)$ which is not independent of x_i'. It is this omission that leads to a biased and inconsistent estimator of β. However, the magnitude and direction of the bias is unknown without making further assumptions. Goldberger (1981) assumed that the x_i', excluding the first element which is assumed to be constant, are normally distributed and evaluated the asymptotic bias of the least squares estimator. Specifically, Goldberger (1981) defined (2.1) as

$$y_i^* = \beta_0 + x_i'\beta_1 + u_i \tag{2.7}$$

where x_i' is assumed normally distributed with mean zero and variance Σ and is distributed independently of u_i. Note that the assumption of zero mean does not involve any loss of generality since any non-zero mean can be absorbed by the constant term. Goldberger (1981) then obtained

$$\text{plim } \hat{\beta}_1 = \frac{1-\gamma}{1-\rho^2\gamma}\beta_1 \tag{2.8}$$

where

$$\gamma = \sigma_y^{-1}\lambda(\beta_0/\sigma_y)[\beta_0 + \sigma_y\lambda(\beta_0/\sigma_y)] \quad \text{and} \quad \rho^2 = \sigma_y^{-2}\beta_1'\Sigma\beta_1,$$

where $\sigma_y^2 = \sigma^2 + \beta_1'\Sigma\beta_1$.

Based on this result it can be shown that $0 < \gamma < 1$ and $0 < \rho < 1$, thus (2.8) implies that $\hat{\beta}_1$ shrinks the estimate of β_1 to zero with the degree of shrinkage being uniform for all elements of β_1. However, if x_i' is not normal the result may not hold.

Similarly, Greene (1981a) showed the biasedness of the least squares estimator when applied to the model using all the limit and non-limit observations. In this case we consider the unconditional expectation which yields the equation

$$E[y_i] = F(z_i)(x_i'\beta) + \sigma f(z_i) \tag{2.9}$$

Following the same assumptions as Goldberger (1981), and expressing (2.1) as (2.7), Greene (1981a) obtained the following result:

$$\text{plim } \tilde{\beta}_1 = F(\beta_0/\sigma_y).\beta_1 \tag{2.10}$$

where $\tilde{\beta}_1$ is the least squares estimator of β_1 from the regression of y_i on x_i' using all the observations. The result in (2.10) is an interesting result since it implies that β_1 can be consistently estimated by $(N/N_1)\tilde{\beta}_1$, which is sometimes referred to as the corrected least squares estimator (COLS). A similar consistent estimator can be obtained for β_0. A Monte Carlo study by Flood (1985) who compared Greene's (1981a) COLS estimator and the maximum likelihood estimator of the tobit model indicates that, under normality of the error terms, the COLS estimator is biased even when the sample size increases. However, the bias of the COLS disappears when the exogenous variables are generated from a normal distribution, a situation which is unlikely to be the case in applied research.

In general we conclude that the least squares estimators using both the positive observations on y_i, or all the N observations, do not provide good estimators for the tobit model. Thus, alternative estimators have been devised; they are discussed in the following sections.

2.4 The Maximum Likelihood Estimator (MLE)

The maximum-likelihood procedure can be applied to obtain consistent estimators of the parameters, β and σ^2. Maximum likelihood estimation of the standard tobit model proceeds as follows. For convenience we assume that, without loss of generality, the first N_1 observations contain the non-zero observations and the remaining N_0 observations contain zero observations on y_i. We know that for the observations for which y_i are zero, we have

$$\Pr(y_i = 0) = \Pr(u_i < -x_i'\beta) = 1 - F_i \tag{2.11}$$

since the normal distribution is symmetric. For the observations for which y_i are greater than zero, we have

$$\Pr(y_i > 0).f(y_i|y_i > 0) = F_i \frac{f(y_i - x_i'\beta, \sigma^2)}{F_i} \tag{2.12}$$

$$= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-(y_i - x_i'\beta)^2/2\sigma^2\} \tag{2.13}$$

Note that f_i and F_i are as defined in (2.3)-(2.4) the density and distribution functions of the standard normal distribution, respectively.

Thus the likelihood function is given by

$$L = \prod_0 [1 - F_i] \prod_1 \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-(y_i - x_i'\beta)/2\sigma^2\} \tag{2.14}$$

where the first product, $\prod_0$, is evaluated over the N_0 observations for which $y_i = 0$ and the second product, $\prod_1$, is evaluated over N_1 observations for which $y_i > 0$.

The log-likelihood function is

$$\log L = \sum_0 \log(1 - F_i) + \sum_1 \log \frac{1}{(2\pi\sigma^2)^{1/2}} - \sum_1 \frac{1}{2\sigma^2}(y_i - x_i'\beta)^2 \tag{2.15}$$

where $\sum_0$ is the summation over the N_0 observations for which $y_i = 0$ and $\sum_1$ is the summation over the N_1 observations for which $y_i > 0$.

The first derivatives of $\log L$ for a maximum are

$$\frac{\partial \log L}{\partial \beta} = -\frac{1}{\sigma}\sum_0 \frac{f_i x_i}{1 - F_i} + \frac{1}{\sigma^2}\sum_1 (y_i - x_i'\beta)x_i \tag{2.16}$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^3}\sum_0 \frac{(x_i'\beta)f_i}{1 - F_i} - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_1 (y_i - x_i'\beta)^2 \tag{2.17}$$

and the second derivatives of $\log L$ are given by

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta'} = -\frac{1}{\sigma}\sum_0 \frac{f_i}{(1 - F_i)^2}\left[\frac{f_i}{\sigma} - \frac{1}{\sigma^2}(1 - F_i)x_i'\beta\right] x_i x_i' - \frac{1}{\sigma^2}\sum_1 x_i x_i' \tag{2.18}$$

$$\frac{\partial^2 \log L}{\partial \sigma^2 \partial \beta'} = -\frac{1}{2\sigma^3}\sum_0 \frac{f_i}{(1 - F_i)^2}\left[\frac{1}{\sigma^2}(1 - F_i)(x_i'\beta)^2 - (1 - F_i) - \frac{x_i'\beta f_i}{\sigma}\right] x_i$$
$$-\frac{1}{\sigma^4}\sum_1 (y_i - x_i'\beta)x_i \tag{2.19}$$

$$\frac{\partial^2 \log L}{\partial (\sigma^2)^2} = \frac{1}{4\sigma^5}\sum_0 \frac{f_i}{(1 - F_i)^2}\left[\frac{1}{\sigma^2}(1 - F_i)(x_i'\beta)^3 - 3(1 - F_i)(x_i'\beta)\right.$$
$$\left. - \frac{(x_i'\beta)^2 f_i}{\sigma}\right] + \frac{N_1}{2\sigma^4} - \frac{1}{\sigma^6}\sum_1 (y_i - x_i'\beta)^2. \tag{2.20}$$

where the following were used to obtain the above results,

$$\frac{\partial F_i}{\partial \beta} = \frac{1}{\sigma} f_i x_i,$$

$$\frac{\partial F_i}{\partial \sigma^2} = -\frac{1}{2\sigma^3} x_i'\beta f_i,$$

$$\frac{\partial f_i}{\partial \beta} = -\frac{1}{\sigma^3} x_i'\beta f_i x_i,$$

$$\frac{\partial f_i}{\partial \sigma^2} = \frac{(x_i'\beta)^2 - \sigma^2}{2\sigma^5} f_i.$$

The maximum-likelihood estimators of the parameters of the tobit model are defined as a solution to the equations obtained by equating the partial derivatives (2.16) and (2.17) to zero. These equations are nonlinear in the parameters and hence must be solved using iterative methods. Amemiya (1973) proved that the tobit MLE estimators are strongly consistent and asymptotically normal with the asymptotic variance-covariance given by the inverse of the information matrix, defined by

$$\begin{aligned} V(\theta) &= \left\{ E\left[-\frac{\partial^2 \log L}{\partial\theta\partial\theta'} \right] \right\}^{-1} \\ &= \begin{bmatrix} \sum_{i=1}^{N} a_i x_i x_i' & \sum_{i=1}^{N} b_i x_i \\ \sum_{i=1}^{N} b_i x_i' & \sum_{i=1}^{N} c_i \end{bmatrix}^{-1} \end{aligned} \qquad (2.21)$$

where

$$\theta = (\beta', \sigma^2)',$$

$$a_i = -\frac{1}{\sigma^2}\left(z_i f_i - \frac{f_i^2}{1-F_i} - F_i\right),$$

$$b_i = \frac{1}{2\sigma^3}\left(z_i^2 f_i + f_i - \frac{z_i f_i^2}{1-F_i}\right),$$

$$c_i = -\frac{1}{4\sigma^4}\left(z_i^3 f_i + z_i f_i - \frac{z_i^2 f_i^2}{1-F_i} - 2F_i\right).$$

Amemiya (1973) showed that the tobit likelihood function is not globally concave with respect to the original parameters, β and σ^2. However, Olsen (1978) proved the global concavity of $\log L$ based on the reparameterization $\alpha = \beta/\sigma$ and $h = 1/\sigma$. This implies that a standard iterative method such as Newthon-Raphson or the method of scoring always converges to the global maximum of $\log L$. In terms of the new parameters, $\log L$ is written as

$$\log L = \sum_0 \log[1 - F(x_i'\alpha)] + N_1 \log h - \frac{1}{2}\sum_1 (hy_i - x_i\alpha)^2, \tag{2.22}$$

The normal equations of the reparameterized version are

$$\frac{\partial \log L}{\partial \alpha} = -\sum_0 \frac{f(-x_i'\alpha)}{F(-x_i'\alpha)} x_i + \sum_1 (hy_i - x_i'\alpha)x_i \tag{2.23}$$

$$\frac{\partial \log L}{\partial h} = \frac{N_1}{h} - \sum_1 (hy_i - x_i'\alpha)y_i \tag{2.24}$$

The second derivatives of the likelihood function based on the new parameters are less cumbersome than the corresponding expressions (2.18) through (2.20). These are

$$\frac{\partial^2 \log L}{\partial \alpha \partial \alpha'} = \sum_0 \frac{f(-x_i'\alpha)}{F(-x_i'\alpha)}\left(x_i'\alpha - \frac{f(-x_i'\alpha)}{F(-x_i'\alpha)}\right) x_i x_i' - \sum_1 x_i x_i' \tag{2.25}$$

$$\frac{\partial^2 \log L}{\partial\alpha\partial h} = \sum_1 y_i x_i \tag{2.26}$$

$$\frac{\partial^2 \log L}{\partial h^2} = -\frac{N_1}{h^2} - \sum_1 y_i^2 \tag{2.27}$$

From which Olsen (1978) obtained the matrix of second derivatives as

$$\begin{bmatrix} \frac{\partial^2 \log L}{\partial\alpha\partial\alpha'} & \frac{\partial^2 \log L}{\partial\alpha\partial h} \\ \frac{\partial^2 \log L}{\partial h\partial\alpha'} & \frac{\partial^2 \log L}{\partial h^2} \end{bmatrix} = \begin{bmatrix} \sum_0 \frac{f_i}{1-F_i}\left(x_i'\alpha - \frac{f_i}{1-F_i}\right) x_i x_i' & 0 \\ 0 & -\frac{N_1}{h^2} \end{bmatrix} + \begin{bmatrix} -\sum_1 x_i x_i' & \sum_1 x_i y_i \\ \sum_1 y_i x_i' & -\sum_1 y_i^2 \end{bmatrix} \tag{2.28}$$

Note that in order for $\log L$ to have a global maximum the matrix given by (2.28) must be negative definite. Since the expression $x_i'\alpha - [1 - F(x_i'\alpha)]^{-1} f(x_i'\alpha) < 0$, the right hand side of (2.28) is the sum of two negative definite matrices and hence is a negative definite matrix. The implication of this result is that the likelihood function of the tobit model has a single maximum. Recently, Greene (1990) investigated the possibility of multiple roots of the tobit loglikelihood function based on the original parameters and showed that the problem of multiple roots in the tobit model is less obvious than suggested in earlier literature. Further, because of the invariance property of maximum likelihood estimators one can obtain a unique solution in terms of the original parameters of the tobit model.

To obtain the maximum likelihood estimates standard iterative procedures such as Newton-Raphson or the method of scoring can be used. The latter procedure uses the information matrix in place of the negative of the Hessian. However, the well known Newton-Raphson requires the second partial derivatives. If we define $\hat{\theta}_2$ to be the second-round estimate, given an initial estimate $\hat{\theta}_1$, the Newton-Raphson iteration is defined by

$$\hat{\theta}_2 = \hat{\theta}_1 - \left[\left. \frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'} \right|_{\hat{\theta}_1} \right]^{-1} \left. \frac{\partial \log L(\theta)}{\partial\theta} \right|_{\hat{\theta}_1} \tag{2.29}$$

Once $\hat{\theta}_2$ is obtained the iteration may continue for $\hat{\theta}_3$ and so on, until the iteration converges to a desired degree of precision. Tobin (1958) suggested the same procedure and in order to speed up convergence of the estimation process, he proposed an initial estimator based on a linear approximation of the reciprocal of the Mill's ratio. Tobin's (1958) procedure was as follows: Equate the right hand side of equation (2.16) to 0 to obtain

$$-\sigma \sum_0 \frac{f_i x_i}{1 - F_i} + \sum_1 (y_i - x_i'\beta)x_i = 0 \tag{2.30}$$

Premultiply (2.30) by $\beta'/2\sigma^4$ and add it to the equation obtained by equating (2.17) to 0. The result is

$$\sigma^2 = \frac{1}{N_1} \sum_1 (y_i - x_i'\beta)y_i \tag{2.31}$$

Equation (2.30) involves a non-trivial function $f_i/(1 - F_i)$ which is the inverse of the Mill's ratio. Tobin (1958) approximates $f_i/(1 - F_i)$ by a linear function of the form $a + b(x_i'\beta/\sigma)$ and substitutes in the left hand side of equation (2.30) , to obtain

$$-\sigma \sum_0 [a + b(x_i'\beta/\sigma)x_i] + \sum_1 (y_i - x_i'\beta)x_i = 0 \tag{2.32}$$

Then, solving for β from equation (2.32) and inserting it in equation (2.31) yields a quadratic function of σ. Once an estimate of σ is obtained β can be estimated linearly from equation (2.32). However, Amemiya (1973) showed that Tobin's (1958) initial estimator is inconsistent, and he proposed an alternative consistent estimator. He also showed that the second-round estimator based on his initial estimator is consistent and asymptotically normal. Amemiya's (1973) estimator was simple:

Consider equation (2.6) which is given by

$$E(y_i|y_i > 0) = x_i'\beta + \sigma\lambda(x_i'\beta/\sigma) \tag{2.33}$$

and

$$E(y_i^2|y_i > 0) = (x_i'\beta)^2 + \sigma(x_i'\beta)\lambda(x_i'\alpha) + \sigma^2 \tag{2.34}$$

in which from (2.33) and (2.34), Amemiya (1973) obtained

$$E(y_i^2|y_i > 0) = (x_i'\beta)E(y_i|y_i > 0) + \sigma^2 \tag{2.35}$$

Equation (2.35) can be written as

$$y_i^2 = y_i x_i'\beta + \sigma^2 + \zeta_i, \text{ for } i \text{ such that } y_i > 0 \tag{2.36}$$

where $E(\zeta_i|y_i > 0) = 0$.

Equation (2.36) is a linear regression model containing an endogenous explanatory variable. Thus, although it does not contain terms involving $\lambda(.)$, applying OLS directly to (2.36) leads to an inconsistent estimator because of the violation of the assumption of independence between the explanatory variable and the disturbance term of the model. Amemiya (1973) then proposed the estimation of equation (2.36) using an instrumental variable such that the explanatory variable $(\hat{y}_i x_i, 1)$ is used in place of $(y_i x_i, 1)$, where $\hat{y}_i$ is the predictor of y_i obtained by regressing y_i on x_i using N_1 observations. Amemiya's (1973) initial estimator is important especially in more complicated tobit models. However, based on a simulation study of a single replication of sample sizes of 1000 and 5000, Wales and Woodland (1980) indicated that it is rather inefficient.

Fair (1977) suggested an alternative procedure for obtaining the maximum likelihood estimators of β and σ^2 based on the normal equations (2.16) and (2.17). His procedure was as follows:

For the purpose of tractability let us rewrite equation (2.31)

$$\sigma^2 = \frac{1}{N_1}\sum_1 (y_i - x_i'\beta)y_i \tag{2.37}$$

Further, multiply equation (2.16) throughout by σ and equate to zero, to obtain

$$-\sum_0 \frac{f_i x_i}{1-F_i} + \frac{1}{\sigma}\sum_1 (y_i - x_i'\beta)x_i = 0 \tag{2.38}$$

Finally, solving for β from equation (2.38) yields

$$\beta = \left(\sum_1 x_i'x_i\right)^{-1}\sum_1 x_i'y_i - \sigma\left(\sum_1 x_i'x_i\right)^{-1}\sum_0 x_i'\gamma_i \tag{2.39}$$

where $\gamma_i = \sigma f_i/(1-F_i)$.

In matrix notation this can be written as

$$\begin{aligned}\beta &= (X'X)^{-1}X'Y - \sigma(X'X)^{-1}X_0'\gamma_0 \\ &= \beta_R^{LS} - \sigma(X'X)^{-1}X_0'\gamma_0 \end{aligned} \tag{2.40}$$

where Y is an $(N_1\text{x}1)$ vector of positive observations on the endogenous variable and X is the corresponding $(N_1\text{x}k)$ matrix of explanatory variables, X_0 is an $(N_0\text{x}k)$ matrix of explanatory variables corresponding to the zero observations of Y, γ_0 is an $(N_0\text{x}1)$ vector of the γ_i's for $y_i = 0$ and β_R^{LS} is the ordinary least squares estimate of β for the non-zero observations.

The expression given by (2.40) shows the relationship between the ordinary least squares estimator based on the non-zero observations and the tobit ML estimator explicitly.

Fair (1977) suggested the following procedure for computing the tobit ML estimates.

1. Compute β_R^{LS} and $(X'X)^{-1}X_0'$.
2. Choose a value of β, say, $\beta^{(1)}$, and compute σ^2 from (2.37). If this value of σ^2 is less than or equal to zero, take for the value of σ^2 some small positive number. Let $\sigma^{(1)}$ denote the square root of the chosen value of σ^2.
3. Compute the vector γ_0 using $\beta^{(1)}$ and $\sigma^{(1)}$. Denote this vector as $\gamma_0^{(1)}$. (A standard FORTRAN function is available to compute the distribution function F_i.)
4. Compute β from (2.40) using $\sigma^{(1)}$ and $\gamma_0^{(1)}$. Denote this value as $\tilde{\beta}^{(1)}$. Let

   $$\beta^{(2)} = \beta^{(1)} + \delta(\tilde{\beta}^{(1)} - \beta^{(1)}), \text{ where } 0 < \delta \leq 1.$$

   where δ is a damping factor which is useful in this sort of procedure.
5. Using $\beta^{(2)}$ as the new value of β, go to step (2) and repeat the process until the iteration converges.

Fair (1977) provided an empirical example in which two samples of size 601 (of which 150 are non-zero and 451 are zero observations) and 6366 (of which 2053 are non-zero and 4313 are zero observations) were involved. He analyzed his data using a program called LIMDEP, which is fairly widely used, and found that both Newton's and his method converged to the same answer. He also noted that although

Newton's method required fewer iterations to converge, the computer time needed in his procedure was much less than the iteration time needed using Newton's procedure. However, the convergence and speed of Fair's procedure is sensitive to changes in the starting value of λ and some times even convergence may not be possible. If, however, it converges, then the variance-covariance matrix can be obtained using (2.21).

Given the assumptions of normality and homoscedasticity, as assumed in the tobit model, the maximum likelihood estimators are consistent and asymptotically efficient. Further, assuming that the residuals are independent the MLE remains consistent but not efficient under serial correlation (see Robinson (1982)). However, the properties are sensitive to non-normality and heteroscedasticity of the errors. If the assumption of normality of the disturbances is violated, as Goldberger (1980) and Arabmazar and Schmidt (1982) indicate, the maximum likelihood estimators may lead to inconsistent estimates. Further, as shown by Hurd (1979), Maddala and Nelson (1975) and Arabmazar and Schmidt (1981), heteroscedasticity of the disturbances can cause inconsistency of the parameter estimates, even when the shape of the error density is correctly specified. Arabmazar and Schmidt (1981) also noted that the asymptotic biases of the censored regression model are not as large as those obtained by Hurd (1979) for the truncated model.

In general, the maximum likelihood estimators are not robust to the assumptions of the model. This is in contrast to the classical regression model which is generally consistent under a wide variety of conditions. Further, except for the simplest cases, the derivatives of the loglikelihood function of the tobit model are very complicated and hence the computational cost of the maximum likelihood estimators can be very high.

2.5 Heckman's Two-step Estimator (H2S)

Heckman (1976, 1979) proposed an alternative estimator which yields consistent estimates of the parameters based on a two-step procedure. Heckman's two-step estimator (H2S) was originally suggested for a system of two equations, but can be used in a single equation with some adjustments. Heckman's paper appears to generalise earlier studies in the economics literature which include those of Gronau (1974) and Lewis (1974). The two-step estimator uses the observations for which $y_i > 0$ and proceeds as follows.

Consider the conditional expectation for which $y_i > 0$. That is, rewrite equation (2.6) as

$$E[y_i|y_i > 0] = x_i'\beta + \sigma\lambda(x_i'\alpha), \tag{2.41}$$

where $\lambda(x_i'\alpha) = f(x_i'\alpha)/F(x_i'\alpha)$ and $\alpha = \beta/\sigma$ are as defined before.

Equation (2.41) can be written as

$$y_i = x_i'\beta + \sigma\lambda(x_i'\alpha) + \varepsilon_i \tag{2.42}$$

where $\varepsilon_i = y_i - E(y_i|y_i > 0)$, $E(\varepsilon_i) = 0$, $E(\varepsilon_i\varepsilon_j) = 0$, $i \neq j$ and the variance of ε_i is given by

$$V(\varepsilon_i) = \sigma^2 - \sigma^2 x_i'\alpha\lambda(x_i'\alpha) - \sigma^2\lambda(x_i'\alpha)^2 \tag{2.43}$$

Note that, as discussed earlier in this Chapter, α and hence $\lambda(.)$ are unknown in equation (2.42). Thus, ordinary least squares estimates from regressing y_i on x_i' are biased estimates. Heckman (1976, 1979) treated the bias as a result of an omitted variable and he suggested a two-step procedure which involves the estimation of the omitted variable using the probit maximum likelihood estimator in the first step, and then the application of ordinary least squares in the second step after replacing the omitted variables by their consistent estimates. Specifically, Heckman's two-step estimator proceeds as follows:

Step 1. Estimate α by the probit maximum likelihood estimator, say $\hat{\alpha}$. The likelihood function is given by

$$L = \prod_0 [1 - F(x_i'\alpha)] \prod_1 F(x_i'\alpha) \tag{2.44}$$

where the products $\prod_0$ and $\prod_1$, as defined before, are evaluated over the N_0 and N_1 observations, respectively.

Then, the ratio $\alpha = \beta/\sigma$ can be easily estimated by maximising (2.44) using standard iterative procedures. Note that one can only estimate α, not β and σ separately. Once $\hat{\alpha}$ is estimated then $\hat{\lambda}(x_i'\hat{\alpha})$ can be obtained by straightforward substitution. These values are consistent since the probit maximum likelihood estimator, $\hat{\alpha}$, is consistent. Further, it can be shown that $\hat{\alpha}$ is asymptotically normal and the asymptotic variance-covariance matrix is given by [see Amemiya (1978, p.1196)]

$$V(\hat{\alpha}) = (\underset{\sim}{X}'D_1\underset{\sim}{X})^{-1}. \tag{2.45}$$

where $\underset{\sim}{X}$ is an $(N\text{x}k)$ matrix of all observations and D_1 is an $(N\text{x}N)$ diagonal matrix whose diagonal elements are given by $F(x_i'\alpha)^{-1}[1 - F(x_i'\alpha)]^{-1}f(x_i'\alpha)^2$.

Step 2. Replace $\lambda(x_i'\alpha)$ by $\hat{\lambda}(x_i'\hat{\alpha})$ in equation (2.42) and then regress y_i on x_i' and $\hat{\lambda}(x_i'\hat{\alpha})$ to obtain consistent estimates of β and σ, based on the observations for which $y_i > 0$, i.e., using only the N_1 observations.

To discuss Heckman's two-step estimator further, equation (2.42) can be written as

$$y_i = x_i'\beta + \sigma[\lambda(x_i'\alpha) - \hat{\lambda}(x_i'\hat{\alpha})] + \sigma\hat{\lambda}(x_i'\hat{\alpha}) + \varepsilon_i \tag{2.46}$$

which is equivalent to

$$y_i = x_i'\beta + \sigma\hat{\lambda}(x_i'\hat{\alpha}) + \varepsilon_i + \eta_i \tag{2.47}$$

where $\eta_i = \sigma[\lambda(x_i'\alpha) - \hat{\lambda}(x_i'\hat{\alpha})]$.

Further, using matrix notation, equation (2.47) can be presented as

$$Y = X\beta + \sigma\hat{\lambda} + (\varepsilon + \eta) \tag{2.48}$$

where Y is an $(N_1\text{x}1)$ vector of the non-zero observations on the dependent variable, X is a $(N_1\text{x}k)$ matrix of explanatory variables corresponding to Y, $\hat{\lambda}$ is an $(N_1\text{x}1)$ vector whose elements are $\hat{\lambda}(x_i'\hat{\alpha})$, and ε and η are vectors of N_1 elements of ε_i and η_i, respectively.

Further, equation (2.48) can be written as

$$Y = \hat{Z}\gamma + (\varepsilon + \eta), \tag{2.49}$$

where $\hat{Z} = (X, \hat{\lambda})$ and $\gamma = (\beta', \sigma)'$.

Thus, Heckman's 2-step estimator of γ is defined by

$$\hat{\gamma} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'Y. \tag{2.50}$$

Heckman (1976, 1979) showed that $\hat{\gamma}$ is consistent and asymptotically normal with mean γ and the asymptotic variance-covariance matrix given by [see also Amemiya (1984), p.13]

$$V_{\hat{\gamma}} = \sigma^2(Z'Z)^{-1}Z'[\Sigma + (I - \Sigma)X(\underset{\sim}{X}'D_1\underset{\sim}{X})^{-1}X'(I - \Sigma)]Z(Z'Z)^{-1} \tag{2.51}$$

where $\sigma^2\Sigma = E(\varepsilon\varepsilon')$ is an $(N_1 \text{x} N_1)$ diagonal matrix whose diagonal elements are $\text{var}(\varepsilon_i)$ as defined in (2.43) and the matrices X and D_1 are defined after (2.45).

Note that the expression given by (2.51) may be estimated consistently by replacing the unknown parameters by their consistent estimates or by $(Z'Z)^{-1}Z'AZ(Z'Z)^{-1}$, where A is a diagonal matrix with the i^{th} diagonal element given by $[y_i - x_i'\hat{\beta} - \hat{\sigma}\lambda(x_i'\hat{\alpha})]^2$, which follows the idea of White (1980b).

It should be noted that the expression in the square brackets in the right hand side of equation (2.51) arose because λ is unknown and had to be estimated. However, if λ were known, equation (2.42) could be estimated directly using least squares estimates and the exact variance-covariance matrix would be $\sigma^2(Z'Z)^{-1}Z'\Sigma Z(Z'Z)^{-1}$.

The H2S estimator is consistent and has been used widely in applied research because of its simplicity. However, it is inefficient. Wales and Woodland (1980) and Nelson (1984) presented some evidence on the inefficiency of the H2S estimator relative to the maximum likelihood estimator. One of the main reasons for the inefficiency is the existence of strong multicollinearity between the explanatory variables and the estimated hazard function, $\hat{\lambda}(x_i'\hat{\alpha})$, in the model. A recent paper by Nawata (1993) shows that there almost always exists a high (negative) correlation between the explanatory variables, X's, and the estimated hazard function, $\hat{\lambda}(x_i'\hat{\alpha})$, which is the main cause of the inefficiency of the H2S estimates. This is an important line of discussion because the improvements that are suggested later in this study are basically designed to avoid the multicollinearity problem; they will be discussed thoroughly in the next Chapter.

Below, we consider alternative techniques of estimating equation (2.42); specifically, we discuss the estimation of the model using weighted least squares, nonlinear least squares and nonlinear weighted least squares based on the observations for which $y_i > 0$.

2.6 Weighted Heckman's two-step Estimator

It should be noted that equation (2.42) represents a regression model with a heteroscedastic error variance, implying that the Heckman two-step estimates are not efficient. Heckman (1976, 1979) suggested that asymptotically more efficient estimates can be obtained using weighted least squares in the second step of the procedure with the weights given by (2.43). This estimator is referred to as the weighted Heckman's two-step (WH2S) estimator.

Let the resulting estimator using the N_1 observations be denoted by $\hat{\gamma}_W$. It is consistent and asymptotically normal with mean γ and the asymptotic variance-covariance matrix given by [see Amemiya (1984), pp. 12-16]

$$V_{\hat{\gamma}_W} = \sigma^2\{Z'[\Sigma + (I-\Sigma)X(\underset{\sim}{X}'D_1\underset{\sim}{X})^{-1}X'(I-\Sigma)]^{-1}Z\}^{-1} \tag{2.52}$$

where $\sigma^2\Sigma$, Z, D_1 are as defined in (2.43), (2.49) and (2.45), respectively.

Note that the weighted Heckman's two-step estimator provides asymptotically more efficient estimates than its counterpart, the Heckman two-step estimator. However, neither the H2S estimator nor the WH2S avoid the multicollinearity problem discussed in the preceding Section. Thus, given that multicollinearity is a serious problem in the model, the gains in efficiency from the WH2S relative to the H2S estimator may not be that large.

Further, the expression for the weights in (2.43) involves a non-trivial element $\lambda(.)$

which in turn involves the estimation of the density and cumulative distribution functions of the normal random variable. Thus, the estimates of the weights themselves may not be consistent if the normality assumption does not hold in practice.

2.7 Nonlinear Estimation based on Conditional Expectation

It is clear that the expression given by (2.42) can be viewed as a nonlinear problem. Thus, one can estimate the model using nonlinear least squares based on the positive (NLSP) observations on y_i. That is, β and σ can be estimated simultaneously by minimizing the sum of squares

$$SS(\beta', \sigma) = \sum_{i=1}^{N_1} [y_i - x_i'\beta - \sigma\lambda(x_i'\beta/\sigma)]^2 \tag{2.53}$$

with respect to the parameters, β and σ.

It should be noted that even if $x_i'\beta$ is linear in β it involves a nonlinear estimation problem in view of the dependence of $\lambda(.)$ on β and σ. The properties of nonlinear least squares are well established in the econometric literature [see for example, Jennrich (1969), Amemiya (1983a)]. The asymptotic properties of nonlinear least squares estimators are generally obtained in such a way that results for the linear regression model hold asymptotically for a nonlinear regression model by treating the partial derivatives of the nonlinear regression as the regression matrix.

Let $\hat{\gamma}_N$ be the nonlinear least squares estimator obtained by minimizing equation (2.53) based on the N_1 observations. Hartley (1976) proved the asymptotic properties and showed that $\hat{\gamma}_N$ is consistent and asymptotically normally distributed with mean

γ and its asymptotic variance-covariance matrix given by [see also Amemiya (1984, 1985)]

$$V_{\hat{\gamma}_N} = \sigma^2(S'S)S'\Sigma S(S'S)^{-1} \tag{2.54}$$

where

$$S = (\Sigma X, D_2\lambda), \tag{2.55}$$

where D_2 is an $(N_1 \text{x} N_1)$ diagonal matrix whose i^{th} element is given by $[1 + (x_i'\alpha)^2 + (x_i'\alpha)\lambda(x_i'\alpha)]$ and $\sigma^2\Sigma$ is as defined in (2.43).

Alternatively, the parameters in (2.42) can be estimated using nonlinear weighted least squares with the weights given by (2.43). That is to say, the parameters β and σ can be estimated simultaneously by minimizing with respect to the parameters the weighted sum of squares of the residuals which is given by

$$WSS(\beta', \sigma) = \sum_{i=1}^{N_1} \left(\frac{[y_i - x_i'\beta - \sigma\lambda(x_i'\alpha)]^2}{1 - x_i'\alpha\lambda(x_i'\alpha) - \lambda(x_i'\alpha)^2} \right) \tag{2.56}$$

Similarly, if we let $\hat{\gamma}_{NW}$ to be the nonlinear weighted least squares estimator based on the N_1 observations, it can be shown that it is asymptotically normal with mean γ and the asymptotic variance-covariance matrix is given as [see Hartley (1976), Amemiya (1984, 1985)]

$$V_{\hat{\gamma}_{NW}} = \sigma^2(S'\Sigma^{-1}S)^{-1} \tag{2.57}$$

where S is given by (2.55) above.

The minimization of these nonlinear functions involves partial derivatives which are not in a closed form. Thus the solutions of the normal equations are obtained using standard iterative procedures such as the Newton-Raphson method. In practice, the nonlinear estimators of the tobit model have not been used, mainly because they are not as easy as the MLE. Convergence may not be guaranteed in the nonlinear estimation procedure.

Note that the above estimators are based on the conditional expectation of the model given by (2.42). However, one may also obtain similar estimators using the unconditional expectation of the model given by (2.9). These estimators are discussed in the section below.

2.8 Two-step Estimators based on Unconditional Expectation

In the preceding section we discussed estimators of the model which are developed using the conditional expectation of the model and which use only those observations for which $y_i > 0$. Wales and Woodland (1980) suggested that a similar procedure may be applied using the unconditional expectation which uses all the observations. That is, including the observations for which $y_i > 0$ as well as those observations for which $y_i = 0$. Further discussions of the estimators based on all observations, and their asymptotic distributions, include those of Stapleton and Young (1984) and Amemiya (1984, 1985). The estimation procedure using all observations proceeds as follows:

Consider the unconditional expectation of y_i, given in (2.9) as

$$E[y_i] = F(x_i'\alpha).(x_i'\beta) + \sigma f(x_i'\alpha) \tag{2.58}$$

Equation (2.58) can be written as

$$y_i = F(x_i'\alpha)[(x_i'\beta) + \sigma\lambda(x_i'\alpha)] + \delta_i \tag{2.59}$$

where $\delta_i = y_i - E[y_i]$ such that $E[\delta_i] = 0,\ E[\delta_i\delta_j] = 0,\ i \neq j,$ and

$$V(\delta_i) = \sigma^2 F(x_i'\alpha)\{(x_i'\alpha)^2 + (x_i'\alpha)\lambda(x_i'\alpha) + 1 - F(x_i'\alpha)[x_i'\alpha + \lambda(x_i'\alpha)]^2\} \qquad (2.60)$$

Further, equation (2.59) can be written as

$$y_i = F(x_i'\hat{\alpha})[x_i'\beta + \sigma\hat{\lambda}(x_i'\hat{\alpha})] + [F(x_i'\alpha) - F(x_i'\hat{\alpha})]x_i'\beta + \sigma[f(x_i'\alpha) - f(x_i'\hat{\alpha})] + \delta_i \quad (2.61)$$

which is equivalent to

$$y_i = F(x_i'\hat{\alpha})[x_i'\beta + \sigma\hat{\lambda}(x_i'\hat{\alpha})] + \delta_i + \xi_i \qquad (2.62)$$

where

$$\xi_i = [F(x_i'\alpha) - F(x_i'\hat{\alpha})]x_i'\beta + \sigma[f(x_i'\alpha) - f(x_i'\hat{\alpha})]$$

In matrix notation (2.62) can be expressed by

$$Y = \hat{D}\hat{\underline{Z}}\gamma + \delta + \xi \qquad (2.63)$$

where $\hat{D}$ is an $(N\text{x}N)$ diagonal matrix whose elements are $F(x_i'\hat{\alpha})$, $\hat{\underline{Z}} = (\underline{X}, \hat{\lambda})$ is an $(N\text{x}\ k+1)$ matrix, δ and ξ are vectors of order N whose elements are δ_i and ξ_i, respectively, and $\underline{X}$ and γ are as defined in (2.45) and (2.49), respectively.

Note that the models given by (2.59) and (2.63) have the same form as the previous models given by (2.42) and (2.49), respectively. One noticeable difference between the two sets of models is that while models (2.42) and (2.49) were derived using N_1 observations, the other two are based on all observations, N. Next, we consider the models (2.59) and/or (2.63) and apply estimators which are analogous to those discussed in the preceding section, namely, the least squares, weighted least squares and nonlinear least squares estimators.

2.8.1 Heckman's two-step Estimator based on the Unconditional Expectation (H2SU) of the Model

One way of estimating model (2.63) is to apply ordinary least squares in the second step of the procedure. This estimator is analogous to the H2S estimator discussed in Section 2.4 and was first suggested by Wales and Woodland (1980). It should be clear that the first step still involves the estimation of α using the probit MLE. Once α is estimated then one can use the method of least squares to estimate the parameters in (2.63).

Let $\tilde{\gamma}$ be the H2SU estimator of γ based on all observations. Then, from (2.63), it is defined as

$$\tilde{\gamma} = (\hat{\underline{Z}}'\hat{D}^2\hat{\underline{Z}})^{-1}\hat{\underline{Z}}'\hat{D}y \tag{2.64}$$

where $\hat{D}$ is defined in (2.63).

Stapleton and Young (1984) derived the asymptotic properties of $\tilde{\gamma}$. They proved that $\tilde{\gamma}$ is consistent and asymptotically normally distributed with mean γ and the asymptotic variance-covariance matrix is given by [see also Amemiya (1984), pp. 14-15].

$$V_{\tilde{\gamma}} = \sigma^2(\underline{Z}'D^2\underline{Z})^{-1}Z'D^2\Omega\underline{Z}(\underline{Z}'D^2\underline{Z})^{-1} \tag{2.65}$$

where $\sigma^2\Omega$ is an $(N\mathrm{x}N)$ diagonal matrix whose i^{th} elements are $\mathrm{Var}(\delta_i)$ given by (2.60).

Again, the model given by (2.59) or (2.63) has a heteroscedastic error variance implying that more efficient estimates can be obtained using weighted least squares in the second step of the procedure, with the weights given by (2.60).

Let the weighted Heckman's two-step estimator based on the unconditional expectation (WH2SU) of the model be denoted by $\tilde{\gamma}_W$. Then $\tilde{\gamma}_W$ is consistent and asymptotically normal with the asymptotic variance-covariance matrix given by

$$V_{\tilde{\gamma}_W} = \sigma^2(\underline{Z}'D^2\Omega^{-1}\underline{Z})^{-1} \tag{2.66}$$

where $\sigma^2\Omega$ is as defined in (2.65) above.

These estimators have not been used in applied research because they are inefficient relative to the maximum likelihood estimator under the assumptions of the model. It is also clear that, similar to the H2S and WH2S estimators discussed above, these estimators are likely to suffer from multicollinearity problems. Nevertheless, it is important to examine their relative performance under various conditions.

2.8.2 Nonlinear Estimation based on the Unconditional Expectation

Clearly, the model given by (2.59) can be treated as a nonlinear problem in the parameters. Hence, nonlinear least squares and nonlinear weighted least squares estimators can be applied to estimate the parameters of the model.

The nonlinear least squares estimator, denoted by $\tilde{\gamma}_N$, of γ is obtained by minimizing the sum of squares

$$SSU(\beta', \sigma) = \sum_{i=1}^{N}[y_i - F(x_i'\beta/\sigma).x_i'\beta - \sigma\, f(x_i'\beta/\sigma)]^2 \tag{2.67}$$

with respect to the parameters, β and σ.

Alternatively, one can estimate the weighted nonlinear least squares estimator,

$\tilde{\gamma}_{NW}$, by minimizing the following weighted sum of squares with respect to the parameters. That is,

$$SSW(\beta', \sigma) = \sum_{i=1}^{N} \left(\frac{[y_i - F(x_i'\beta/\sigma).x_i'\beta - \sigma\, f(x_i'\beta/\sigma)]^2}{w_i} \right) \tag{2.68}$$

where the weights=w_i are given by (2.60).

These estimators are consistent; however, since δ_i is not normal the nonlinear procedure is not maximum likelihood. Further, they are not computationally easier than the MLE. The asymptotic properties of these estimators were provided by Stapleton and Young (1984) following White's (1980c) results for nonlinear least squares estimators in the presence of heteroscedasticity. Interestingly, Stapleton and Young (1984) showed that the nonlinear least squares estimator, $\tilde{\gamma}_N$, has the same asymptotic distribution as $\tilde{\gamma}$, the Heckman's two-step estimator based on all observations. Similarly, the weighted nonlinear least squares estimator $\tilde{\gamma}_{NW}$, has the same asymptotic distribution as $\tilde{\gamma}_W$ [see also Amemiya (1984, 1985)].

Finally, given the estimators of the tobit model discussed in the Sections above, it is worth noting some of the important points:

The two-step estimators, $\hat{\gamma}$ and $\tilde{\gamma}$ cannot be ranked on the basis of their asymptotic covariance matrices. This is because the difference between the matrices given by (2.51) and (2.65) is generally neither positive nor negative definite. This implies that a preference for one of the two estimators depends on parameter values.

Similarly, one cannot make definite comparisons between the covariance matrices given by (2.51) and (2.54), or between the covariance matrices given by

(2.52) and (2.57). Therefore, the choice between the corresponding estimators $\hat{\gamma}$ and $\hat{\gamma}_N$ and $\hat{\gamma}_W$ and $\hat{\gamma}_{NW}$, respectively, depends on the empirical values.

Further, the estimators $\tilde{\gamma}_N$ and $\tilde{\gamma}_{NW}$ have the same asymptotic distributions as $\tilde{\gamma}$ and $\tilde{\gamma}_W$, respectively. Thus, it would be interesting to see their relative performance in finite samples. In general, little is known about the relative performances of all these estimators in finite samples under various specifications of the model, a gap which this study attempts to fill.

The following Sections of this Chapter provide some highlights of other estimators; semi-parametric, bounded-influence and Bayes estimators of the standard tobit model are considered.

2.9 Highlights of Other Estimators

The literature on tobit models is quite extensive. Thus, it is difficult in this study to provide an exhaustive review of all the available estimators and their properties. It is also important to recall that one of the main objectives of this study is to investigate the small sample properties of the estimators that could be used in applied research with reasonable computational skill and resources. Thus the computational ease and availability of statistical/econometric packages needs to be taken into consideration.

On the other hand, it is important to provide some highlights of the recent developments and citations of the relevant literature. Some of these developments include the semi-parametric estimators, the bounded-influence estimators and the Bayesian estimation of the tobit model. A short discussion of these estimators follows.

2.9.1 Semi-Parametric Estimators of the Model

In a standard linear regression model, the least squares estimators are unbiased and consistent for a wide class of distributions of the disturbances. However, the situation is quite different for tobit models. That is, the assumption of normality of the disturbances, which is a common feature for both the MLE and Heckman's two-step estimators, is essential for the proofs of consistency. In general, these proofs are sensitive to violations of the assumptions of the model. This situation led to the development of estimators which are robust (or less sensitive) to the functional form of the distribution of the disturbances. Contributions to the development of the semi-parametric estimation methods include those of Manski (1975, 1985) and Cosslett (1983) for qualitative response models, and those of Powell (1984, 1986a, 1986b), Duncan (1986) and Ruud (1986) for tobit models.

Specifically, Powell (1984) proposed the least absolute deviations (LAD) estimator which is obtained by minimizing with respect to β the function

$$S_N(\beta) = (1/N)\sum_{i=1}^{N}|y_i - \max\{0, x_i'\beta\}| \qquad (2.69)$$

Powell (1984) proved that the LAD estimator is consistent under the assumption that the conditional error distribution has median zero. The estimator is a non parametric estimator due to the fact that the median (which is the solution to $S_N(\beta)$ in this case) of the censored variable does not depend, unlike the mean, on the functional form of the density function of the disturbances.

However, although consistent, equation (2.69) involves serious computational problems for applied work. Note that (2.69) is not a differentiable function which implies that conventional gradient vectors cannot be used directly to solve for the parameters, β. Solving (2.69) by converting into a nonlinear program also involves some numerical problems such as the minimand may admit several minima or it may not have a unique minimum.

Paarsch (1984) conducted a Monte Carlo study in which he compared the tobit MLE, Heckman's two-step estimator, and Powell's LAD estimator under the normal, Laplace and Cauchy distributions. It was found that, for normal and Laplace distributions, the MLE performed better than the other two estimators. Powell's LAD estimator performed better than the tobit MLE for the Cauchy distribution. However, when the sample size is small the LAD estimator appeared to be neither accurate nor stable. Nawata (1992) proposed an alternative algorithm based on a linear search procedure and evaluated the LAD estimator using a Monte Carlo experiment designed after Paarsch's (1984) experiment. Nawata (1992) also noted that the LAD estimator appears to do well under the Cauchy distribution, given that the sample size is very large. However, the LAD estimator is very unstable even for moderate sample sizes and the computational cost of the LAD estimator as compared to the MLE or H2S estimators remains incomparably high. Nawata's (1992) paper also questions the reliability of the results obtained for the LAD estimator in Paarsch's (1984) paper. In general, one major limitation of Powell's (1984) LAD estimator is its computational difficulty even if the model is simple.

Alternatively, Powell (1986b) proposed a symmetrically censored least squares (SCLS) estimator which is based upon symmetric censoring or truncation of the upper tail of the distribution of the dependent variable. The SCLS estimator, although useful, is more relevant to the truncated model than the censored regression model.

The SCLS is defined by minimizing the following function with respect to β

$$R_N(\beta) = \sum_{i=1}^{N}[(y_i - \max\{y_i/2, x_i'\beta\})^2 + I(y_i > 2x_i'\beta)[(y_i/2)^2 - (\max\{0, x_i'\beta\})^2]] \quad (2.70)$$

where $I(A)$ is an indicator which takes the value 1 if A is positive 0 otherwise.

Like the LAD estimator, the SCLS does not require the assumption of identically and independently distributed Gaussian errors. It is consistent and asymptotically normal for a wide class of (symmetric) error distributions.

In addition to robustness to non-normality of the disturbances, the censored least absolute deviations (LAD) estimator and the symmetrically censored least squares (SCLS) estimator have an additional desirable property of robustness to heteroscedasticity. However, these estimators can be inefficient under the correct specification of the model (i.e., when the assumptions of the model actually hold), because they do not make full use of the information on the parameters.

Other papers which are related to the semi-parametric estimation of the tobit model include, among others, those of Buckly and James (1979), Horowitz (1986), Duncan (1986), Moon (1989) and Blundell and Smith (1994). Interested readers may refer to these and the references there in for details. These estimators, although useful, are computationally difficult and their use in applied research is very limited. More importantly, the small sample properties of these estimators are not known.

2.9.2 Bounded Influence Estimators of the Model

As discussed earlier in this study, the tobit MLE, although widely used in applied research, is not robust to violations of the assumptions made for the disturbances.

On the other hand, semi-parametric estimators such as the Powell's (1984) least absolute deviations (LAD) estimator and Powell's (1986b) symmetrically censored least squares (SCLS) have certain robustness properties, but since they disregard entirely the information contained in the parametric assumptions, they can be inefficient when the assumptions of the tobit model hold.

Bounded-influence estimators provide a compromise between efficiency and robustness. These estimators make use of the parametric assumptions and hence attain high efficiency when the model is correctly specified, but are robust in Hampel's (1971) sense; that is, their probability distribution changes only slightly for small changes in the underlying probability distribution of the observation. These estimators are also referred to as 'optimal bounded-influence estimators', since they have a bounded influence function [Hampel (1974)] which guarantees protection from the effects of small departures from the assumptions of the model. They, therefore, attain the best trade-off between efficiency and robustness. They can also be viewed as weighted maximum likelihood estimators, with weights depending on the particular choice of the efficiency or robustness criteria. Earlier works on bounded-influence estimators include those of Hampel (1978), Krasker (1980) and Krasker and Welsh (1982) for the traditional linear regression model and the bounded-influence estimator for the logit model which was proposed recently by Stefanski, Carroll and Ruppert (1986).

More recently, Peracchi (1990) introduced a new class of bounded-influence estimators for the standard tobit model. He defined a class of bounded-influence estimators which are all based on a score function related to the tobit model. Peracchi (1990) provided an empirical example using household expenditure survey data of the Sudan and compared the results with other tobit estimators including the maximum likelihood estimator (MLE) and Powell's least absolute deviations (LAD) and symmetrically censored least squares (SCLS) estimators. Peracchi (1990) noted that the bounded-influence estimators provided results which are generally close to those of the Powell's LAD and SCLS estimators. However, the bounded-influence estimators appear to be relatively more precise than the semi-parametric estimators. The results for the MLE were sensitive to a few extreme observations and look very poor in some cases. Finally, Peracchi's results imply that the bounded-influence weights may provide useful diagnostic information for identifying potential problems such as outliers and influential observations.

2.9.3 Bayesian Estimation of the Tobit Model

Bayesian methods are widely used in theoretical econometrics and statistics. Their use in applied research, although relatively limited, has also increased rapidly in recent years [see Koop (1994) who provides a survey on the applications of Bayesian techniques in applied research]. The rapid growth in applications of Bayesian methods is associated with advances in computer power and the availability of various Bayesian techniques which make use of the computer power.

The main paper which focusses on Bayesian estimation of the tobit model is that of Chib (1992). Other related works include those of Carriquiry et el. (1987) and Sweeting (1987) in the biomedical sciences and Zellner and Rossi (1984) for the probit model.

Chib (1992) considered the standard tobit model defined in this study and analyzed the model in a Bayesian context. He provided a simple condition for the existence of posterior moments. He also developed suitable Monte Carlo procedures based on symmetric multivariate-t distributions and Laplacian approximations. Ideas such as data generation and augmentation and Gibbs sampling were also developed and discussed. Chib's paper provided an example and demonstrates the feasibility of Bayesian techniques for estimation of the model. The reader may refer to Chib (1992) and the references therein for further discussion.

2.10 Some Useful Results

As a final remark regarding the use and the interpretation of results, it is important to understand when and how the results of tobit models are used in applied research. Recall that there are three regression functions which are associated with the model. These are:

$$E[y_i^*] = x_i'\beta \tag{2.71}$$

$$E[y_i|y_i > 0] = x_i'\beta + \sigma\, f(x_i'\alpha)/F(x_i'\alpha) \tag{2.72}$$

$$\begin{aligned} E[y_i] &= F(x_i'\alpha).\, E[y_i|y_i > 0] \\ &= F(x_i'\alpha)\,[x_i'\beta + \sigma\, f(x_i'\alpha)/F(x_i'\alpha)] \end{aligned} \tag{2.73}$$

Further, if one is interested in the effects of a unit change in an explanatory variable on any one of the above expectations then one can make use of the following useful results.

(i) The effects of unit change in x_j (suppressing observation subscript) on the latent variable is given by

$$\frac{\partial E[y_i^*]}{\partial x_j} = \beta_j \tag{2.74}$$

(ii) The effects of a unit change in x_j on y_i, $E[y_i]$ is given as

$$\frac{\partial E[y_i]}{\partial x_j} = F(x_i'\alpha).\beta_j \tag{2.75}$$

(iii) Similarly, the effects of a unit change in x_j on the conditional expectation of y_i is computed by

$$\frac{\partial E[y_i|y_i > 0]}{\partial x_j} = \beta_j \left[1 - (x_i'\alpha).\frac{f(x_i'\alpha)}{F(x_i'\alpha)} - \left(\frac{f(x_i'\alpha)}{F(x_i'\alpha)} \right)^2 \right]. \tag{2.76}$$

Thus, the use and interpretation of tobit results depends on the type of outcomes that may be required from the research. It is also important to note that, unlike the traditional regression model, all coefficients including the constant term play an important role in computing the responses, as shown in items (ii) and (iii) above.

2.11 Summary and Conclusions

Tobit models refer to regression models in which the observations on the dependent variable are restricted to a specific range. Since Tobin's (1958) paper where he suggested the standard tobit model, many types of tobit models have appeared in the literature, ranging from a single equation tobit model to more complex simultaneous equation tobit models. As a result, various types of estimators have been suggested to estimate the parameters of the model.

In this Chapter, we reviewed those estimators of the model which are particularly relevant to the estimation of the parameters of the standard tobit model which is usually referred to simply as the tobit model. However, some of the properties of these estimators may also apply to other types of tobit models with some adjustments.

These estimators include, among others, the maximum likelihood estimator (MLE), the Heckman's 2-step (H2S) and its weighted version, the weighted Heckman's 2-step (WH2S) estimator, the Heckman-type 2-step estimator based on the unconditional expectation of the model and its weighted version, and nonlinear estimators based on the conditional as well as unconditional expectations of the model. Furthermore, some highlights on recent developments with regard to semi-parametric, bounded-influence and Bayesian estimation of the parameters of the model and the relevant citations are provided. In general, comparison between these estimators using analytical methods is either difficult or impossible.

Of these estimators, the MLE and the H2S estimator are widely used in applied research. The MLE provides consistent and asymptotically more efficient estimates, provided that the model is correctly specified. However, the MLE is not robust if the assumptions of the model are violated. For example, some studies have indicated that the MLE is not only inefficient but can also be inconsistent under non-normality of the error terms of the model, which is in contrast to the traditional regression model where the MLE is consistent under a wide variety of conditions. Despite these warnings, the MLE of the tobit model is used to estimate the parameters of the model in most applied papers.

Another estimator which is frequently used in applied research is the H2S estimator. The H2S estimator is consistent and usually preferable for its simplicity, especially for tobit models involving simultaneous equations. However, it performs poorly in finite samples because of unavoidable multicollinearity between the explanatory variables and the estimated inverse of the Mill's ratio. An improved estimator along the lines of Heckman's estimator is suggested to avoid this problem and will be discussed in the next Chapter.

Other estimators have been used little in applied research, perhaps for a variety of reasons. For example, among other things, the nonlinear least squares estimators involve non-trivial complex functions and hence are not computationally easier than either the MLE or H2S estimators. Also convergence may not be guaranteed. Similarly, the semi-parametric estimators are computationally burdensome even for the simplest cases. This is coupled with lack of available statistical/econometric packages that incorporate these estimators. Most of all, there appears to be lack of a clear evidence on the finite sample properties of most of the estimators of the model.

It is also important to note that the semi-parametric and Bayesian estimators of the tobit model have been hardly used in applied research and no further analysis of these estimators will be provided in this study. However, they have some potentially attractive characteristics and there appears to be a need for further research in this direction. For example, as shown by Newey (1987) and Peters and Smith (1991), the semi-parametric estimators can be used for pre-testing for the normality of the errors of the model.

The purpose of this study is, therefore, to provide unified and relatively comprehensive Monte Carlo evidence regarding the relative finite sample performance of most of the estimators so that they can be used as a practical guide (indicators) in applied research.

Chapter 3

An Improved Heckman Estimator and its Properties

3.1 Introduction

In Chapter 2, we discussed most of the estimators of the standard tobit model. Of these estimators, the H2S estimator is relatively simple and usually preferred in applied research for its computational ease. However, some studies have indicated that the H2S estimator performs relatively poorly in finite samples [Wales and Woodland (1980), Nelson (1984), Paarsch (1984), Nawata (1993, 1994)]. One of the main reasons, among others, is the presence of strong and often unavoidable collinearity between the explanatory variables and the estimated inverse of the Mill's ratio (hazard function).

In this Chapter, we introduce an improved estimator which is along the lines of Heckman's two-step estimator and which is referred to as the three-step estimator

(3SE). The 3SE, although similar to the H2S estimator, does not suffer the serious multicollinearity problem which characterizes the H2S estimator. Computationally, the 3SE preserves the simplicity of the H2S estimator.

Section 3.2 discusses the main steps involved in the 3SE and the likely advantages as compared to the H2S estimator. Then, the asymptotic properties (i.e., consistency and the variance-covariance matrix) of the 3SE are derived in Section 3.3. Section 3.4 provides some generalisations of the 3SE. Finally, Section 3.5 presents the conclusions.

3.2 The Three-step Estimator (3SE)

Recall the conditional expectation of the model which is defined by

$$y_i = x_i'\beta + \sigma\ \lambda(x_i'\alpha) + \varepsilon_i \tag{3.1}$$

where the various components of the model are as defined in Section 2.4.

Given this model, the H2S estimates are obtained by estimating α, say $\hat{\alpha}$, and hence $\hat{\lambda}(x_i'\hat{\alpha})$ by the probit maximum likelihood estimator in the first step of the procedure. Then, the coefficients of the model are estimated directly from (3.1) by regressing y_i on the x's and $\hat{\lambda}(x_i'\hat{\alpha})$, using only those observations for which y_i is positive.

However, the H2S estimates are likely to be imprecise because of the multicollinearity problem between the explanatory variables of the model. That is, as can be seen from (3.1), $\lambda(x_i'\alpha)$ is expressed as a function of the x's and hence there is an inherent problem of multicollinearity resulting from the particular form of the model. But most importantly, the severity of the multicollinearity problem arises because $\lambda(.)$ can be approximated by a linear function of the form $\lambda(x) = a + b(x)$ over a wide range of observations, where the values of a and b depend on the observations on the x's [see Johnston and Kotz (1970, p. 123), Tobin (1958)].

In general, multicollinearity is unavoidable and often strong in the second step of the H2S procedure and often leads to estimates which are relatively unreliable [see Wales and Woodland (1980), Nelson (1984), Paarsch (1984), Nawata (1993, 1994)]. In particular, Nawata (1993) showed that there almost always exists a very high (negative) correlation between the explanatory variables and the estimated inverse of Mill's ratio, $\hat{\lambda}(x_i'\hat{\alpha})$, and it is this correlation which causes the inefficiency of the H2S estimator in most cases.

Another, perhaps less important problem, is that the estimated value of σ obtained directly from (3.1) is not guaranteed to be positive, which is contrary to theoretical expectations. Although there appears to be very little practical importance in the interpretation of the estimated value of σ, it indicates the degree of unreliability involved in the H2S procedure.

Below, we propose an improved 3-step estimator (3SE), which is along the lines of Heckman's estimator, but which avoids the above problems. The three major steps involved in the 3SE are as follows:

Step 1. The first step of the 3SE, similar to the H2S estimator, involves the estimation of α, say $\hat{\alpha}$, using the probit maximum likelihood estimator. The main departure of the 3SE is in the following two steps.

Step 2. The model in (3.1) can be rearranged as

$$y_i = \sigma[x_i'\alpha + \lambda(x_i'\alpha)] + \varepsilon_i \tag{3.2}$$

If the quantity in square brackets is known, equation (3.2) is a simple linear regression model with no constant term. Thus, once the right hand side observations are estimated in Step 1, one can estimate σ by regressing y_i on $[x_i'\hat{\alpha} + \hat{\lambda}(x_i'\hat{\alpha})]$, using the N_1 observations for which y_i is positive. It is important to note that both the left, y_i, and right hand side, $[x_i'\hat{\alpha} + \hat{\lambda}(x_i'\hat{\alpha})]$, variables in equation (3.2) are positive. This implies that the estimated value of σ will be positive, a case which is not guaranteed directly from the H2S estimator [see Heckman (1976, p.482)]. Given this, the coefficients of the model, β's, can be estimated consistently by adding one more step as follows.

Step 3. Let $\hat{\sigma}_{3S}$ be an estimate of σ from equation (3.2) in Step 2. Then substituting $\hat{\sigma}_{3S}$ in equation (3.1) and rearranging the model gives

$$y_i - \hat{\sigma}_{3S}\hat{\lambda}(x_i'\hat{\alpha}) = x_i'\beta + \eta_i + \varepsilon_i \tag{3.3}$$

where $\eta_i = \sigma\lambda(x_i'\alpha) - \hat{\sigma}_{3S}\hat{\lambda}(x_i'\hat{\alpha})$.

Equation (3.3) can be written as

$$\tilde{y}_i = x_i'\beta + \eta_i + \varepsilon_i \tag{3.4}$$

where $\tilde{y} = y_i - \hat{\sigma}_{3S}\hat{\lambda}(x_i'\hat{\alpha})$.

Thus, the 3S estimates of the model are obtained by applying the method of least squares on equation (3.4). That is, one can estimate the parameters of the model by regressing $\tilde{y}_i$ on the x's.

Given this, the 3-step estimator (3SE) has two important advantages as compared to the H2S procedure. These are:

(i) As can be seen from Step 2 above, the estimate of σ from (3.2) is not only positive, but can be estimated more precisely even if $x_i'\hat{\alpha}$ and $\hat{\lambda}(x_i'\hat{\alpha})$ are indistinguishable.

(ii) The most important aspect of the 3SE is that, unlike the H2S estimator, equation (3.4) consists only of the x's in the right hand side of the model and does not involve $\hat{\lambda}(x_i'\hat{\alpha})$, which is the main source of the multicollinearity problem under the H2S procedure.

Before discussing the asymptotic properties of the 3S estimator let's consider the following.

Consider equation (3.4), and using matrix notation, we have

$$\tilde{Y} = X\beta + \eta + \varepsilon \tag{3.5}$$

where $\tilde{Y} = Y - \hat{\sigma}_{3S}\hat{\lambda}$ is an N_1x1 vector of observations on $\tilde{y}_i$, Y and $\hat{\lambda}$ are N_1x1 vectors whose elements are y_i and $\hat{\lambda}(x_i'\hat{\alpha})$, respectively, X is an N_1xk matrix of explanatory variables corresponding to Y, and η and ε are vectors of order N_1 whose elements are η_i and ε_i, respectively.

Let $\hat{\beta}_{3S}$ be the 3S estimator of β. Then, from (3.5), it is defined by

$$\hat{\beta}_{3S} = (X'X)^{-1}X'\tilde{Y} \tag{3.6}$$

which can be expressed equivalently as

$$\begin{aligned} \hat{\beta}_{3S} &= (X'X)^{-1}X'Y - \hat{\sigma}_{3S}(X'X)^{-1}X'\hat{\lambda} \\ &= \hat{\beta}_{OLSP} - \hat{\sigma}_{3S}\,\hat{\gamma}_{OLSP} \end{aligned} \tag{3.7}$$

where $\hat{\beta}_{OLSP}$ is the least squares estimator obtained by regressing y_i on x's using only the observations for which y_i is positive (OLSP); similarly $\hat{\gamma}_{OLSP}$ is obtained by regressing $\hat{\lambda}(x_i'\hat{\alpha})$ on the x's using the N_1 observations and $\hat{\sigma}_{3S}$ is the OLS estimator of σ obtained from (3.2), in Step 2.

The second term on the right hand side of (3.7) can be referred to as the correction factor or the bias of the OLSP estimator. The expression (3.7) is also similar to that of Fair's (1977) MLE estimator of the tobit model discussed in Chapter 2 of this study. There are two points which are worth noting regarding the relationship given by equation (3.7). These are: (i) It is clear that once $\hat{\lambda}$ and $\hat{\sigma}_{3S}$ are estimated from the first and second step, respectively, the 3S estimates of β, $\hat{\beta}_{3S}$, may be obtained directly from (3.7) by correcting the bias of the OLSP estimator. (ii) More importantly, Equation (3.7) shows the relationship between the OLSP and the 3SE explicitly. It is also clear that the magnitude and the direction of the bias of the OLSP estimator can be determined directly based on this relationship. Note that previous studies have noted that one cannot determine the magnitude and the direction of bias of the OLSP estimator without making further assumptions about the model [see Goldberger (1981)].

3.3 Asymptotic Properties of the 3S estimator

3.3.1 Consistency

Note that, as discussed above, the first step of the three-step procedure (which is the same as the first step in the H2S estimator) involves the estimation of α, say $\hat{\alpha}$, using the probit maximum likelihood estimator. It is a well known result that $\hat{\alpha}$

is a consistent estimator of α [see Chapter 2, Section 2.4 of this study]. The main purpose here is to show that the estimators in the second and the third step of the 3S procedure, $\hat{\sigma}_{3S}$ and $\hat{\beta}_{3S}$, are also consistent. To do this we proceed as follows:

First, let us assume the usual standard assumptions as follows:

Assumption 1: X is uniformly bounded, where X is, as defined in (3.5), an N_1xk matrix of explanatory variables.

Assumption 2: $\lim_{N_1\to\infty}\left(\frac{X'X}{N_1}\right)$ exists.

Assumption 3: $\operatorname{plim} X'\varepsilon/N_1 = 0$, where ε is defined in (3.5).

Result 1:

Let $\hat{\sigma}_{3S}$ be the 3-step estimator of σ, then $\hat{\sigma}_{3S}$ is consistent.

Proof:

Consider equation (3.2), given by

$$y_i = \sigma\, z_i + \varepsilon_i \tag{3.8}$$

where $z_i = x_i'\alpha + \lambda(x_i'\alpha)$. Define $\hat{z}_i = x_i'\hat{\alpha} + \lambda(x_i'\hat{\alpha})$.

Then we have

$$y_i = \sigma \hat{z}_i + \nu_i + \varepsilon_i \tag{3.9}$$

where $\nu_i = (z_i - \hat{z}_i)\sigma$.

Regressing y_i on $\hat{z}_i$, we obtain

$$\begin{aligned} \hat{\sigma}_{3S} &= (\hat{z}'\hat{z})^{-1}\hat{z}'y \\ &= \sigma + (\hat{z}'\hat{z})^{-1}\hat{z}'(\nu+\varepsilon) \end{aligned} \tag{3.10}$$

Now, $\hat{\sigma}_{3S}$ is consistent if the following three conditions hold:

$$\text{plim}\left(\frac{\hat{z}'\hat{z}}{N_1}\right) \text{ is finite,}$$

$$\text{plim}\left(\frac{\hat{z}'\nu}{N_1}\right) = 0,$$

$$\text{plim}\left(\frac{\hat{z}'\varepsilon}{N_1}\right) = 0.$$

Let $\hat{z} = \hat{Z}\hat{\delta}$ where $\hat{Z} = [X, \hat{\lambda}]$ and $\hat{\delta} = (\hat{\alpha}', 1)'$ implying $\hat{z}'\hat{z} = \hat{\delta}'\hat{Z}'\hat{Z}\hat{\delta}$.

Thus,

$$\text{plim}\left(\frac{\hat{z}'\hat{z}}{N_1}\right) = \text{plim}\left[\hat{\delta}'\left(\frac{\hat{Z}'\hat{Z}}{N_1}\right)\hat{\delta}\right] \tag{3.11}$$

We know that, since the probit MLE $\hat{\alpha}$ is a consistent estimator of α, we have (see Amemiya (1985) p. 369)

$$\text{plim}\left(\frac{\hat{Z}'\hat{Z}}{N_1}\right) = \lim_{N_1\to\infty}\left(\frac{Z'Z}{N_1}\right) \text{ exists}$$

and therefore

$$\text{plim}\left(\frac{\hat{z}'\hat{z}}{N_1}\right) = \text{plim}\left[\hat{\delta}'\left(\frac{\hat{Z}'\hat{Z}}{N_1}\right)\hat{\delta}\right] = \left[\delta'\left(\lim_{N_1\to\infty}\frac{Z'Z}{N_1}\right)\delta\right] \text{ is finite.} \tag{3.12}$$

For the second condition,

$$\begin{aligned}
\text{plim}\left(\frac{\hat{z}'\nu}{N_1}\right) &= \text{plim}\frac{1}{N_1}\left[(\hat{\alpha}' \;\; 1)\begin{pmatrix} X' \\ \hat{\lambda}' \end{pmatrix}\left[(X \;\; \lambda)\begin{pmatrix} \alpha \\ 1 \end{pmatrix} - (X \;\; \hat{\lambda})\begin{pmatrix} \hat{\alpha} \\ 1 \end{pmatrix}\right]\right]\sigma \\
&= \text{plim}\left[(\hat{\alpha}' \;\; 1)\begin{pmatrix} \frac{X'X}{N_1} & \frac{X'\lambda}{N_1} \\ \frac{\hat{\lambda}'X}{N_1} & \frac{\hat{\lambda}'\lambda}{N_1} \end{pmatrix}\begin{pmatrix} \alpha \\ 1 \end{pmatrix}\right. \\
&\quad \left. - (\hat{\alpha}' \;\; 1)\begin{pmatrix} \frac{X'X}{N_1} & \frac{X'\hat{\lambda}}{N_1} \\ \frac{\hat{\lambda}'X}{N_1} & \frac{\hat{\lambda}'\hat{\lambda}}{N_1} \end{pmatrix}\begin{pmatrix} \hat{\alpha} \\ 1 \end{pmatrix}\right]\sigma \\
&= 0
\end{aligned} \tag{3.13}$$

Finally, the third condition is given by

$$\text{plim}\left(\frac{\hat{z}'\varepsilon}{N_1}\right) = \text{plim}\left(\frac{\hat{\alpha}'X'\varepsilon}{N_1} + \frac{\hat{\lambda}'\varepsilon}{N_1}\right) \tag{3.14}$$

$$= \text{plim}\left(\frac{\hat{\alpha}'X'\varepsilon}{N_1}\right) + \text{plim}\left(\frac{\hat{\lambda}'\varepsilon}{N_1}\right) \tag{3.15}$$

Now, we know that ε is uncorrelated with $\hat{\alpha}$ (see Amemiya 1985). Also, ε and X are uncorrelated (by Assumption 3). Thus, the first term in the right hand side of (3.15) goes to 0. Further, in the view that $\hat{\lambda}$ is expressed as a function of X and $\hat{\alpha}$ it is clear that $\hat{\lambda}$ and ε are not correlated.

Thus,

$$\text{plim}\left(\frac{\hat{z}'\varepsilon}{N_1}\right) = 0 \tag{3.16}$$

Therefore, from (3.12), (3.13) and (3.16) it follows that $\hat{\sigma}_{3S}$ is a consistent estimator of σ.

Result 2:

Let $\hat{\beta}_{3S}$ be the 3S estimator of β, then $\hat{\beta}_{3S}$ is consistent.

Proof:

Consider (3.5), which can be written as

$$Y - \hat{\sigma}_{3S}\hat{\lambda} = X\beta + \eta + \varepsilon, \tag{3.17}$$

where $\eta = \sigma\lambda - \hat{\sigma}_{3S}\hat{\lambda}$ and $\lambda = \lambda(x_i'\alpha)$.

Then, $\hat{\beta}_{3S}$ is defined by

$$\hat{\beta}_{3S} = \beta + (X'X)^{-1}X'(\eta + \varepsilon) \tag{3.18}$$

Thus, $\hat{\beta}_{3S}$ is consistent if the following conditions hold:

$$\text{plim}\left(\frac{X'\varepsilon}{N_1}\right) = 0 \tag{3.19}$$

$$\text{plim}\left(\frac{X'\eta}{N_1}\right) = 0 \tag{3.20}$$

The first condition (3.19) is true by Assumption 3. Because $\hat{\sigma}_{3S}$ is consistent from the previous argument, it follows that

$$\text{plim}\left(\frac{X'\eta}{N_1}\right) = \text{plim}\left(\frac{\sigma X'\lambda}{N_1} - \frac{\hat{\sigma}_{3S}X'\hat{\lambda}}{N_1}\right) = 0$$

Therefore, $\hat{\beta}_{3S}$ is a consistent estimator of β.

3.3.2 Asymptotic Distributions of $\hat{\sigma}_{3S}$ and $\hat{\beta}_{3S}$

Consider (3.9), given by

$$y_i = \sigma\hat{z}_i + \nu_i + \varepsilon_i \tag{3.21}$$

where $\nu_i = (z_i - \hat{z}_i)\sigma$.

$\hat{\sigma}_{3S}$ is defined as

$$\hat{\sigma}_{3S} = \sigma + (\hat{z}'\hat{z})^{-1}\hat{z}'(\nu + \varepsilon),$$

which then implies

$$\sqrt{N_1}(\hat{\sigma}_{3S} - \sigma) \stackrel{a}{=} \sqrt{N_1}(\hat{z}'\hat{z})^{-1}\hat{z}'(\nu + \varepsilon) \tag{3.22}$$

The expression in (3.22) implies that the right hand side has the same limiting distribution as the left hand side.

Define

$$\hat{a}' = (\hat{z}'\hat{z})^{-1}\hat{z}' \tag{3.23}$$

Then, (3.22) can be written as

$$\sqrt{N_1}(\hat{\sigma}_{3S} - \sigma) \overset{a}{=} \sqrt{N_1}\hat{a}'(\nu + \varepsilon) \tag{3.24}$$

Now, consider

$$\begin{aligned} \nu &= \sigma(z - \hat{z}) = \sigma(Z\delta - \hat{Z}\hat{\delta}) \\ &= \sigma\left[(X \;\; \lambda)\begin{pmatrix} \alpha \\ 1 \end{pmatrix} - (X \;\; \hat{\lambda})\begin{pmatrix} \hat{\alpha} \\ 1 \end{pmatrix}\right] \\ &= \sigma[X(\alpha - \hat{\alpha}) + (\lambda - \hat{\lambda})] \end{aligned} \tag{3.25}$$

By Taylor expansion of $\hat{\lambda}$ around λ, we have (see also Amemiya (1985, p.369))

$$(\lambda - \hat{\lambda}) = -\frac{\partial \lambda}{\partial \alpha'}(\hat{\alpha} - \alpha) + O_p(N_1^{-1}) \tag{3.26}$$

Substituting (3.26) into (3.25), we have

$$\nu = -\sigma[X(\hat{\alpha} - \alpha) + \frac{\partial \lambda}{\partial \alpha'}(\hat{\alpha} - \alpha)] + O_p(N_1^{-1}) \tag{3.27}$$

Also using (3.17),

$$\frac{\partial \lambda}{\partial \alpha'} = (\Sigma - I)X, \tag{3.28}$$

where Σ is an $N_1 \mathrm{x} N_1$ matrix defined in (2.54) and I stands for identity matrix.

And hence substituting (3.28) into (3.27) gives

$$\begin{aligned} \nu &= -\sigma[X(\hat{\alpha} - \alpha) + (\Sigma - I)X(\hat{\alpha} - \alpha)] + O_p(N_1^{-1}) \\ &= -\sigma\Sigma X(\hat{\alpha} - \alpha) + O_p(N_1^{-1}). \end{aligned} \tag{3.29}$$

Substituting (3.29) in (3.24) and because

$$\begin{aligned} \hat{a}'\sigma\Sigma X &= a'\sigma\Sigma X + O_p(N_1^{-1/2}) \quad \text{and} \\ \hat{\alpha} - \alpha &= O_p(N_1^{-1/2}) \end{aligned} \tag{3.30}$$

It follows that

$$\sqrt{N_1}(\hat{\sigma}_{3S} - \sigma) \stackrel{a}{=} \sqrt{N_1}\, a'[-\sigma\Sigma X(\hat{\alpha} - \alpha) + \varepsilon] + O_p(N_1^{-1/2}). \tag{3.31}$$

Further, as shown in Amemiya (1985, pp. 366-370), $(\hat{\alpha}-\alpha)$ and ε are uncorrelated. It follows that,

$$\sqrt{N_1}\, a'\nu \xrightarrow{d} N(0, \sigma^2 \lim_{N_1\to\infty} \left(\frac{a'(\Sigma X V(\hat{\alpha}) X'\Sigma)a}{N_1}\right)) \tag{3.32}$$

and

$$\sqrt{N_1}\, a'\varepsilon \xrightarrow{d} N(0, \sigma^2 \lim_{N_1\to\infty} \left(\frac{a'\Sigma a}{N_1}\right)) \tag{3.33}$$

Using (3.31), (3.32) and (3.33), it follows that

$$\sqrt{N_1}(\hat{\sigma}_{3S} - \sigma) \xrightarrow{d} N(0, \sigma^2 Q^{-1} R Q^{-1}), \tag{3.34}$$

where

$$Q = \lim_{N_1\to\infty} \left(\frac{z'z}{N_1}\right) \quad \text{and}$$

$$R = \lim_{N_1\to\infty} \left(\frac{z'(\Sigma X V(\hat{\alpha}) X'\Sigma + \Sigma)z}{N_1}\right).$$

Therefore, the asymptotic variance of $\hat{\sigma}_{3S}$ is given by

$$V(\hat{\sigma}_{3S}) = \sigma^2 a'(\Sigma X V(\hat{\alpha}) X'\Sigma + \Sigma)a. \tag{3.35}$$

where $a' = (z'z)^{-1}z'$, $V(\hat{\alpha})$ is the covariance matrix of the probit estimator, $\hat{\alpha}$, and is defined by (2.48), in Chapter 2.

Now, to derive the asymptotic distribution of $\hat{\beta}_{3S}$, consider (3.17), given by

$$Y - \hat{\sigma}_{3S}\hat{\lambda} = X\beta + \eta + \varepsilon, \tag{3.36}$$

where $\eta = \sigma\lambda - \hat{\sigma}_{3S}\hat{\lambda}$.

Then, $\hat{\beta}_{3S}$ is defined by

$$\hat{\beta}_{3S} = \beta + (X'X)^{-1}X'(\eta + \varepsilon)$$

which implies

$$\sqrt{N_1}(\hat{\beta}_{3S} - \beta) \overset{a}{=} \sqrt{N_1}(X'X)^{-1}X'(\eta + \varepsilon) \tag{3.37}$$

Now, consider

$$\begin{aligned} \eta &= \sigma\lambda - \hat{\sigma}_{3S}\hat{\lambda} \\ &= \sigma(\lambda - \hat{\lambda}) - \hat{\lambda}\hat{a}'(\nu + \varepsilon) \end{aligned} \tag{3.38}$$

Again, from (3.26), using Taylor's expansion of λ around $\hat{\lambda}$, we have

$$\eta = -\sigma\frac{\partial\lambda}{\partial\alpha'}(\hat{\alpha} - \alpha) - \lambda\hat{a}'(\nu + \varepsilon) + O_p(N_1^{-1}) \tag{3.39}$$

substituting (3.28) in (3.39) and from (3.31), we have

$$\eta = \sigma(I - \Sigma)X(\hat{\alpha} - \alpha) + \sigma[a'\Sigma X(\hat{\alpha} - \alpha)]\lambda - (a'\varepsilon)\lambda + O_p(N_1^{-1}) \tag{3.40}$$

Thus, combining (3.37) and (3.40), we get

$$\begin{aligned} \sqrt{N_1}(\hat{\beta}_{3S} - \beta) &\overset{a}{=} \sqrt{N_1}\,(X'X)^{-1}X'v \\ &= \left(\frac{X'X}{N_1}\right)^{-1}\frac{X'v}{\sqrt{N_1}} \end{aligned} \tag{3.41}$$

where

$$\begin{aligned} v &= v_1 + v_2 + O_p(N_1^{-1}) \quad \text{and} \\ v_1 &= \varepsilon - (a'\varepsilon)\lambda \\ v_2 &= \sigma\{(I - \Sigma)X(\hat{\alpha} - \alpha) + [a'\Sigma X(\hat{\alpha} - \alpha)]\lambda\}. \end{aligned}$$

Now, consider the properties of v_1 and v_2,

$$E(v_1) = E[\varepsilon - (a'\varepsilon)\lambda] = 0$$

and

$$\begin{aligned} E(v_2) &= E[\sigma\{(I-\Sigma)X(\hat{\alpha}-\alpha) + [a'\Sigma X(\hat{\alpha}-\alpha)]\lambda\}] \\ &= \sigma\{(I-\Sigma)X\,E(\hat{\alpha}-\alpha) + [a'\Sigma X\,E(\hat{\alpha}-\alpha)]\lambda\} = 0 \qquad (3.42) \end{aligned}$$

Further,

$$\begin{aligned} E(v_1 v_1') &= E[\varepsilon - (a'\varepsilon)\lambda][\varepsilon - (a'\varepsilon)\lambda]' \\ &= E[\varepsilon\varepsilon' - \varepsilon(\varepsilon'a)\lambda' - \lambda(a'\varepsilon)\varepsilon' + (a'\varepsilon)(\varepsilon'a)\lambda\lambda'] \\ &= \sigma^2[\Sigma - \Sigma a\lambda' - \lambda a'\Sigma + (a'\Sigma a)\lambda\lambda'] \qquad (3.43) \end{aligned}$$

Similarly,

$$\begin{aligned} E(v_2 v_2') &= E[\sigma\{(I-\Sigma)X(\hat{\alpha}-\alpha) + [a'\Sigma X(\hat{\alpha}-\alpha)]\lambda\}] \\ &\quad [\sigma\{(I-\Sigma)X(\hat{\alpha}-\alpha) + [a'\Sigma X(\hat{\alpha}-\alpha)]\lambda\}]' \\ &= \sigma^2[(I-\Sigma)XV(\hat{\alpha})X'(I-\Sigma) + (I-\Sigma)XV(\hat{\alpha})X'\Sigma a\lambda' \\ &\quad +\lambda a'\Sigma XV(\hat{\alpha})X'(I-\Sigma) + (a'\Sigma XV(\hat{\alpha})X'\Sigma a)\lambda\lambda'] \qquad (3.44) \end{aligned}$$

and

$$\begin{aligned} E[v_1 v_2] &= E[\varepsilon - (a'\varepsilon)\lambda][\sigma\{(I-\Sigma)X(\hat{\alpha}-\alpha) + [a'\Sigma X(\hat{\alpha}-\alpha)]\lambda\}] \\ &= \sigma E[\{(I-\Sigma)X(\hat{\alpha}-\alpha)\varepsilon + [a'\Sigma X(\hat{\alpha}-\alpha)\varepsilon]\lambda\} \\ &\quad -(a'\varepsilon\lambda[\{(I-\Sigma)X(\hat{\alpha}-\alpha)\varepsilon + [a'\Sigma X(\hat{\alpha}-\alpha)\varepsilon]\lambda\}] \\ &= 0. \qquad (3.45) \end{aligned}$$

That is, because ε and $(\hat{\alpha} - \alpha)$ are uncorrelated (see Amemiya (1985), p. 369-370), v_1 and v_2 are uncorrelated.

Thus, following Amemiya (1985), we have

$$\begin{aligned} \sqrt{N_1}(\hat{\beta}_{3S} - \beta) &\stackrel{a}{=} \sqrt{N_1}\,(X'X)^{-1}X'v \\ &= (N_1^{-1}X'X)^{-1}\,(N_1^{-1/2}X'v_1 + N_1^{-1/2}X'v_2) \end{aligned} \tag{3.46}$$

It follows that,

$$\sqrt{N_1}\;X'v_1 \xrightarrow{d} N(0, \sigma^2 \lim_{N_1\to\infty}\left(\frac{X'H_1^*X}{N_1}\right)) \tag{3.47}$$

and

$$\sqrt{N_1}\;X'v_2 \xrightarrow{d} N(0, \sigma^2 \lim_{N_1\to\infty}\left(\frac{X'H_2^*X}{N_1}\right)) \tag{3.48}$$

Thus,

$$\sqrt{N_1}(\hat{\beta}_{3S} - \beta) \xrightarrow{d} N(0, \sigma^2 R^{-1} T R^{-1}), \tag{3.49}$$

where

$$R = \lim_{N_1\to\infty}\left(\frac{X'X}{N_1}\right) \text{ and}$$

$$T = \lim_{N_1\to\infty}\left(\frac{X'(H_1^* + H_2^*)X}{N_1}\right).$$

and

$$H_1^* = \Sigma - \Sigma a\lambda' - \lambda a'\Sigma + (a'\Sigma a)\lambda\lambda', \tag{3.50}$$

$$\begin{aligned} H_2^* &= (I - \Sigma)XV(\hat{\alpha})X'(I - \Sigma) \\ &\quad +(I - \Sigma)XV(\hat{\alpha})X'\Sigma a\lambda' \\ &\quad +\lambda a'\Sigma XV(\hat{\alpha})X'(I - \Sigma) \\ &\quad +(a'\Sigma XV(\hat{\alpha})X'\Sigma a)\lambda\lambda' \end{aligned} \tag{3.51}$$

Thus the asymptotic covariance matrix of $\hat{\beta}_{3S}$ is given by

$$V(\hat{\beta}_{3S}) = \sigma^2(X'X)^{-1}X'[H_1^* + H_2^*]X(X'X)^{-1} \tag{3.52}$$

The expression (3.52) can be estimated consistently by replacing the unknown parameters by their consistent estimates, or following White's (1980b) idea, using $(X'X)^{-1}X'BX(X'X)^{-1}$, where B is a diagonal matrix whose i^{th} diagonal element is given by $[y_i - \hat{\sigma}_{3S}\hat{\lambda}_i - x_i'\hat{\beta}_{3S}]^2$. Similarly, one can estimate the variance of $\hat{\sigma}_{3S}$ (which is given by (3.35) in a similar way.

The analogous expressions of the asymptotic variance of the H2S estimator $\hat{\sigma}_{2S}$ and the covariance of $\hat{\beta}_{2S}$ can be obtained straightforwardly using inverse matrix methods from equation (2.54) and are given by

$$V(\hat{\sigma}_{2S}) = \sigma^2 w'[\Sigma X V(\hat{\alpha})X'\Sigma + \Sigma]w \tag{3.53}$$

and

$$V(\hat{\beta}_{2S}) = \sigma^2(X'X)^{-1}X'[H_1 + H_2]X(X'X)^{-1} \tag{3.54}$$

where

$$\begin{aligned} H_1 &= \Sigma - \Sigma w\lambda' - \lambda w'\Sigma + (w'\Sigma w)\lambda\lambda', \qquad (3.55)\\ H_2 &= (I-\Sigma)XV(\hat{\alpha})X'(I-\Sigma)\\ &\quad +(I-\Sigma)XV(\hat{\alpha})X'\Sigma w\lambda'\\ &\quad +\lambda w'\Sigma XV(\hat{\alpha})X'(I-\Sigma)\\ &\quad +(w'\Sigma XV(\hat{\alpha})X\Sigma w)\lambda\lambda' \qquad (3.56)\end{aligned}$$

where $w' = (\lambda' M\lambda)^{-1}\lambda' M$ and $M = I - X(X'X)^{-1}X'$.

In finite samples $V(\hat{\sigma}_{3S})$ and $V(\hat{\beta}_{3S})$ are expected to be more efficient than $V(\hat{\sigma}_{2S})$ and $V(\hat{\beta}_{2S})$, respectively, because the former avoids the multicollinearity problem and because of the restriction imposed in the second step of the 3SE estimator.

3.4 Some Generalizations of the 3S Estimator

Note that we have shown that the 3S estimators of σ and β are consistent and that we have derived their limiting distributions. It should also be noted that, similar to the H2S estimator, the 3SE is based on a model which has a heteroscedastic error term. Thus, one can use weighted versions of (3.2) and (3.4) to obtain more efficient estimates of σ and β which are analogous to the weighted Heckman's 2-step estimators discussed in Chapter 2.

Let the weighted 3-step estimators (W3SE) of σ and β be denoted by $\hat{\sigma}_{W3}$ and $\hat{\beta}_{W3}$, respectively.

Then, they are defined as

$$\hat{\sigma}_{W3} = (z'\hat{\Psi}^{-1}z)^{-1}z'\hat{\Psi}^{-1}Y \tag{3.57}$$

where $\Psi = \sigma^2[\Sigma X V(\hat{\alpha})X'\Sigma + \Sigma]$.

and

$$\hat{\beta}_{W3} = (X'\hat{\Delta}^{-1}X)^{-1}X'\hat{\Delta}^{-1}\tilde{Y} \tag{3.58}$$

where $\Delta = \sigma^2[H_1^* + H_2^*]$, , $\tilde{Y} = Y - \hat{\sigma}_{3S}\hat{\lambda}$ (as defined in (3.5)) and H_1^* and H_2^* are given by (3.50) and (3.51), respectively.

It is straightforward to show that, under certain assumptions and following Amemiya (1984), the W3SE estimators are consistent and have normal limiting distribution (see also White (1984), Judge et al. (1988)), and their respective asymptotic variance and covariance can be obtained using standard procedures as follows:

$$\begin{aligned} V(\hat{\sigma}_{W3}) &= (z'\Psi^{-1}z)^{-1} \\ &= \sigma^2\left\{z'[\Sigma X V(\hat{\alpha})X'\Sigma + \Sigma]^{-1}z\right\}^{-1}. \end{aligned} \tag{3.59}$$

and

$$\begin{aligned} V(\hat{\beta}_{W3}) &= (X'\Delta^{-1}X)^{-1} \\ &= \sigma^2 \left\{ X'[H_1^* + H_2^*]^{-1}X \right\}^{-1}. \end{aligned} \tag{3.60}$$

Note that the 3S estimators and their weighted version, W3SE, discussed above, are all based on the conditional expectation of the model and use only the N_1 observations for which $y_i > 0$. On the other hand, one can also apply similar procedures using all observations. That is, one can use the 3-step procedure in order to improve the 2-step estimators based on the unconditional expectation of the model. That is, recall the model (see Section 2.8)

$$y_i = F(x_i'\alpha)\,[x_i'\beta + \sigma\,\lambda(x_i'\alpha)] + \delta_i \tag{3.61}$$

where $E[\delta_i] = 0$, $E[\delta_i\delta_j] = 0$, $i \neq j$, and $V(\delta_i) = \sigma^2\Omega$.

It is clear that this model is likely to suffer from multicollinearity as the explanatory variables are linearly related. That is, the arguments made against the H2S estimator also apply similarly for (3.61). On the other hand, it is also possible to apply the 3S estimator which avoids this problem. The specific steps required to obtain improved 3S estimates from (3.61) are illustrated below:

Step 1. Estimate α using probit MLE, $\hat{\alpha}$, as before.

Step 2. Estimate σ, say $\tilde{\sigma}_{3S}$, by regressing y_i on $F(x_i'\hat{\alpha})[x_i'\hat{\alpha} + \lambda(x_i'\hat{\alpha})]$.

Step 3. Estimate the β's, say $\tilde{\beta}_{3S}$, by regressing $[y_i - \tilde{\sigma}_{3S} f(x_i'\hat{\alpha})]/F(x_i'\hat{\alpha})$ on x_i'.

Note that steps 2 and 3 avoid the multicollinearity problem. The consistency and the asymptotic distributions of these estimators can be obtained using arguments which are similar to those used earlier in this Chapter. Furthermore, more efficient estimates can be obtained by applying weighted least squares, analogously to the weighted Heckman's 2-step based on the unconditional (WH2SU) expectation of the model (see Section 2.8).

Note that, in line with the objectives of the study, the estimators discussed so far are concerned with estimation of the standard tobit model. However, although there are many alternative estimators of the model, in practice almost all applied papers use either the maximum likelihood or the Heckman's two-step estimator to estimate the parameters of the model. On the other hand, while the MLE is widely used in the estimation of the standard tobit model (Type-I Tobit), the H2S estimator is more important in the estimation of more general models (for example, Type-II Tobit models) where estimation using the MLE is computationally difficult or highly costly [see Amemiya (1984), Maddala (1983)]. Thus the importance of the 3S estimator should be viewed not only as an alternative to the MLE for the simplest cases, but with regard to its potential in the estimation of more complicated models, with its advantage of avoiding the problem of multicollinearity. However, it is also important to note that this does not necessarily imply that the 3SE can be applied directly to all models that appear in the sample selection literature and that involve the H2S procedure.

Below, we consider an example, other than the standard tobit model, and demonstrate how the 3SE can be applied directly to improve the H2S estimator.

Consider the two-limit tobit model which is defined by

$$y_i^* = x_i'\beta + u_i, \qquad i = 1, 2, ..., N. \tag{3.62}$$

$$\begin{aligned} y_i &= L_{1i} \text{ if } y_i^* \leq L_{1i}, \\ &= y_i^* \text{ if } L_{1i} < y_i^* < L_{2i}, \\ &= L_{2i} \text{ if } y_i^* \geq L_{2i}. \end{aligned} \tag{3.63}$$

where y_i^* and y_i are, as defined in Chapter 2, the latent and the observed dependent variables, respectively. L_{1i} and L_{2i} are, respectively, the lower and upper limits.

In practice, such models can be used to solve many economic problems. For example, many insurance companies have a minimum and a maximum coverage and values in between them. In commodity trading markets, there exist upper and lower limits of price movements which are usually fixed based on the previous day's closing prices [see Nelson (1976), Maddala (1983)].

The likelihood function of the model is defined by [see Maddala (1983, p.161)]

$$\begin{aligned} L(\beta', \sigma | y_i, x_i, L_{1i}, L_{2i}) &= \prod_{y_i = L_{1i}} F\left(\frac{L_{1i} - x_i'\beta}{\sigma}\right) \prod_{y_i = y_i^*} \frac{1}{\sigma} f\left(\frac{y_i - x_i'\beta}{\sigma}\right) \\ &\quad \prod_{y_i = L_{2i}} \left[1 - F\left(\frac{L_{2i} - x_i'\beta}{\sigma}\right)\right]. \end{aligned} \tag{3.64}$$

where f_i and F_i are, respectively, the probability and cumulative density functions of the standard normal random variable.

Given this, it is straightforward to obtain the first and second partial derivatives for optimization. And standard iterative procedures can be used to estimate the parameters of the model [see Nelson (1976) for further discussion on computer programs].

Alternatively, one can use a simpler two-step procedure to estimate the parameters of the model as follows:

Consider the conditional expectation of y_i

$$\begin{aligned} E(y_i|L_{1i} < y_i^* < L_{2i}) &= x_i'\beta + E(u_i|L_{1i} - x_i'\beta < u_i < L_{2i} - x_i'\beta) \\ &= x_i'\beta + \sigma\Phi_i \end{aligned} \tag{3.65}$$

where

$$\Phi_i = \frac{f[(L_{1i} - x_i'\beta)/\sigma] - f[(L_{2i} - x_i'\beta)/\sigma]}{F[(L_{1i} - x_i'\beta)/\sigma] - F[(L_{2i} - x_i'\beta)/\sigma]}$$

Now, the model in (3.65) can be used for a 2-step procedure. That is, given that L_{1i} and L_{2i} are known, a two-limit probit model that uses just the number of observations at the limits, L_{1i} and L_{2i}, and the number of observations between the limits, provides consistent estimates of β/σ. The two-limit probit model has been discussed in more detail by Rosset and Nelson (1975) and its likelihood function is given as [see also Maddala (1983)]

$$L(\beta', \sigma) = \prod_{n_1} F_{1i} \prod_{n_2} (F_{2i} - F_{1i}) \prod_{n_3} (1 - F_{2i}) \tag{3.66}$$

where $F_{1i} = F[(L_{1i} - x_i'\beta)/\sigma]$, $F_{2i} = F[(L_{2i} - x_i'\beta)/\sigma]$, and n_1, n_2 and n_3 are, respectively, the number of observations corresponding to the lower limit L_{1i}, between the limits L_{1i} and L_{2i}, and the upper limit L_{2i}.

As shown by Rosset and Nelson (1975), this model can be easily maximized by any of several well known methods and the estimates of the unknown parameters can be estimated. Thus, the H2S estimator can be used because it makes use of this computational advantage. Note that we can only estimate the ratio $\alpha = \beta/\sigma$ from (3.66). Then, the parameters β and σ are estimated by regressing (3.65) after replacing Φ_i by its estimate. A similar procedure can be applied to the unconditional expectation of the model.

However, the main point of interest here is that the estimates from the second step are likely to be imprecise if there exists a high correlation between the X's and the estimated Φ_i in (3.65), as has been the case for the standard tobit model. If so, then one can readily apply the 3S procedure to avoid the multicollinearity problem. In order to see whether there exists multicollinearity between the explanatory variables we use an empirical example which is similar to that of Nawata (1993). Specifically, we generated random variables, ξ_i, from a uniform distribution over the interval [-10,10] and L_1 and L_2 were fixed at 0 and 4, respectively. Then we computed the correlation between ξ_i and $\Phi(\xi_i)$. The results are given in Table 3.1 below.

Table 3.1: Correlation between ξ and $\Phi(\xi)$

$[a, b]$	Correlation Coefficient, ρ
$[-10, 10]$	-0.9916
$[-09, 09]$	-0.9917
$[-08, 08]$	-0.9925
$[-06, 06]$	-0.9867
$[-05, 05]$	-0.9772
$[-05, 10]$	-0.9912
$[0, 10]$	-0.9815

Clearly, Table 3.1 depicts that the correlation between the explanatory variables in the second step, i.e., equation (3.65), of the procedure can be very high in magnitude.

This simply implies that the 2-step estimates are likely to be inefficient because of the presence of such multicollinearity. Therefore, the 3S estimator can be used directly to avoid the multicollinearity problem and hence obtain more efficient estimates. In fact, as shown earlier in this Chapter, there are no additional costs incurred in using the 3S estimator directly in place of the H2S estimator without even checking the presence of multicollinearity, given that it is directly applicable to the model as is the case for the two-limit tobit model.

In general, the 3S estimator, while preserving the simplicity of the H2S estimator, avoids the multicollinearity problem and may be used in place of the H2S procedure in many tobit models. The examples provided above demonstrate the useful potential for the 3S estimator in applied research. However, its use in models involving simultaneous equations (for example, Type-II Tobit models) is a subject of future research.

3.5 Summary and Conclusions

The Heckman's 2-step (H2S) estimator of the tobit model is widely used in applied research. However, it has poor finite sample properties because of unavoidable multicollinearity between the explanatory variables and the estimated inverse of Mill's ratio.

In this Chapter, we proposed an improved Heckman-type estimator which is referred to as the 3-step estimator (3SE). This estimator, while it preserves the simplicity of the frequently used H2S estimator, avoids the problem of multicollinearity which characterizes the H2S estimator. Furthermore, one can use a weighted version of the 3SE, the W3SE, to obtain more efficient estimates of the parameters of the model.

It is also shown that the 3SE estimator is consistent and has a normal limiting distribution. It is expected that the 3SE will perform better than the H2S estimator in finite samples because of its avoidance of the multicollinearity problem. Similarly, the W3SE is expected to perform better in finite samples than its counterpart, the weighted Heckman's 2-step (WH2S) estimator.

Further, the 3S estimator can be applied to improve the 2-step estimators based on the unconditional expectation of the model. Some generalizations of the 3S estimator with regard to its applications to other types of tobit models are also discussed and specific examples provided. For example, the 3S estimation procedure can be readily used in the estimation of the two-limit tobit model instead of using the H2S estimator which again suffers from serious multicollinearity problems.

Finally, it would be interesting to see if the 3S estimator could be extended to more general models, for example, the sample selection models considered in Manning, Duan and Rogers (1987) and Leung and Yu (1994) or other models in general. Research is underway in this direction.