Chapter 4

Small Sample Properties of Tobit Models: Relevant Monte Carlo/Simulation Studies

4.1 Introduction

In recent years a number of papers have appeared describing theoretical and empirical applications of tobit models. Most of the theoretical papers are concerned with the asymptotic properties of alternative estimators as well as with the development of asymptotic test statistics. On the other hand, there have been only a few Monte Carlo and/or simulation studies concerned with the small sample properties of the various estimators and test statistics suggested in the literature. In other words, little is known about the small sample properties of the estimators of the model. Some of the few finite sample studies of the estimators of the standard tobit model include those of Wales and Woodland (1980), Paarsch (1984), Moon (1989) and Nawata (1993).

In this Chapter, we discuss these and other available literature on the small sample properties of estimators which are related to the standard tobit model, which is defined in Chapter 2 of this study.

4.2 Small Sample Studies of Tobit Models

The first contribution to this information was made by Wales and Woodland (1980) who studied various methods of estimating labour supply functions. In their study they included the estimation of labour supply models using tobit models of the form (2.1)-(2.2). Specifically, Wales and Woodland (1980) considered a model of the form

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \qquad i = 1, 2, ..., N. \tag{4.1}$$

$$\begin{aligned} y_i &= y_i^* \text{ if } y_i^* > 0, \\ &= 0 \text{ if } y_i^* \leq 0. \end{aligned} \tag{4.2}$$

and

$$x_{1i} = \gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + e_i, \qquad i = 1, 2, ..., N. \tag{4.3}$$

where u_i and e_i are normally distributed with zero means and variances σ_u^2 and σ_e^2, respectively, and correlation coefficient ρ. In their experiment the exogenous variables x_{2i}, z_{1i} and z_{2i} were independently distributed as uniform random variables with the range [-1,1]. The β's and γ's are unknown parameters to be estimated.

Given this model, Wales and Woodland (1980) considered various estimators of the tobit model including the tobit maximum likelihood estimator (MLE), Amemiya's instrumental variables estimator and the nonlinear least squares estimator based on the conditional expectation of the model, i.e., using the observations for which y_i is positive. Further, they considered sample sizes of 1000 and 5000 and an approximate degree of censoring of 30%.

Wales and Woodland (1980) estimated the parameters of the model based on a single replication and observed that the maximum-likelihood estimator was superior to the others. It should be noted that, as discussed in Chapter 2, if the errors are normally distributed the maximum-likelihood estimator is consistent and asymptotically efficient. Hence the results reported by Wales and Woodland (1980) may not be surprising. However, Wales and Woodland (1980) did not consider cases where the assumption of normality did not hold. Further, although useful, their conclusion was based on a single simulation rather than on repeated samples as in the case of a Monte Carlo study.

One of the frequently cited papers with regard to the small sample properties of the estimators of the standard tobit model is that of Paarsch (1984). Paarsch (1984) studied the finite sample properties of tobit model estimators which included the maximum likelihood estimator (MLE), Heckman's two-step estimator (H2S), Powell's (1984) censored least absolute deviations (LAD) estimator and the ordinary least squares estimator based upon the observations for which the dependent variable, y_i, is positive (OLSP).

Paarsch's model was of the form

$$
\begin{aligned}
y_i^* &= \beta_0 + \beta_1 x_{1i} + u_i, \qquad i = 1, 2, ..., N. && (4.4)\\
y_i &= y_i^* \quad \text{if} \quad y_i^* > 0 \\
&= 0 \quad \text{if} \quad y_i^* \leq 0. && (4.5)
\end{aligned}
$$

where x_{1i}'s are the exogenous variables, β_0 and β_1 are the unknown parameters to be estimated and u_i's are independently and identically distributed random variables which are assumed to be normally distributed with mean zero and variance σ^2 so that (4.4)-(4.5) becomes the standard tobit model.

Given this model, Paarsch (1984) investigated the small sample effects of the following on the estimators of the model.

(i) The effects of different distributions of the error term.

(ii) The effects of sample sizes.

(iii) The effects of degree of censoring.

The effects of distributional assumptions of the error term were investigated by considering a variety of distributions, namely, the normal, Laplace and the Cauchy distributions. He considered sample sizes of 50, 100 and 200 representing low, medium and large sizes, respectively, to investigate the effects of sample size. To examine the effects of the degree of censoring he considered 25% and 50% degrees of censoring. Further, the explanatory variable, x_{1i}, was generated from the interval [0,20] by placing each observation equidistantly, the distance depending on the sample size. Finally, while β_1 was fixed at one throughout the experiment, the constant term β_0 was used to adjust the degree of censoring and takes different values depending on the type of the error distribution and the required degree of censoring.

Paarsch (1984), based upon 100 replications of Monte Carlo results, pointed out the following important points.

> Under normally distributed errors, the tobit MLE performed better than both the H2S estimator and Powell's LAD estimator. The latter two appeared to be biased in small samples sizes, although the bias shrinks quickly in large samples.
>
> The LAD estimator appeared to be neither accurate nor stable for sample sizes less than 100 and a high degree of censoring. But, for large sample sizes, the LAD estimator performed much better than the H2S estimator under any of the distributions considered in the study. He also indicated that the LAD estimator performed better than the MLE when the errors are Cauchy. Note that Powell's LAD estimator is computationally burdensome and hence its use in applied research is very limited. Paarsch (1984) noted that three hundred calculations for Powell's LAD estimator took 500-550 CPU minutes compared to 60-70 CPU minutes for three hundred MLE and 30-40 CPU seconds each for the OLSP and H2S estimators.
>
> Finally, but not surprisingly, Paarsch (1984) reported that the least squares estimator based on the observations for which the dependent variable is positive (OLSP) performed poorly in all cases.

Note that Paarsch's results have been very influential in the finite sample studies of the sample selection literature. Other studies which are similarly designed to that in Paarsch's paper include those of Moon (1989) and Nawata (1993) and are discussed later in this Chapter. But first it is very important to examine the outcomes in Paarsch's experiment and their implications for the tobit estimators.

Clearly, the results for the MLE under the correct specification of the model are as anticipated. That is, the MLE performs better under normally distributed error terms. This outcome is quite consistent with our expectations and is in agreement with other similar studies such as those of Wales and Woodland (1980), Flood (1985) and Moon (1989).

However, what is a most interesting result in Paarsch's paper is that the MLE is preferable to both the H2S and the LAD estimators under the Laplace distribution. Surprisingly, this result indicates the robustness of the MLE to non-normality of the error terms. This is an outcome which is in contrast to previous claims such as those of Goldberger (1980) and Arabmazar and Schmidt (1982) who indicated that the MLE is not robust to the violations of assumptions about the error term. Specifically, these studies indicated that the MLE is not only inefficient but also inconsistent if the assumption of normality does not hold. Obviously, there seems to be quite a significant deviation between these studies and those of Paarsch's (1984); this deviation suggests the need for further research regarding the properties of the MLE of the tobit model under non-normal distributions.

On the other hand, Paarsch's paper depicts that the MLE under the Cauchy distribution did not perform as well as under the Laplace distribution and, in this case, is particularly inferior compared to the semi-parametric LAD estimator. The reasons for these substantial differences in the finite sample performance of the MLE are not clearly explained. One possible explanation could be that the MLE did not perform well because the Cauchy distribution, unlike the Laplace, does not have finite moments.

It would be interesting to see whether the results for the Cauchy distribution hold for other similar distributions, but with finite first and second moments.

Another important aspect which is worth discussing in a study of this nature is the design of the experiment. It is clear that the outcome of any Monte Carlo experiment is likely to be influenced by the design of its experiment, although the significance of this influence may vary from one experiment to the other. One of the problems with Paarsch's (1984) results is related to his experimental design. That is, some of the outcomes in Paarsch's paper are attributed to a critically deficient experimental design. For example, one of the estimators included in the experiment, the H2S estimator, is known to be very sensitive to the multicollinearity between the explanatory variables and the estimated inverse of Mill's ratio (see Chapters 2 and 3). This implies that an experimental design which yields a high correlation between the explanatory variables and the estimated hazard function, unless specifically designed for that purpose, is likely to contribute to negative outcomes for this estimator. That is exactly what has happened in Paarsch's experiments. A quick examination of his experimental design reveals that almost all setups in the experiment yield a correlation between the explanatory variable, x_{1i}, and the estimated hazard function, $\hat{\lambda}(x'_{1i}\hat{\alpha})$ which is close to negative one (i.e., $|\hat{\rho}_{x\lambda}| \simeq 1$). Thus, it is not surprising that the H2S estimator performed so badly in all cases as compared to both the MLE and LAD estimators. For example, as reported in Paarsch (1984, pp. 210-212), given the normal distribution and 25% degree of censoring, the Monte Carlo variances of the coefficients for the H2S estimator are reported to be up to 75 times larger than those of the MLE estimates and more than 30 times larger than those of the LAD estimators for the large (200) sample size. The difference between the Monte Carlo variances of the MLE and the H2S estimators increased dramatically to several hundreds for the smaller sample sizes (i.e., for sample sizes of 50 or 100) and 50% degree of censoring. The main reason for this is that there exists almost exact collinearity between the explanatory variable and the estimated inverse of Mill's ratio in the experiment. However, no explanation has been provided for such unbelievably large differences. This, however, is not to argue in favour of the H2S estimator but to indicate the severity and the consequences of biases that may be introduced with or without the knowledge of the researcher in any experiment. Further, Paarsch (1984) used only 100 replications which may indicate that the accuracy (quality) of the results of the experiment may be questionable. Finally, although the above discussion focusses on specific experimental results in the literature, it is possible similar comments can be made about many other studies [see Leung and Yu (1994) for further comments in this direction]. This implies that an important consideration should be given to the design of any experiment before one realizes the outcomes.

Flood (1985) compared the small sample properties of the maximum likelihood estimator (MLE) and the corrected least squares estimator (COLS), proposed by Greene (1981a, 1983) of the tobit model. He considered a model of the form (4.4)-(4.5) and studied the effects of (i) non-normally distributed error terms, (ii) sample size, (iii) degree of censoring, and (iv) normally distributed exogenous variables. The non-normal distributions were generated from a mixture of normally distributed random variables.

Flood's (1985) study involved actual data from the 1975-1976 survey of time allocation among American adults containing a total number of 766 observations. Flood's paper depicts the following point:

Similar to the results of Wales and Woodland (1980) and Paarsch (1984), when the errors are normally distributed, the MLE is superior to the COLS estimator. Bias appears to be a problem for the MLE only for the smallest sample sizes. However, the COLS is biased even when the sample size increases. Further, there exists some evidence that the bias in COLS increases with the increase in the proportion of limit observations (degree of censoring) in the sample. It is important, however, to note that most of the bias in the COLS disappears when the exogenous variables are generated from a normal distribution, which is very unlikely to be the case in applied research.

There appears to be some bias for the MLE under non-normality of the error terms. However, the COLS estimates are not sensitive to the particular choice of distribution whether normal or non-normal. The COLS estimates appear to be superior to those of the MLE for highly skewed distribution.

More recent studies of the small sample properties of tobit estimators include those of Moon (1989) and Nawata (1993, 1994). Moon's (1989) paper was concerned with a comparison of the semi-parametric estimators of the tobit model. These estimators include Powell's least absolute deviations (LAD) estimator, the Buckley-James estimator, Horowitz's distribution-free least squares estimator and its conditional version which was proposed by the author. The tobit MLE was also included in the comparison. Similar to that of Paarsch (1984), Moon (1989) considered three distributions, namely, the normal, Laplace and Cauchy distributions for the error term. In general, his design of the experiment was the same as that of Paarsch (1984) except that the number of replications were increased from 100 to 500.

Interestingly, Moon's paper also depicts that the MLE performs well under the Laplace distribution when compared to the semi-parametric estimators of the model. Obviously, the MLE is the best under normality of the error terms.

On the other hand, the LAD estimator seems to perform better under the Cauchy distribution. However, bias appears to be a problem for small sample sizes and higher degrees of censoring. These results are quite similar to those of Paarsch (1984). Another, more serious problem with the semi-parametric estimators is their computational difficulty which makes them undesirable in applied research.

More recently, Nawata (1993, 1994) studied the finite sample properties of the estimators of the tobit model but with a slightly different objective than most of the papers discussed above. In particular, Nawata's (1993, 1994) papers are aimed at investigating the effects of correlation between the explanatory variables and the estimated inverse of Mill's ratio on the performance of the H2S estimator. His experiments, also similar in design to that of Paarsch (1984), reveal that the poor finite samples performance of the H2S estimator (for example, Wales and Woodland (1980), Nelson (1984), Paarsch (1984)) can, in most cases, be attributed to the existence of a high correlation between the explanatory variables and the estimated hazard function, $\hat{\lambda}(.)$. Specifically, Nawata (1993) pointed out that there almost always exists a high (and negative) correlation between the explanatory variables and the hazard function and that this is the reason why the H2S estimator is less efficient.

Other related Monte Carlo studies in the sample selection literature include, among others, those of Hartman (1991), Manning, Duan and Rogers (1987) and

Leung and Yu (1994). Although similar, these studies involve sample selection models involving simultaneous equations in which the H2S estimator plays an important role in the estimation of the parameters of the model. In general, these studies also indicate that the H2S estimator can be inefficient in finite samples. In particular, as discussed earlier in this Chapter, Leung and Yu (1994) argued that the negative results regarding the H2S estimator which reported by many Monte Carlo studies including those of Paarsch (1984) and Manning, Duan and Rogers (1987) are usually exaggerated by a critical design problem that produces a serious multicollinearity problem.

4.3 Summary and Conclusions

In recent years, a number of analytical methods and econometric techniques have been developed to address the problems of estimation and/or test statistics of the censored regression (tobit) model. However, most of these studies are concerned with the asymptotic properties of the various estimators or test statistics of the model. On the other hand, there exists only a few studies which are concerned with the small sample properties of the various estimators of the model; this study aims to contribute in this direction.

In this Chapter, we discussed the relevant small sample studies and their likely implications in applied research. Those estimators of the model which have been studied most frequently in the finite sample studies include the maximum likelihood estimator (MLE), the Heckman's two-step (H2S) estimator and the semi-parametric least absolute deviations (LAD) estimator. The relative performance of these estimators were compared under the normal and non-normal distributions of the error term as well as different sample sizes and degrees of censoring. The main results and their implications may be summarized as follows.

> As anticipated, the MLE is superior to all other estimators of the model, provided that the errors are normally distributed. However, under non-normality of the error terms, the MLE appears to provide mixed results. For instance, under the Laplace distribution, the MLE performed better than both the H2S and LAD estimators in two separate studies. This is quite surprising because the MLE is not expected to be robust under non-normality of the error terms. However, the situation is different under the Cauchy distribution where the MLE is not as good as it is under the normal or Laplace distributions. Further, there is very little evidence regarding its performance under skewed distributions of the error terms.
>
> On the other hand, although the LAD estimator appears to do well under the Cauchy distribution as compared to the MLE, it has other serious consequences. It is inefficient not only when the model is correctly specified (i.e., under normality) but also under the Laplace distribution. More importantly, it is neither reliable nor stable for small samples and higher degrees of censoring. It is also computationally burdensome even for the simplest cases, making it very unattractive for applied research.
>
> The H2S estimator appears to be relatively inefficient compared to the MLE in

almost all cases. However, it is also important to note that most of the inefficiency of the results have been inflated by serious design problems. A design problem which produces serious correlation between the explanatory variables and the estimated inverse of Mill's ratio.

In general, although there are many estimators of the model, the evidence on small sample properties of many of the estimators is either very limited or non-existent. Further, almost all of the studies discussed in this Chapter have been obtained based on a limited number of Monte Carlo replications (samples) which makes the accuracy of many of the outcomes questionable.

Finally, it is important to note that almost all the finite sample studies have been concerned with the bias and efficiency of the point estimates of the parameters of the model. In other words, none of these studies have actually attempted to investigate the performance of the various estimators when they are used for hypothesis testing and/or confidence interval estimation for the parameters of the model. Note that many researchers are more interested in reliability as reflected by interval estimates than in single point estimates. This study considers most of the points discussed above and is aimed to narrow the gap between the asymptotic results and the finite sample properties of estimators of the standard tobit model.

Chapter 5

The Design of the Monte Carlo Experiment

5.1 Introduction

The Monte Carlo method is a scientific tool which is used to solve many complex problems which may or may not be possible to solve analytically. Although the term *Monte Carlo* is said to have been introduced by Metropolis and Ulam (1949), its use has become increasingly important in recent years with advances in computer technology and power. This is because the Monte Carlo approach is relatively capital intensive, requiring a great deal of computer power and time. The Monte Carlo methodology is not specific to econometrics or statistics and is widely used in other disciplines of science. For example, it is used to evaluate complex multidimensional integrals and/or to approximate certain integral equations in physics. Useful materials which are related to the Monte Carlo methodology include those of Hammersley and

Handscomb (1964), Rubinstein (1981), Kalos and Whitlok (1986) and Lewis and Orav (1989).

In econometrics, Monte Carlo methods are used frequently to examine, among other things, the finite sample properties of various estimators (or an estimator) under some prespecified error distributions, the size of a test statistic under the null hypothesis and the power of a test statistic under some specified alternative hypothesis. Hendry (1984) provides useful discussion on the use of Monte Carlo experimentation in econometrics [see also Smith (1973) and Sowey (1973)]. It is important, however, to note that this Chapter is not concerned with Monte Carlo methodology in general but with the design of the experiment used in this study.

It is clearly known that the design of the experiment is an integral part of a Monte Carlo study and that it may influence the outcomes of the experiment. That is, the outcomes of a Monte Carlo experiment depend on the specifications of the model (models) being studied and the various specific details of the experimental design which are usually determined by the experimenter. This implies that, like many other scientific tools, the Monte Carlo experiment also has its limitations. The two most frequently cited criticisms of Monte Carlo studies are the problems of specificity and precision [see Hendry (1984) for more discussion]. However, the impact of these problems, although not entirely avoidable, can be reduced by examining closely the various components of the experimental design as well as by learning from previous studies, some of which are discussed in the preceding Chapter. Thus, the objective of this Chapter is to provide the details of the design of the experiment of the study, a design which tries to keep the various problems to a minimum.

Specifically, Section 5.2 begins with the specific form of the econometric model to be investigated in the Monte Carlo experiment. It then details the various distributional assumptions, the levels of censoring and the sample sizes considered in the experiment. The data generation and Monte Carlo estimation process of the experiment is discussed in Section 5.3. This section includes, the specification of the data generation mechanism which is a special case of the econometric model, the generation of the exogenous variables of the model, the determination of the parameters (true values) and the generation of the random variates in the experiment. Further important items such as the number of replications, output statistics to be computed and other related matters are included in this section. Section 5.4 concludes.

5.2 The Specification of the Model

The specific model to be investigated in this study is of the form:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \quad i = 1, ..., N. \tag{5.1}$$

$$\begin{aligned} y_i &= y_i^* \text{ if } y_i^* > 0 \\ &= 0 \text{ if } y_i^* \leq 0 \end{aligned} \tag{5.2}$$

where X=(1, x_1, x_2) is an (Nx3) matrix of observations containing a column vector of 1's corresponding to the constant term and observations on the explanatory variables x_1 and x_2,

y^* , the latent variable, is an (Nx1) vector which is assumed to be observed only if it is positive,

y is an (Nx1) vector of observations on the dependent variable consisting of N_1 positive (non-limit) observations corresponding to the positive values of y^* and $N_0 = N - N_1$ zero (limit) observations,

β=(β_0, β_1, β_2)$'$ is a (3x1) vector of unknown parameters to be estimated, and u is an (Nx1) vector of identically and independently distributed random errors with mean 0 and variance σ^2. Note that if u_i's are normally distributed model (5.1)-(5.2) becomes the standard tobit model defined by (2.1)-(2.2).

As discussed in Chapter 4, most of the Monte Carlo studies of the estimators of the model have been based on a relatively simpler model which contains one explanatory variable. Thus the model in (5.1)-(5.2) can be considered slightly more general when compared to these studies. The model is also convenient if one needs to study, say, the effects of correlation or misspecification.

Given this model, the objectives of the Monte Carlo experiment are to investigate the effects of the following on the estimators which we discussed in Chapter 2.

1. The effects of changes in distributional assumptions for the disturbance term. That is, to investigate, more clearly, the effects of violating the assumptions of the error distribution of the model.
2. The effects of the degree of censoring.
3. The effects of sample size.
4. The effects of multicollinearity between the explanatory variables. More specifically, the aim is to investigate the effects of correlation between the X's and the estimated inverse of Mill's ratio on the performance of the H2S and the 3SE estimators. A more elaborate discussion will be provided later in the respective Chapters of the study.

The following Sections of this Chapter provide the specific details considered in the Monte Carlo experiment in order to achieve these objectives.

5.2.1 The Effects of Distributional Assumptions

In almost all cases the assumption of normally distributed error terms is used to derive the asymptotic properties of the various estimators of the tobit model. However, in reality, this assumption may not hold. Therefore, it is the purpose of this study to investigate the performance of the estimators under several distributions for the disturbances of the model. In other words, the disturbances of the model are first generated from a presumed (known) error distribution and then the model is estimated by assuming (pretending) that the errors are normal.

In order to do this we consider three major distributional assumptions in our experiment. The details are given as follows:

1. Our first objective is to evaluate the finite sample performance of the different estimators when the assumption of normality of disturbances holds. This is important for two reasons: (i) for the comparison of the various estimators when the assumption is actually true and (ii) it can be used as a basis for comparison between estimators when the assumption of normality does not hold. To attain this, the standard normal distribution is considered, i.e., $u_i \sim N(0, 1)$.

2. The second major objective is to investigate symmetric departures from normality. That is, a situation where the error terms are symmetric but have a fat (wide) tailed distribution. To pursue this objective the students'-t distribution with three degrees of freedom is considered and its effects investigated.

3. The third main purpose is to examine the finite sample performance of the estimators when the disturbances have a skewed distribution. One way of achieving

this objective is by considering a skewed distribution such as the Chi-square distribution. In this particular study we considered a chi-square distribution with four degrees of freedom.

Note that the distributional assumptions given in items (2) and (3), that is, the students'-t distribution and the chi-square distribution, are designed to represent two major departures from the usual assumption, these departures being (i) symmetric but fat tailed and (ii) skewed distributions. The usual assumption is also represented by the standard normal distribution given by item (1) above.

In general, these distributions represent, in terms of shape, the possible violations of the assumptions of the error term in applied research as well as a situation where the assumption actually holds. The non-normal distributions considered in this study are also different from those of Paarsch (1984) and Moon (1989) who used the Laplace and the Cauchy distributions in their experiment. However, unlike this study, their experiments involve only symmetric distributions for the error term. It is interesting to see if Cauchy results from others also hold for the t-distribution with three degrees of freedom, where first and second moments are finite. It is also important to note that, although many empirical observations (economic, social, etc.) may follow any of these or similar distributional structures (assumptions), it is also true that any study of this nature is limited by the problem of specificity [see Hendry (1984)].

5.2.2 The Effects of the Degree of Censoring

The degree of censoring indicates the percentage (ratio) of limit (zero) observations to the total number of observations on the explanatory variable, y_i. For example, a 25% degree of censoring means that out of, say, 100 observations only 25 of them correspond to zero values on y_i and the remaining 75 observations are positive.

The effects of the degree of censoring may vary from one estimator to another and across various distributions and sample sizes. But in general, where the severity of the degree of censoring influences performance, it is useful to examine changes associated with an increase in the number of limit observations. The effects of the degree of censoring in the performance of the estimators of the tobit model are therefore examined at three levels, namely, 25 percent, 50 percent and 75 percent degrees of censoring. These levels of degree of censoring are chosen to represent a wide range of economic or other data to which the model may be applied.

Another point which is worth mentioning here is that, as discussed in Chapter 3, the correlation between the explanatory variables and the inverse of Mill's ratio increases with higher levels of the degree of censoring and hence affects the two-step estimators of the model. In other words, a change in the level of censoring also implies a change in the correlation between the explanatory variables and the estimated Mill's ratio. Therefore, taking this into consideration, it becomes very important to study these effects using a wide range of censoring levels.

5.2.3 The Effects of Sample Size

One of the fundamental objectives of finite sample studies of this nature is to examine the results for the estimators based on sample sizes which are likely to be representative of real life situations. The choice of these sample sizes therefore may depend, among other things, on the type of the model and its use in applied research, technical difficulties involved in the experiment, the comparability of the expected outcomes to other related studies, availability of resources and so on.

Given this, three levels of sample sizes are considered in order to examine the finite sample properties of the estimators of the model. These levels include the sample sizes of 100, 200 and 400 which correspond to small, medium and large sizes, respectively. In particular, it is important to note that the choice of these sizes is made by taking into consideration factors such as the degree of censoring. For example, given the small sample size of 100, the actual number of positive (non-limit) observations on the explanatory variable y_i reduces to 25, 50 and 75 observations, respectively, for 75 percent, 50 percent and 25 percent degrees of censoring. Similar interpretations can be made for the medium and large sample sizes. In general, this experiment involves sample sizes ranging from a minimum of 25 to a maximum of 375 non-limit observations.

5.3 Data Generation and Estimation Process

Given the specification of the model and the various assumptions discussed in the preceding sections, this Section presents the details involved in the data generation and Monte Carlo estimation process of the experiment. This process may include the determination of the explanatory variables, the true values for the parameters, the generation of the random variates (i.e., the standard normal, the students'-t and the Chi-square distributed random variates) and other things which may be relevant for the data generation process of the experiment. This is followed by a discussion on the Monte Carlo estimation procedure which may include the determination of the number of runs (replications) used in the experiment, parameters to be estimated, outputs (or statistics) to be obtained for the comparison of the various estimators and related discussions.

5.3.1 The Data Generation Process

The data generation process of this experiment is based on the model defined in (5.1) which is given by

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \qquad i = 1, ..., N. \tag{5.3}$$

where the y_i^*'s, the X's, the β's and the u_i's are as defined in (5.1)-(5.2).

This is a well known linear regression model with two explanatory variables, three unknown parameters, β's, and a random disturbance term. The crucial point which makes model (5.3) different from the classical linear regression model, as defined in (5.1)-(5.2), is that the dependent variable y_i^* is assumed to be observed only when it is positive.

Given (5.3), the various components of the model are determined as follows:

(i) One of the important components in the data generation process of a Monte Carlo experiment is the generation of the explanatory variables in the model. There are generally two main ways of obtaining these variables, each of which have their own advantages and disadvantages. One is to use a real data set which contains all the variables and sample the variables repeatedly. The advantage of this kind of procedure is that it enables one to see the applicability of the different estimators in real life situations. However, one of the drawbacks of such an experiment is that it is sometimes very difficult to identify whether some of the differences that may be observed between the estimators are a result of the specific data set, which is not controlled by the researcher, or due to other reasons. Of course, another major problem is to get an actual economic or other data set which is large enough to conduct a Monte Carlo experiment.

The second alternative is to use hypothetical (or computer generated) data. Although the data may not be generally representative of real life situations, this procedure allows the researcher to identify the reasons that may have caused any differences between the estimators. This is because the researcher has relatively full knowledge about the experiment. It is also possible to use both actual and computer generated data in the experiment. But, again, the availability of real data and time and resource constraints are the major problems.

In this study, we have chosen the second alternative, i.e., we used computer generated variables because of their flexibility and mainly due to lack of a large actual data set. The variables are generated as follows:

The observations on the explanatory variable x_{1i}, $i = 1, ..., N$, are generated from the interval [0,4] equidistantly where the distance depends on the sample size.

Similarly, the observations on the second explanatory variable, x_{2i}, $i = 1, ..., N$, are generated uniformly from the interval [-1,1] and independently from x_{1i}. Once the values of the explanatory variables are determined from their respective intervals they remain the same throughout the experiment.

Note that there is one important point which should be taken into consideration while generating the explanatory variables of the model. That is, that some of the estimators of the tobit model are likely to be sensitive to multicollinearity (e.g., H2S). Thus, the explanatory variables are generated so that this is not always the case. As discussed in Chapter 4, this is useful in terms of avoiding the bias against some of the estimators as a result of the design of the experiment. On the other hand, a different data generation mechanism will be discussed at a later stage that will be used to investigate the effects of multicollinearity between the explanatory variables and the inverse of Mill's ratio on the performance of the estimators [see Chapter 8].

(ii) The parameters β_1 and β_2 are set to be equal to one, (i.e., $\beta_1 = \beta_2 = 1$) in all cases.

(iii) The degree of censoring is determined by varying the value of β_0 which takes different values depending on the particular level of degree of censoring and type of distribution. For example, given (i) and (ii), if the disturbances are normally distributed with mean zero and variance equal to one, i.e., the standard normal distribution, then a value of $\beta_0 = -0.75$ yields approximately 25 percent degree of censoring. Similarly, the approximate levels of degree of censoring of 50 percent and 75 percent can be obtained by setting β_0 equal to -2.00 and -3.25, respectively. Some preliminary experiments were conducted to determine the value of β_0 for the three types of distributions and degrees of censoring.

In general, items (i) to (iii) are interrelated and determine the systematic component of the data generation process in the experiment. The remaining component, the generation of the random variable, u_i in (5.3) is discussed below.

5.3.2 The Generation of Random Variates

As we have discussed in the preceding sections, this experiment involves three distributional assumptions for the random disturbances of the model, namely, the normal, the students'-t and the chi-square distribution. Note that, hereafter, the names normal, students'-t and chi-square distributions in this particular design refer, respectively, to the standard normal distribution, students'-t distribution with 3 degrees of freedom and the chi-square distribution with 4 degrees of freedom. The random variates, more accurately referred to as pseudo-random variates, must therefore be generated from their respective distributions.

Note that there are several methods of generating random variates and many algorithms have been suggested based on the various methods [see for example, Rubinstein (1981), Lewis and Orav (1980), and Forsythe (1972)]. These, however, are not all equal in terms of efficiency and quality and therefore are subject to the choice of the researcher. However, in recent years the choice of the random generators has become less important because of the availability of statistical/econometric packages such as SHAZAM which incorporate random variable generators which are widely used and well tested. Below, we provide a brief discussion on the random variate generation for the three distributions listed above.

1. **The Normal Distribution**

Independent standard normal variates are derived from independent standard uniform variates. These random variates can be obtained using several procedures. One way of generating the random variates is using the inverse transformation method. This method requires that the probability distribution function of the random variable must be invertible. However, for many probability distributions such as the normal distribution this is either impossible or difficult. Further, the inverse transformation method is not also necessarily the most efficient [see Rubinstein (1981), Knuth (1981) and Lewis and Orav (1989)].

Alternatively, one of the widely applied methods is that of Box and Muller (1958) which is used to obtain a pair of independent normal random variates from a pair of independent uniform random variates as follows.

Let U_1 and U_2 be independent random variates from U(0,1), then the variates

$$Z_1 = (-2\ln U_1)^{1/2}\cos 2\pi U_2 \tag{5.4}$$

$$Z_2 = (-2\ln U_1)^{1/2}\sin 2\pi U_2 \tag{5.5}$$

are independent standard normal variates. Thus one can obtain U_1 and U_2 from U(0,1) and compute Z_1 and Z_2 simultaneously by substituting U_1 and U_2 in the system of equations (5.4)-(5.5). However, one major problem of the Box-Muller procedure is it heavily relies on the independence of U_1 and U_2. Thus, if the uniform random variables are not based on a good uniform random generator, the properties of the normal random variates will suffer.

Another frequently and widely applied procedure for generating random variates is based on the acceptance-rejection method. This method is originally due to von Neumann (1951) and was later generalized by Forsythe (1972). Further extensions of the method for the normal distribution are provided by Ahrens and Dieter (1973). The method consists of sampling a random variate from a certain probability distribution and then subjecting it to a test for acceptance [see also Rubinstein (1981)]. This method is quite efficient and frequently used in applied research. For example, the random number generator in SHAZAM is based on the acceptance-rejection method. Specifically, the algorithm used to generate the normal random variates is based on an improved version proposed by Brent (1974). As noted by Brent (1974, p.705), the algorithm is exact, relatively efficient and is preferable to methods which depend on central limit theorems or use approximations to the inverse of the distribution function. Further, the chi-square tests for independence of pairs of successive random variates also show that they are highly independent at a 5 percent significant level.

Once the normal random variates are generated, random variates from any other distributions, including the chi-square and the students'-t, can be generated using fundamental statistical relationships as follows [see Rubinstein (1981), Leemis (1986)].

2. **The Chi-Square Distribution**

Let $Z_1, ..., Z_k$ be standard normal random variates from N(0,1). Then

$$Y = \sum_{i=1}^{k} Z_i^2 \tag{5.6}$$

has the chi-square distribution with k degrees of freedom and is denoted by $\chi^2_{(k)}$. It has mean k and variance $2k$. That is, the sum of the squares of independent standard normal variates has a chi-square distribution with degrees of freedom equal to the number of terms in the sum.

In particular, if Z_1, Z_2, Z_3 and Z_4 are independent standard normal variates then $Y = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2$ has a Chi-square distribution with four degrees of freedom with mean 4 and variance equal to 8. Thus, once we generate the Z's from (1), Y can be easily generated using (5.6). Further, the location and scale of the chi-square random variates is changed so that their mean and variance become zero and one, respectively. This is done so that the effects of the distributions on the estimators of the model can be attributed to their shapes not to variations in the mean and variance of the distribution.

3. **The Students'-t Distribution**

   Similarly, the random variates from a students'-t distribution are derived using the following statistical relationship.

   Let Z have a standard normal distribution, let Y have a chi-square distribution with k degrees of freedom, and let Z and Y be independent, then

$$X = \frac{Z}{\sqrt{(Y/k)}} \tag{5.7}$$

   has a students'-t distribution with k degrees of freedom, denoted by $t_{(k)}$, and has mean zero and variance given by $k/(k-2)$. Thus we generate X by first generating Z and Y using definitions (1) and (2), respectively. Again, the students'-t distributed errors are rescaled so that their variance is equal to one.

4. Finally, since the explanatory variables are fixed, the coefficient of determination, R^2, is controlled by setting the variance of the disturbance term to a

certain constant which in this case is set to be equal to one, i.e., $\sigma^2 = 1$. This also implies that the signal to noise ratio is constant, the signal being the variation of the dependent variable due to variations of the systematic component and the noise represents the variance of the error term of the model.

Note that it is also important to determine the magnitude of the variance of the error term in relation to the variations in the explanatory variables. One of the reasons for this is that if the variation in the explanatory variables is large compared to the variance of the error term then the correlation between the explanatory variables and the inverse of Mill's ratio becomes high. Similar problems arise if the variance of the error term is large relative to small variations in the explanatory variables [see Leung and Yu (1994)].

Given the data generation process discussed above, the correlations that exist between the explanatory variables and the inverse of Mill's ratio are illustrated as follows. For example, the correlation matrices below summarize the simple correlations between the variables, x_1, x_2 and $\lambda(.)$ for the three levels of censoring, given normally distributed errors and a sample size of 200.

Degree of censoring=25%

	x_1	x_2	$\lambda(.)$
x_1	1.00	−0.12	−0.74
x_2	−0.12	1.00	0.29
$\lambda(.)$	−0.74	0.29	1.00

(5.8)

Degree of censoring=50%

	x_1	x_2	$\lambda(.)$
x_1	1.00	−0.30	−0.81
x_2	−0.30	1.00	0.33
$\lambda(.)$	−0.81	0.33	1.00

(5.9)

Degree of censoring=75%

	x_1	x_2	$\lambda(.)$
x_1	1.00	−0.10	−0.85
x_2	−0.10	1.00	0.49
$\lambda(.)$	−0.85	0.49	1.00

(5.10)

As shown above, there is very little correlation between the explanatory variables x_1 and x_2. Similarly, the correlation between x_2 and $\lambda(.)$ is quite small. However, the correlation between x_1 and $\lambda(.)$ is relatively larger in magnitude and varies from around -0.75 for the low level of censoring to -0.85 for the high level of censoring. This indicates that even if no correlation exists between x_1 and x_2 the correlation between the X's and the inverse of Mill's ratio can be substantial. The above correlations between the variables remain around the same level for all samples and errors, except for the small sample size (100) and high degree of censoring (75%) where the correlation between x_1 and $\lambda(.)$ is around -0.95.

Note that the above matrices also reveal that the higher the degree of censoring the bigger (stronger) is the correlation between the explanatory variables and the inverse of the Mill's ratio. However, in practice, this may not be always the case. That is, there may be situations where the degree of censoring is high but with a low or moderate correlation between the explanatory variables. Alternatively, the correlations between the variables can be high while the degree of censoring is low. These and other related questions will be examined further by considering a modified experimental design later in this study (see Chapter 8).

5.3.3 The Estimation (Monte Carlo) Process

Two important considerations in the Monte Carlo experiment are the number of replications (samples) used in the experiment and the output statistics to be computed from the experiment based on these replications.

In general, the precision of the Monte Carlo results depends on the number of replications used in the experiment. Given the data generation process discussed above, a preliminary experiment was conducted based on 1000, 2000,..., 5000 replications. It was observed that there was no significant difference up to three decimal points between the results for 3000 and 5000 replications. Thus, given the large amount of experiments required in this study, the number of replications is determined to be 3000. Each experiment in this study involves combinations of the following:

(i) Sample size: 100, 200, 400

(ii) Degree of Censoring: 25%, 50%, 75%

(iii) Type of Distribution: Normal, Student's-t, Chi-Square

Another important aspect of the estimation process is concerned with final outputs (usually averages from the 3000 replications) to be computed to compare the various estimators. That is, the types of statistics used to compare the relative performance of the various estimators of the model. In general, the computation of any output statistics from a Monte Carlo experiment depends on the purpose and type of the study. Next, we present the output statistics that will be computed from the Monte Carlo experiment of this study.

5.3.4 Output Statistics

This study involves a wide ranging comparison between estimators and includes statistics relating to both point estimates (such as bias and efficiency) and to hypotheses testing and/or confidence intervals. Some of the output statistics computed are listed below.

Let $\hat{\beta}_k$ be an estimator of β_k $(k = 0, 1, 2.)$ and $M = 3000$ be the number of replications (or samples) in the experiment. Then the following statistics (outputs) are used to compare the finite sample properties of the various estimators in the experiment. These outputs are frequently used in studies of this nature.

1. **Estimated Mean (EM):**

   The first step of the procedure is to compute the estimated value of the mean for $\hat{\beta}_k$ from the 3000 samples. It is defined as

   $$EM(\hat{\beta}_k) = \hat{E}(\hat{\beta}_k) = \sum_{m=1}^{M} \hat{\beta}_{km}/M \tag{5.11}$$

   where $\hat{\beta}_{km}$ is the estimate of β_k from the m^{th} sample.

   Thus, the estimated mean (EM) is the average of the estimates obtained from the 3000 samples. Given this, the following statistics are used to compare the various estimators.

2. **Bias (BIAS):**

The bias is used to measure the difference between the true value of β_k and its estimated mean and is defined as

$$\widehat{BIAS}(\hat{\beta}_k) = \hat{E}(\hat{\beta}_k) - \beta_k \tag{5.12}$$

where $\hat{E}(\hat{\beta}_k)$ is as defined in (5.11).

If $\hat{\beta}_k$ is an unbiased estimator of β_k, then the bias is zero, i.e., $E(\hat{\beta}_k) = \beta_k$. However, in finite samples, this may not be always the case. Thus, the magnitude of the bias is important in comparing one estimator with another.

3. **Variance of $\hat{\beta}_k$, Var($\hat{\beta}_k$):**

The variance of $\hat{\beta}_k$ measures the spread of the distribution of $\hat{\beta}_k$ about its mean. Its Monte Carlo-based estimate is defined by

$$\begin{aligned} \widehat{Var}(\hat{\beta}_k) &= \hat{E}[\hat{\beta}_k - \hat{E}(\hat{\beta}_k)]^2 \\ &= \sum_{m=1}^{M} [\hat{\beta}_{km} - \hat{E}(\hat{\beta}_k)]^2 / M \end{aligned} \tag{5.13}$$

and its square-root, the **standard error** measures the precision of $\hat{\beta}_k$. That is,

$$SE(\hat{\beta}_k) = \sqrt{\sum_{m=1}^{M} [\hat{\beta}_{km} - \hat{E}(\hat{\beta}_k)]^2 / M} \tag{5.14}$$

The smaller the variance of the sampling distribution of $\hat{\beta}_k$, the greater is the precision of the estimator. That is, the chance of a sample estimate lying within some specified interval about the true value is greater.

4. **Mean Square Error (MSE):**

An important statistic which is frequently used for comparison of various estimators (or models) is the mean square error. The MSE measures the spread of the estimates, $\hat{\beta}_{km}$'s, around the true values, β_k. The MSE is given by

$$\begin{aligned} MSE(\hat{\beta}_k) &= E[\hat{\beta}_k - \beta_k]^2 \\ &= E\{[\hat{\beta}_k - E(\hat{\beta}_k)] + [E(\hat{\beta}_k) - \beta_k]\}^2 \\ &= E[\hat{\beta}_k - E(\hat{\beta}_k)]^2 + [E(\hat{\beta}_k) - \beta_k]^2 \\ & \quad 2E[\hat{\beta}_k - E(\hat{\beta}_k)][E(\hat{\beta}_k) - \beta_k] \end{aligned} \tag{5.15}$$

since the cross product is zero, we have

$$MSE(\hat{\beta}_k) = Var(\hat{\beta}_k) + [BIAS(\hat{\beta}_k)]^2 \tag{5.16}$$

where Monte Carlo estimates of the variance, $Var(\hat{\beta}_k)$, and the bias, $BIAS(\hat{\beta}_k)$, are computed from (5.13) and (5.12), respectively.

Alternatively, the MSE can be estimated directly as

$$\widehat{MSE}(\hat{\beta}_k) = \sum_{m=1}^{M} (\hat{\beta}_{km} - \beta_k)^2 / M \tag{5.17}$$

and the root mean square error (RMSE) is defined by

$$R\widehat{MSE}(\hat{\beta}_k) = \sqrt{\sum_{m=1}^{M} (\hat{\beta}_{km} - \beta_k)^2 / M} \tag{5.18}$$

Note that one can use either the MSE or RMSE for comparison purposes. The RMSE is frequently employed because it uses the same unit of measurement as the mean or standard error. The RMSE criteria is used in this study.

As shown above, the MSE (RMSE) takes into account both the variability and bias of the estimator which makes it often a preferable yardstick in the comparison of estimators or models. The smaller the MSE (RMSE) the better is the estimator.

5. **Comparison of Variances and Hypothesis Testing:**

One of the major objectives of this study is to compare the various estimators in terms of their performance for hypothesis testing and/or construction of confidence intervals for the coefficients of the model. Thus, the following related outputs are computed in addition to those discussed above,

(i) Asymptotic variances of the estimators are computed by inserting the actual (known) values into the their respective asymptotic expressions for the covariance matrices provided in Chapters 2 and 3 of this study. For example, the asymptotic variances of the maximum likelihood estimators are computed by substituting the actual values in equation (2.21) of Chapter 2, i.e, the inverse of the information matrix. The asymptotic variances of other estimators are computed in a similar way using their respective covariance matrices.

These asymptotic variances are then compared with their corresponding true (Monte Carlo) variances obtained by using (5.13). The main purpose of this comparison is to examine whether the asymptotic variances are accurate (or good) estimates of the finite sample variances of the estimators. This comparison has important implications in applied research. That is, assuming asymptotic variance expressions are used to compute standard errors, asymptotic variances which are larger than their corresponding variances generally imply confidence intervals which are wider than they should be. On the other hand, asymptotic variances which are smaller than their true variances imply confidence intervals which are narrower (or overprecise) than the desired level; or the probability of rejecting a true hypothesis becomes larger than would be expected. These implications are studied in detail using the following statistics.

(ii) To examine the implications of the asymptotic variances of the estimators for hypothesis testing, we test the hypotheses:

$$H_0 : \beta_k = 1 \quad (5.19)$$

$$H_1 : \beta_k \neq 1, \quad k = 1, 2. \quad (5.20)$$

To test the hypotheses we use the test statistic:

$$t = \frac{\hat{\beta}_k - 1}{\sqrt{\overline{Var}(\hat{\beta}_k)}} \quad (5.21)$$

where, under the null hypothesis, the statistic t is asymptotically distributed as a standard normal random variable, $\hat{\beta}_k$ is the sample estimate of β_k and $\overline{Var}(\hat{\beta}_k)$ is an estimate of the variance of $\hat{\beta}_k$ based on the asymptotic variance expression.

It is important to note that the estimated variance of $\hat{\beta}_k$, $\overline{Var}(\hat{\beta}_k)$, is obtained by substituting sample estimates into the diagonal elements of the asymptotic expression of the covariance matrices of the estimators. That is, while the asymptotic variances discussed in (i) are obtained using the actual values, the variance of $\hat{\beta}_k$ in (5.21) is obtained by substituting the sample estimates into the diagonal elements of the asymptotic results or expressions.

Unless otherwise specified, a nominal 5% level of significance is considered in all cases so that the expected percentage of rejections whenever the null hypothesis is true is equal to 5%.

Or equivalently, a 95% confidence interval can be constructed such that:

$$P[\hat{\beta}_k - z \times s.e.(\hat{\beta}_k) < \beta_k < \hat{\beta}_k + z \times s.e.(\hat{\beta}_k)] = 0.95 \tag{5.22}$$

where $s.e.(\hat{\beta}_k) = \sqrt{\widehat{Var}(\hat{\beta}_k)}$.

which is equivalent to

$$P(-1.96 < t < 1.96) = 0.95 \tag{5.23}$$

where t is defined by (5.21) and the standard z value at a 5% significant level is approximately 1.96 for large N.

Thus, the percent of coefficients contained in the 95% confidence interval is obtained for the estimators. The results are then examined to see whether the different estimators provide the desired (expected) level or not.

Note that all or some of these output statistics can be used in the analysis of results depending on the necessity and purpose of the comparison. Furthermore, although most of the outputs and other details of the experiment are provided in this Chapter, additional explanations will be provided in the subsequent Chapters when the need arises.

Finally, given the data generation process in this Chapter, the computer program for this Monte Carlo study is entirely written using the econometric computer program called SHAZAM, see White (1993), and processed on a mainframe computer. Any problems relating to the program or estimation (Monte Carlo) process are discussed along with the results of the experiment in the next Chapters.

5.4 Summary and Conclusions

In this Chapter we presented the details of the design of the experiment which is an integral part of the Monte Carlo study. The quality of the results and their applicability to more general situations depends on the design of the experiment. One of the most important aspects discussed in the design of this experiment is the data generation process of the experiment. The data generation process involves the generation of the explanatory variables, the determination of the levels of censoring and sample sizes, as well as the determination of the true values of the parameters of the model. Further, the various distributions of the error term and their generation mechanism are provided and discussed in this Chapter.

The design of this Monte Carlo experiment is not generally different from the previous studies discussed in Chapter 4. However, there are few noticeable differences compared to those studies. These are:

- The specific form of the model can be considered relatively more general compared to most of the previous studies discussed in Chapter 4.

- It involves a wide range of degrees of censoring and sample sizes.

- The distributional assumptions of the error term are selected to represent both symmetric and skewed distributions.

- The data generation process is designed so that some undesirable properties that may result from the design of the experiment are kept minimal.

- It involves a large number of replications which is very important for the accuracy of the outcomes.

- Most importantly, this experiment involves, not only a large number of estimators, but also much wider comparisons between the estimators. That is, most of the estimators of the model are evaluated using several criteria.

Finally, it is important to note that like any other Monte Carlo study, this one is also subject to problems such as specificity. We can only try to minimize these problems.

Chapter 6

Discussion of Results

6.1 Introduction

In this Chapter we discuss the results obtained for the various estimators of the tobit model. These estimators are discussed earlier in Chapters 2 and 3. A total number of 11 estimators are included in this analysis. The results are obtained from a Monte Carlo experiment of 3000 replications based upon the details of the experimental design provided in Chapter 5.

For convenience, the estimators of the tobit model can be divided into two main categories: those estimators using only non-limit, N_1, observations; and those which use all N observations. This division is made purely for discussion purposes because of the large number of estimators involved in the study. Further, it is also important to note that all coeffients of the model including the constant term play an important role in deriving useful results that are associated with tobit model. For example, the effects of the dependent variable, y_i, as a result of one unit change in x_{ij}, $i = 1, 2, ..., N$,

$j = 1, 2$ (see Section 2.10, Chapter 2). However, in almost all cases the comparison of the estimators in this Chapter will be based only on the outcomes of the experiment that are related to β_1 and β_2; mainly because of the large number of estimators and hence the large amount of output statistics involved in the comparison. However, once a few estimators are selected on a step by step basis for further analysis, the comparison of results in later Chapters will include results on all coeffients including the constant term of the model.

Given this situation, the anlysis in this Chapter is organized as follows. Section 6.2 presents the results obtained for those estimators using only N_1 observations. Results for the estimators using all N observations are discussed in Section 6.3. Section 6.4 presents a further analysis and comparison of selected estimators. Finally, a short summary and conclusion is presented in Section 6.5.

6.2 Estimators using only N_1 observations

Below, we discuss the Monte Carlo results obtained for the estimators using only the positive observations on y_i. These estimators include:

(i) The ordinary least squares estimator using only positive (non-limit) observations on y_i (OLSP).

(ii) The Heckman's two-step estimator (H2S).

(iii) The weighted Heckman's two-step estimator (WH2S).

(iv) The three-step estimator (3SE).

(v) The weighted three-step estimator (W3SE).

(vi) The nonlinear least squares estimator using only positive observations on y_i (NLSP).

Summary statistics of the Monte Carlo results for these estimators are provided in Tables 6.1-6.7 and Tables A.1-A.5 of Appendix A. As a guide to interpreting the tables, consider Table 6.1 which presents estimated results of estimators using N_1 observations, given a sample size of 100, a 25% degree of censoring and for the three distributions, namely, normal, students'-t and chi-square distributions. The first row of Table 6.1 presents the list of distributions and the parameters (β_1 and β_2) to be estimated under each distribution. The corresponding true values of the parameters are listed in row 3 of Table 6.1 where β_1=β_2=1.000 in all cases. The numbers in brackets indicate the column number in the table. The list of estimators are given in column (1). In column (2) are the variables estimated for each estimator, i.e., the estimated mean (EM), the standard error (SE), the bias (BIAS) and the root mean square error (RMSE). Finally, the corresponding estimates are presented in columns (3)-(8). For example, given a sample size of 100, a 25% degree of censoring and normally distributed error terms, the EM of $\hat{\beta}_1$ using OLSP is equal to 0.760 (see Column(3) of Table 6.1). And this is obtained as follows:

$$EM(\hat{\beta}_1) = \bar{\hat{\beta}}_1 = \frac{1}{3000}\sum_{i=1}^{3000}\hat{\beta}_{1i} = 0.760 \tag{6.1}$$

Similarly, continuing downwards in the same column is its standard error (SE) which is calculated as

Table 6.1: Results for Estimators using only N_1 observations given N=100 and 25% Degree of Censoring for the three Distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.760	0.788	0.825	0.832	0.753	0.768
	SE	0.093	0.167	0.098	0.167	0.110	0.197
	BIAS	-0.240	-0.212	-0.175	-0.168	-0.247	-0.232
	RMSE	0.257	0.270	0.201	0.236	0.271	0.304
H2S	EM	0.989	0.991	1.021	1.016	1.013	0.995
	SE	0.190	0.234	0.171	0.221	0.172	0.232
	BIAS	-0.011	-0.009	0.021	0.016	0.013	-0.005
	RMSE	0.190	0.234	0.173	0.222	0.173	0.232
WH2S	EM	0.987	0.985	1.039	1.036	0.999	0.978
	SE	0.186	0.232	0.189	0.242	0.183	0.242
	BIAS	-0.013	-0.015	0.039	0.036	-0.001	-0.022
	RMSE	0.186	0.234	0.193	0.244	0.183	0.247
3SE	EM	0.989	0.985	1.006	1.004	0.928	0.978
	SE	0.113	0.186	0.117	0.180	0.118	0.202
	BIAS	-0.015	-0.011	0.006	0.004	-0.072	-0.078
	RMSE	0.114	0.186	0.117	0.180	0.138	0.216
W3SE	EM	0.987	0.986	0.994	0.994	0.900	0.886
	SE	0.112	0.187	0.131	0.199	0.129	0.222
	BIAS	-0.013	-0.014	-0.006	-0.006	-0.100	-0.114
	RMSE	0.113	0.187	0.131	0.199	0.163	0.249
NLSP	EM	1.014	1.011	1.078	1.059	1.145	1.108
	SE	0.215	0.291	0.162	0.189	0.267	0.323
	BIAS	0.014	0.011	0.078	0.059	0.145	0.108
	RMSE	0.215	0.291	0.180	0.198	0.304	0.344

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

$$SE(\hat{\beta}_1) = \sqrt{\sum_{i=1}^{3000} (\hat{\beta}_{1i} - \bar{\hat{\beta}}_1)^2 / 3000} = 0.093 \tag{6.2}$$

and its bias is given by

$$\begin{aligned} BIAS(\hat{\beta}_1) &= \bar{\hat{\beta}}_1 - \beta_1 & (6.3) \\ &= 0.760 - 1.000 = -0.240 & (6.4) \end{aligned}$$

Finally, its RMSE is calculated by

$$RMSE(\hat{\beta}_1) = \sqrt{\sum_{i=1}^{3000} (\hat{\beta}_{1i} - \beta_1)^2 / 3000} = 0.257 \tag{6.5}$$

Other tables may be interpreted in a similar way. As discussed in Chapter 2 of this study, it is a well known fact that the ordinary least squares estimator using only positive values of y_i (OLSP) provides estimates which are biased. However, what may be of interest in this case is the degree and the direction of bias which may result due to changes in sample size, degree of censoring and distributional assumptions of the error term, and the relative performance compared with other estimators.

Table 6.1 shows that, given a sample size of 100 and a 25 percent degree of censoring, and using the OLSP estimator, β_1 and β_2 are estimated, respectively, with biases of 24.0% and 21.2% under normally distributed error terms, 17.5% and 16.8% under the t-distribution and 24.7% and 23.2% under the chi-square distribution. These biases are large compared, say, to the H2S estimates in which the coefficients are estimated with at most 2% bias under similar conditions. Further, as shown in Tables 6.2 and 6.3, the bias of the OLSP estimates remains high even for larger sample sizes.

Table 6.2: Results for Estimators using only N_1 observations given N=200 and 25% Degree of Censoring for the three Distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.746	0.782	0.818	0.830	0.725	0.739
	SE	0.064	0.132	0.071	0.126	0.080	0.138
	BIAS	-0.254	-0.218	-0.182	-0.170	-0.275	-0.261
	RMSE	0.262	0.255	0.195	0.212	0.286	0.296
H2S	EM	0.986	0.985	1.037	1.034	1.028	1.023
	SE	0.150	0.185	0.138	0.169	0.134	0.178
	BIAS	-0.014	-0.015	0.037	0.034	0.028	0.023
	RMSE	0.151	0.185	0.143	0.172	0.138	0.181
WH2S	EM	0.985	0.982	1.066	1.063	1.007	1.003
	SE	0.145	0.182	0.162	0.194	0.144	0.193
	BIAS	-0.015	-0.018	0.066	0.063	0.007	0.003
	RMSE	0.146	0.183	0.175	0.204	0.144	0.193
3SE	EM	0.989	0.988	1.013	1.013	0.935	0.936
	SE	0.079	0.147	0.084	0.137	0.087	0.146
	BIAS	-0.011	-0.012	0.013	0.013	-0.065	-0.064
	RMSE	0.080	0.148	0.085	0.137	0.109	0.159
W3SE	EM	0.991	0.989	0.998	0.993	0.907	0.902
	SE	0.078	0.148	0.095	0.158	0.096	0.161
	BIAS	-0.009	-0.011	0.002	0.007	-0.093	-0.098
	RMSE	0.078	0.148	0.095	0.158	0.134	0.188
NLSP	EM	1.010	1.007	1.075	1.062	1.149	1.112
	SE	0.144	0.199	0.100	0.135	0.178	0.234
	BIAS	0.010	0.007	0.075	0.062	0.149	0.112
	RMSE	0.144	0.199	0.126	0.148	0.232	0.259

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

Table 6.3: Results for estimators using only N_1 observations, given N=400 and 25% degree of censoring for the three distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.754	0.782	0.825	0.822	0.731	0.746
	SE	0.045	0.089	0.049	0.090	0.056	0.100
	BIAS	-0.246	-0.218	-0.175	-0.178	-0.269	-0.254
	RMSE	0.250	0.236	0.181	0.200	0.275	0.273
H2S	EM	0.996	0.995	1.037	1.034	1.032	1.017
	SE	0.099	0.124	0.089	0.118	0.093	0.123
	BIAS	-0.004	-0.005	0.037	0.034	0.032	0.017
	RMSE	0.099	0.124	0.096	0.123	0.098	0.124
WH2S	EM	0.996	0.993	1.065	1.065	1.015	0.999
	SE	0.095	0.122	0.110	0.140	0.100	0.132
	BIAS	-0.004	-0.007	0.065	0.065	0.015	-0.001
	RMSE	0.095	0.123	0.128	0.155	0.101	0.132
3SE	EM	0.997	0.996	1.020	1.017	0.939	0.934
	SE	0.056	0.098	0.057	0.098	0.060	0.104
	BIAS	-0.003	-0.004	0.020	0.017	-0.061	-0.066
	RMSE	0.056	0.098	0.061	0.099	0.085	0.123
W3SE	EM	0.998	0.997	1.006	1.000	0.912	0.897
	SE	0.054	0.098	0.067	0.118	0.064	0.116
	BIAS	-0.002	-0.003	0.006	0.000	-0.088	-0.103
	RMSE	0.054	0.098	0.068	0.118	0.109	0.155
NLSP	EM	1.004	1.005	1.088	1.068	1.136	1.111
	SE	0.100	0.130	0.076	0.097	0.119	0.152
	BIAS	0.004	0.005	0.088	0.068	0.136	0.111
	RMSE	0.100	0.130	0.117	0.119	0.181	0.188

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

In Table 6.3 where N=400 and a 25% degree of censoring, the biases of the OLSP estimates lie within the range of 17 to 27 percent compared to less than 2 percent for H2S estimates. In general, our results indicate that the OLSP estimator has the largest bias and this bias does not seem to decline with increases in sample size (see also Tables A.1 and A.2, Appendix A).

Regarding the effects of distributional assumptions about the error term, the OLSP estimates under the chi-square (skewed) distribution appear to be inferior relative to the same estimates under the normal and students'-t (symmetric) distributions. In Table 6.1 the RMSE for β_1 and β_2, respectively, are given by 0.257 and 0.270 for the normal distribution compared to 0.271 and 0.304 for the chi-square distribution. The respective values under the t-distribution are given by 0.201 and 0.236. Results for sample sizes of 200 (Table 6.2) and 400 (Table 6.3) also show that the OLSP estimates are relatively better under symmetric (i.e., normal and students'-t) distributions. Using the RMSE criteria the OLSP estimates, however, generally perform poorly relative to all the estimators except the NLSP estimator which is no better than the OLSP in several cases (see Tables 6.4-6.5).

Further, as anticipated, the biasedness of the OLSP estimator depends largely on the proportion of limit observations relative to the total number of observations in the sample, i.e., the degree of censoring. The effects of the degree of censoring are given in Table 6.4 for a sample of 100 and normally distributed error terms. When the degree of censoring is doubled from 25 to 50 percent the bias of the OLSP estimates increases by about 70 to 80 percent. The OLSP estimates deteriorate further with increases in the number of limit observations to 75 percent and in general the bias of the estimates increases almost linearly with the degree of censoring. Similar conclusions apply for

Table 6.4: The Effects of the Degree of censoring for Estimators using only N_1 observations, given N=100 and normally distributed error terms.

		25%		50%		75%	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.760	0.788	0.574	0.642	0.362	0.390
	SE	0.093	0.167	0.116	0.228	0.199	0.307
	BIAS	-0.240	-0.212	-0.426	-0.358	-0.638	-0.610
	RMSE	0.257	0.270	0.441	0.425	0.668	0.683
H2S	EM	0.989	0.991	0.977	0.976	0.952	0.947
	SE	0.190	0.234	0.406	0.430	1.604	1.561
	BIAS	-0.011	-0.009	-0.023	-0.024	-0.048	-0.053
	RMSE	0.190	0.234	0.406	0.431	1.604	1.562
WH2S	EM	0.987	0.985	0.978	0.976	0.946	0.938
	SE	0.186	0.233	0.382	0.410	1.542	1.494
	BIAS	-0.013	-0.015	-0.022	-0.024	-0.054	-0.062
	RMSE	0.186	0.234	0.382	0.410	1.543	1.494
3SE	EM	0.985	0.989	0.981	0.986	0.999	0.994
	SE	0.113	0.186	0.154	0.269	0.285	0.402
	BIAS	-0.015	-0.011	-0.019	-0.014	-0.001	-0.006
	RMSE	0.114	0.186	0.155	0.269	0.285	0.402
W3SE	EM	0.987	0.986	0.985	0.990	1.018	1.015
	SE	0.112	0.187	0.143	0.261	0.260	0.382
	BIAS	-0.013	-0.014	-0.015	-0.010	0.018	0.015
	RMSE	0.113	0.187	0.144	0.261	0.261	0.382
NLSP	EM	1.014	1.011	1.120	1.104	0.897	1.105
	SE	0.215	0.291	0.447	0.573	1.862	2.353
	BIAS	0.014	0.011	0.120	0.104	-0.103	0.105
	RMSE	0.215	0.291	0.463	0.583	1.864	2.355

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

Table 6.5: Results for Estimators using only N_1 observations given N=100 and 50% Degree of Censoring for the three Distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.574	0.642	0.629	0.656	0.392	0.480
	SE	0.116	0.228	0.176	0.237	0.162	0.265
	BIAS	-0.426	-0.358	-0.371	-0.344	-0.608	-0.516
	RMSE	0.441	0.425	0.410	0.418	0.629	0.579
H2S	EM	0.977	0.976	1.119	1.104	1.044	1.054
	SE	0.406	0.430	0.386	0.422	0.487	0.501
	BIAS	-0.023	-0.024	0.119	0.104	0.044	0.054
	RMSE	0.406	0.431	0.404	0.435	0.489	0.504
WH2S	EM	0.978	0.976	1.167	1.151	0.942	0.964
	SE	0.382	0.410	0.471	0.517	0.521	0.525
	BIAS	-0.022	-0.024	0.167	0.151	-0.058	-0.036
	RMSE	0.382	0.410	0.499	0.539	0.524	0.526
3SE	EM	0.981	0.986	0.959	0.958	0.827	0.867
	SE	0.154	0.269	0.188	0.256	0.193	0.297
	BIAS	-0.019	-0.014	-0.041	-0.042	-0.173	-0.133
	RMSE	0.155	0.269	0.193	0.259	0.259	0.326
W3SE	EM	0.985	0.990	0.907	0.905	0.829	0.870
	SE	0.143	0.261	0.258	0.330	0.196	0.327
	BIAS	-0.015	-0.010	-0.093	-0.095	-0.171	-0.130
	RMSE	0.144	0.261	0.274	0.344	0.260	0.352
NLSP	EM	1.120	1.104	1.295	1.314	1.358	1.393
	SE	0.447	0.573	0.458	0.644	0.600	0.944
	BIAS	0.120	0.104	0.295	0.314	0.358	0.393
	RMSE	0.463	0.583	0.545	0.716	0.699	1.022

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

medium and large samples (see Tables A.3, A.4 and A.5, Appendix A). These results, however, may not be surprising since the OLSP estimator is a simple linear regression based only on the non-limit observations in the sample. Our findings are also similar to those of Paarsch (1984).

Note that our results show that the OLSP provides estimates which are biased to wards zero in all cases, i.e., for all sample sizes, degrees of censoring and distributional assumptions. In general, one cannot determine the direction of bias of the OLSP estimator without making further assumptions. Goldberger (1981) has shown that, if the explanatory variables are normally distributed, OLSP estimators are biased towards zero.

The H2S estimator, compared to the OLSP, provides consistent estimates. As shown in Table 6.1, given a small sample size and a low degree of censoring, the biases of the estimates of β_1 and β_2 using the H2S estimator are, respectively, 1.1% and 0.9% under the normal distribution, 2.1% and 1.6% under the t-distribution and 1.3% and 0.5% for the chi-square distribution. These results are in contrast to the biases of the OLSP estimator which are in the range of about 16 to 25 percent. Bias appears to be relatively small for the H2S estimates for all sample sizes and distributions provided that the number of limit observations remains low (i.e., 25% degree of censoring). However, as the degree of censoring gets higher, bias becomes a problem for the non-normal distributions. This is shown clearly in Tables 6.5 and 6.6. For example, Table 6.6 depicts that given a sample size of 100 and 75% degree of censoring, the biases of the H2S estimates are about 20 and 60 percent under chi-square and t-distributions, respectively, compared to about 5% under the normal distribution. Further, if the sample size is increased to 400, the bias for the normal

Table 6.6: Results for estimators using only N_1 observations, given N=100 and 75% degree of censoring for the three distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.362	0.390	0.160	0.279	0.170	0.240
	SE	0.199	0.307	0.433	0.502	0.235	0.398
	BIAS	-0.638	-0.610	-0.840	-0.721	-0.830	-0.760
	RMSE	0.668	0.683	0.945	0.879	0.863	0.858
H2S	EM	0.952	0.947	1.609	1.535	0.791	0.848
	SE	1.604	1.561	2.668	2.398	1.378	1.375
	BIAS	-0.048	-0.053	0.609	0.535	-0.209	-0.152
	RMSE	1.604	1.562	2.737	2.456	1.394	1.383
WH2S	EM	0.946	0.938	1.660	1.589	0.709	0.750
	SE	1.542	1.494	3.217	2.891	1.458	1.457
	BIAS	-0.054	-0.062	0.660	0.589	-0.291	-0.250
	RMSE	1.543	1.495	3.284	2.950	1.487	1.478
3SE	EM	0.999	0.994	0.784	0.835	0.757	0.808
	SE	0.285	0.402	0.476	0.554	0.307	0.477
	BIAS	-0.001	-0.006	-0.215	-0.165	-0.243	-0.192
	RMSE	0.285	0.402	0.522	0.578	0.391	0.514
W3SE	EM	1.018	1.015	0.790	0.823	0.821	0.858
	SE	0.260	0.382	0.514	0.689	0.301	0.499
	BIAS	0.018	0.015	-0.210	-0.177	-0.179	-0.142
	RMSE	0.261	0.382	0.555	0.711	0.350	0.519
NLSP	EM	0.897	1.105	1.219	1.528	1.249	1.676
	SE	1.862	2.353	2.055	2.550	1.322	2.065
	BIAS	-0.103	0.105	0.219	0.528	0.249	0.676
	RMSE	1.864	2.355	2.066	2.608	1.345	4.723

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

distribution drops to about 1% whereas the biases for the non-normal distributions remain almost the same (see Table A.5, Appendix A). Based on the RMSE criteria, the H2S appears to be robust for normal and non-normal distributions given that the degree of censoring remains low. However, when the degree of censoring increases, the H2S estimator performs better under the normal distribution.

Further, the efficiency of the H2S estimates declines dramatically with higher levels of the degree of censoring even when the sample size becomes large. This is evident from the results summarized in Table 6.7. That is, given a large sample size of 400 and normally distributed error terms, the estimates of the SE's of β_1 and β_2 increased by almost 50% when the proportion of limit observations in the sample increased from 25 to 50 percent, and are six to seven times higher when the degree of censoring is 75%. In general, the effects of the degree of censoring on the performance of the H2S estimator are very severe relative to the effects of the distributional assumptions about the error term. The H2S estimates deteriorate further under non-normal distributions coupled with higher levels of censoring (see also Tables A.3-A.5, Appendix A).

It is important to note that the weighted Heckman's two-step estimator (WH2S) provides more efficient estimates of β_1 and β_2 only when the errors are normally distributed. As can be seen from Tables 6.1-6.4, the WH2S estimates are relatively more efficient than their H2S counterparts under normally distributed error terms. Otherwise, the WH2S estimator gives results which are relatively inferior to the H2S estimates. Similar results can be observed for all sample sizes and degrees of censoring. In general, the WH2S estimator is sensitive to changes in distributional assumptions about the error structure.

Table 6.7: The effects of the degree of censoring for estimators using only N_1 observations, given N=400 and normally distributed error terms.

		25%		50%		75%	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLSP	EM*	0.754	0.782	0.561	0.622	0.361	0.398
	SE	0.045	0.089	0.058	0.107	0.091	0.127
	BIAS	-0.246	-0.218	-0.439	-0.378	-0.638	-0.602
	RMSE	0.250	0.236	0.443	0.393	0.646	0.616
H2S	EM	0.996	0.995	0.993	0.992	0.993	0.989
	SE	0.099	0.124	0.207	0.213	0.669	0.635
	BIAS	-0.004	-0.005	-0.007	-0.008	-0.007	-0.011
	RMSE	0.099	0.124	0.207	0.212	0.669	0.635
WH2S	EM	0.996	0.993	0.990	0.989	0.997	0.993
	SE	0.095	0.122	0.191	0.200	0.599	0.574
	BIAS	-0.004	-0.007	-0.010	-0.011	-0.003	-0.007
	RMSE	0.095	0.123	0.192	0.201	0.599	0.574
3SE	EM	0.997	0.996	0.995	0.995	1.001	0.997
	SE	0.056	0.098	0.078	0.131	0.139	0.179
	BIAS	-0.003	-0.004	-0.005	-0.005	0.001	-0.003
	RMSE	0.056	0.098	0.079	0.131	0.139	0.179
W3SE	EM	0.998	0.997	0.997	0.997	1.004	1.001
	SE	0.054	0.098	0.073	0.128	0.126	0.170
	BIAS	-0.002	-0.003	-0.003	-0.003	0.004	0.001
	RMSE	0.054	0.098	0.073	0.128	0.126	0.170
NLSP	EM	1.004	1.005	-	-	-	-
	SE	0.100	0.130	-	-	-	-
	BIAS	0.004	0.005	-	-	-	-
	RMSE	0.100	0.130	-	-	-	-

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

On the other hand, the 3SE estimator provides results which are superior to the H2S estimator or its weighted version, the WH2S, in all cases. Given a small sample size (i.e., 100), the RMSE for the 3SE estimator is always lower for the three degrees of censoring as shown in Tables 6.1, 6.5 and 6.6. Similarly, the results for the medium and large sample sizes depict that the 3SE estimator gives more efficient estimates under all degrees of censoring. However, what is more important about these results is that the 3SE estimator is not only more efficient but also less sensitive to changes in the degree of censoring in a given sample. For example, Table 6.4 depicts that when the degree of censoring increases from 25 to 75 percent the RMSE for the 3SE of β_1 and β_2 increased by just over two times, i.e., from 0.114 and 0.186 to 0.285 and 0.402, respectively. This result is incomparable to the H2S estimates where the increase in RMSE of the coefficients is six to eight times for the same changes in degrees of censoring. This is also true for the medium and large samples which suggests that the 3SE estimator is much less sensitive to changes in the degree of censoring for a given sample (see Table 6.7 and Tables A.3-A.5 of Appendix A).

As to the effects of distributional assumptions about the error term, the 3SE estimator seems to perform better under normality conditions and gets worse under the chi-square distribution. This is particularly significant for higher degrees of censoring. However, it is important to note that under all conditions the 3SE estimator actually improves on the H2S estimates in terms of reliability (efficiency). Furthermore, the 3SE estimator is much less sensitive to increases in the proportion of limit observations in a sample compared to the H2S estimator. Note that, as discussed in Chapter 5, higher level of censoring means an increase in the correlation between the explanatory variables and the inverse of Mill's ratio, and hence adversely affecting the H2S estimates. A more detailed analysis along this line will be provided later in this study.

In general, the 3SE estimates have the lowest RMSE, except when the errors are normally distributed. In such cases it's weighted version, the W3SE, yields more efficient estimates. Similar to the WH2S estimator, the W3SE estimator is sensitive to changes in distributional assumptions about the error structure. One likely explanation for the sensitiveness of the weighted estimators to non-normality of the error terms is that the expressions for the weights depend on the assumption of normality of the error terms. Specifically, as shown in Chapter 2, the weights involve results such as the probability and cumulative density functions of the standard normal distribution. Hence, any departure from normality of the error terms may result in inefficiency and/or inconsistency of estimates.

Note that, as discussed in Chapter 2, the 3SE estimator guarantees that the estimated value of σ, which is given in the right hand side of equation (2.42), is positive, which is not the case for the H2S estimator. For example, consider Table 6.8 below which depicts that, given a sample size of 100, a 25% degree of censoring and normally distributed error terms, the estimated values of σ, using the H2S estimator, ranges from -1.266 to 4.237 compared to 0.177 to 1.635 for the 3SE estimator. Similar results are observed for all sample sizes, distributions and degrees of censoring. That is, our results depict that, unlike the H2S estimator, the 3SE estimator always guarantees positive estimates of σ. Further, the standard errors of the 3SE estimates are much smaller than those of the H2S estimates implying gains in relative efficiency.

Table 6.8: Comparisons of H2S and 3SE estimates, given a Sample Size of 100, 25% degree of Censoring and the three distributions.

		Normal			Students'-t			Chi-Square		
		β_1	β_2	σ	β_1	β_2	σ	β_1	β_2	σ
True Values		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
H2S	EM	0.989	0.991	0.974	1.021	1.016	0.898	1.013	0.995	1.082
	SE	0.190	0.234	0.672	0.171	0.221	0.659	0.172	0.232	0.537
	MIN*	0.336	0.265	-1.226	0.485	-0.184	-1.138	0.464	0.197	-0.437
	MAX*	1.849	1.869	4.237	1.856	3.006	5.311	1.762	2.192	3.485
3SE	EM	0.985	0.989	0.947	1.006	1.004	0.820	0.928	0.922	0.722
	SE	0.113	0.186	0.179	0.117	0.180	0.207	0.118	0.202	0.134
	MIN	0.635	0.323	0.177	0.309	0.247	0.249	0.482	0.230	0.162
	MAX	1.443	1.609	1.635	1.624	2.494	1.525	1.415	1.623	1.195

—* MIN and MAX stands for the minimum and maximum estimates, respectively, in 3000 replications (samples).

Considering the NLSP estimator, the nonlinear equation which is discussed in Chapter 2 consists of non trivial functions involving probability and cumulative density functions. The complex nature of the function has affected the speed and convergence of the nonlinear estimates in our experiment. In SHAZAM the nonlinear regressions are estimated by maximum likelihood, assuming that the errors are additive and normally distributed. The estimation procedure uses the algorithm known as a Quasi-Newton method. However, although not difficult, the nonlinear estimation in SHAZAM is rather slow and depends on the complexities of the nonlinear function. In this case the NLSP estimator takes significantly more computer time to converge than do other estimators. For example, it takes approximately 4 minutes CPU time to obtain results from 100 samples (replications) using NLSP compared to about 20 seconds CPU time for H2S or 3SE estimators.

Note that convergence is not always guaranteed in nonlinear least squares, even after a large number of iterations. In this experiment we considered a maximum of 100 iterations, and to help speed up the convergence, the true parameter values were used as starting values for the NLSP estimates. However, some did not converge after 100 iterations. For example, given a sample size of 200 observations and a 75% degree of censoring, of the 3000 replications (samples), 4, 2, and 5 percent of the samples did not converge, respectively, for normal, students'-t and chi-square distributions. Convergence is more difficult for small samples and higher degrees of censoring. Results for samples which did not converge are excluded from the experiment.

Given this, our results in general show that the NLSP estimator provides the least efficient estimates. As shown in Table 6.1, given normally distributed error terms and a 25% degree of censoring, the SE's of the estimates of β_1 and β_2 are 0.215 and 0.291, respectively, for a sample size of 100. These results are over 10 percent larger than the corresponding H2S estimates, and a further 50 to 90 percent larger than the 3SE estimates. By increasing the sample size to 200 the NLSP estimates come relatively close to the H2S estimates, but are still much less efficient than the 3SE estimates (see Table 6.2). This is also true for the large sample size (see Table 6.3) even if the degree of censoring remains low. Wales and Woodland (1980) made similar observations based on a single simulation result.

Further, as the proportion of limit observations in the sample increases, the quality of the NLSP estimates deteriorates quickly. Table 6.4 depicts that, given a small sample size and normally distributed error terms, the RMSE of the NLSP estimates almost doubled when the degree of censoring increased from 25% to 50 % and further increased to over eight times for a 75% degree of censoring.

As to the effects of distributional assumptions of the error term, the NLSP estimator performs relatively better under normal conditions, especially for degrees of censoring of 50% and above, where the NLSP provides relatively better results under normality conditions and gets worse under the chi-square (skewed) distribution (see Tables 6.5 and 6.6). In other words, the NLSP estimator is not robust to changes in the error structure of the model.

Finally, the following points can be concluded from the preceding discussions [see also Table 6.9 for some concluding remarks]:

> The 3SE estimator provides estimates which are relatively efficient and less sensitive to changes in the degree of censoring. Bias is not a problem for lower levels of censoring. However, bias becomes a problem for higher degrees of censoring and non-normal error terms. It also seems to perform slightly better under normality conditions. The 3SE estimator appears to be the best, given the estimators which use only the positive observations on y_i.
>
> The H2S estimator is generally less efficient compared to the 3SE estimator and can be very sensitive to changes in the degree of censoring. The H2S estimator seems to be robust to changes in distributional assumptions, provided that the degree of censoring is low. However, bias becomes a serious problem for higher degrees of censoring and non-normal distributions.
>
> It is important to emphasize that the weighted versions of the Heckman's two-step (WH2S) and the three-step (W3SE) estimators may improve the efficiency of their counterparts; i.e., the H2S and 3SE estimators, respectively, if and

Table 6.9: Summary Notes on the Relative Performance of the Various Estimators.

	Degree of	Normal			Students'-t			Chi-square		
Estimator	Censoring	100	200	400	100	200	400	100	200	400
3SE	25%	√	√	√	√	√	√	√	√	√
	50%	√	√	√	√	√	√	♠	♠	♠
	75%	√	√	√	√	√	√	♠	♠	♠
W3SE	25%	√	√	√	†	†	†	†	†	†
	50%	√	√	√	♠	♠	♠	♠	♠	♠
	75%	√	√	√	♠	♠	♠	♠	♠	♠
H2S	25%	√	√	√	√	√	√	√	√	√
	50%	†	†	√	♠	♠	♠	†	†	†
	75%	†	†	†	♠	♠	♠	♠	♠	♠
WH2S	25%	√	√	√	†	†	†	†	†	†
	50%	†	†	√	♠	♠	♠	†	†	†
	75%	†	†	†	♠	♠	♠	♠	♠	♠
NLSP	25%	†	√	√	†	√	√	†	†	†
	50%	♠	†	†	♠	†	†	♠	†	†
	75%	♠	♠	♠	♠	♠	♠	♠	♠	♠
OLSP	Bias proportional to Degree of Censoring									

Note:
√ = performs well.
† = inefficiency appears to be a problem.
♠ = bias and inefficiency appear to be a problem.

only if the errors are normally distributed. Otherwise, the WH2S and W3SE estimators provide estimates which are inferior to the H2S and 3SE estimates, respectively. Thus one should take proper caution before applying the weighted versions of the estimators to obtain more efficient estimates, as the assumption of normality may not actually hold. In other words, preliminary steps such as pre-testing for normality of the error structure may be necessary. On the other hand, it should be noted that, even under normality conditions of the error term, the improvement of the weighted estimates over the unweighted estimates seems to be marginal.

The NLSP estimator is not one of the best estimators for the following reasons. It provides inefficient estimates relative to H2S and 3SE estimates. Further, the NLSP is not robust to changes in distributional assumptions about the error term. Bias seems to be a problem and gets worse for skewed distributions and higher levels of censoring. It is computationally more difficult and may not always converge. Specifically, convergence is difficult for small sample sizes and higher degrees of censoring.

Not surprisingly, the OLSP estimator provides poor estimates in all cases.

6.3 Estimators using all N observations

In this Section we discuss the results for estimators which use all limit and non-limit observations on y_i. These estimators are:

(i) The simple ordinary least squares estimator (OLS) using all observations.

(ii) The maximum likelihood estimator (MLE).

(iii) The Heckman's two-step estimator based on the unconditional expectation of the tobit model (H2SU).

(iv) The weighted Heckman's two-step estimator based on the unconditional expectation of the model (WH2SU). That is, the weighted version of (iii).

(v) The nonlinear least squares estimator applied to the unconditional expectation of the model (NLSU).

Given these estimators, summary statistics of the Monte Carlo results are provided in Tables 6.10-6.14 and Table A.6 of Appendix A. The interpretation of the results is similar to those in the previous tables, except that these estimators utilize all the observations on y_i.

As discussed in Chapter 2, the traditional OLS estimator provides biased estimates and hence is not feasible. The interest in this case is to check its relative performance compared to that of the other estimators. Results for a sample of 100 and a 25% degree of censoring are tabulated in Table 6.10. These results suggest that bias is a serious problem for OLS estimates and averages around 20 percent for all distributions compared to the MLE estimates which range from a minimum of 0.2 percent under normally distributed errors to a maximum of 4.3 percent for the chi-square distribution. The bias of the OLS estimates remains high and does not seem to decline even when the sample size is large (see Tables 6.11 to 6.12). Further, the severity of bias of the OLS estimates increases proportionately with the number of limit observations in the sample, irrespective of the total sample size or distributional assumptions of the error structure of the model. For example, as can be seen from

Table 6.10: Results for Estimators using all observations given N=100 and 25% Degree of Censoring for the three Distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLS	EM*	0.800	0.797	0.808	0.795	0.808	0.791
	SE	0.069	0.170	0.069	0.153	0.073	0.141
	BIAS	-0.200	-0.203	-0.192	-0.205	-0.192	-0.209
	RMSE	0.211	0.263	0.204	0.255	0.205	0.252
MLE	EM	1.002	1.002	1.023	1.025	1.043	1.028
	SE	0.099	0.177	0.099	0.174	0.109	0.196
	BIAS	0.002	0.002	0.023	0.025	0.043	0.028
	RMSE	0.099	0.177	0.102	0.176	0.117	0.198
H2SU	EM	0.811	0.876	0.797	0.849	0.779	0.877
	SE	0.172	0.244	0.117	0.232	0.116	0.196
	BIAS	-0.189	-0.124	-0.203	-0.151	-0.221	-0.213
	RMSE	0.256	0.273	0.234	0.276	0.250	0.232
WH2SU	EM	0.644	0.790	0.663	0.719	0.573	0.767
	SE	0.223	0.281	0.214	0.377	0.186	0.217
	BIAS	-0.356	-0.210	-0.337	-0.281	-0.427	-0.233
	RMSE	0.420	0.251	0.425	0.470	0.466	0.318
NLSU	EM	1.065	1.047	1.058	1.047	1.090	1.055
	SE	0.288	0.280	0.232	0.277	0.238	0.288
	BIAS	0.065	0.047	0.058	0.047	0.090	0.055
	RMSE	0.237	0.284	0.238	0.280	0.255	0.293

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

Table 6.11: Results for Estimators using all observations, given N=400 and Degree of Censoring of 25% for the three Distributions.

		Normal		Students'-t		Chi-Square	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLS	EM*	0.800	0.767	0.833	0.815	0.786	0.765
	SE	0.034	0.071	0.033	0.073	0.036	0.078
	BIAS	-0.200	-0.233	-0.167	-0.185	-0.214	-0.235
	RMSE	0.203	0.244	0.170	0.198	0.217	0.247
MLE	EM	1.000	1.001	1.023	1.019	1.044	1.025
	SE	0.048	0.093	0.051	0.091	0.056	0.103
	BIAS	0.000	0.001	0.023	0.019	0.044	0.025
	RMSE	0.048	0.093	0.056	0.093	0.071	0.106
H2SU	EM	0.808	0.857	0.823	0.857	0.744	0.855
	SE	0.057	0.104	0.067	0.103	0.045	0.108
	BIAS	-0.192	-0.143	-0.177	-0.143	-0.256	-0.145
	RMSE	0.210	0.177	0.189	0.176	0.259	0.181
WH2SU	EM	0.624	0.735	0.591	0.662	0.518	0.731
	SE	0.070	0.124	0.191	0.227	0.117	0.129
	BIAS	-0.376	-0.265	-0.409	-0.338	-0.482	-0.269
	RMSE	0.383	0.293	0.452	0.408	0.496	0.299
NLSU	EM	1.033	1.025	1.016	1.007	1.057	1.047
	SE	0.146	0.153	0.120	0.135	0.123	0.150
	BIAS	0.033	0.025	0.016	0.007	0.057	0.047
	RMSE	0.149	0.155	0.121	0.135	0.136	0.157

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

Table 6.12: The Effects of Sample Size on the Estimators using all observations, given 25% Degree of Censoring and Normally Distributed error terms.

		100		200		400	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLS	EM*	0.800	0.797	0.803	0.789	0.800	0.767
	SE	0.068	0.170	0.050	0.102	0.034	0.071
	BIAS	-0.200	-0.203	-0.197	-0.211	-0.200	-0.233
	RMSE	0.211	0.263	0.204	0.235	0.203	0.244
MLE	EM	1.002	1.002	1.002	1.002	1.000	1.001
	SE	0.099	0.177	0.070	0.129	0.048	0.093
	BIAS	0.002	0.002	0.002	0.002	0.000	0.001
	RMSE	0.099	0.177	0.070	0.129	0.048	0.093
H2SU	EM	0.811	0.876	0.817	0.870	0.808	0.857
	SE	0.172	0.244	0.096	0.151	0.057	0.104
	BIAS	-0.189	-0.124	-0.183	-0.130	-0.192	-0.143
	RMSE	0.256	0.273	0.206	0.199	0.201	0.177
WH2SU	EM	0.644	0.790	0.639	0.762	0.624	0.735
	SE	0.233	0.281	0.117	0.174	0.070	0.124
	BIAS	-0.356	-0.210	-0.361	-0.238	-0.376	-0.265
	RMSE	0.420	0.251	0.379	0.294	0.383	0.293
NLSU	EM	1.065	1.047	1.045	1.029	1.033	1.025
	SE	0.228	0.280	0.193	0.207	0.146	0.153
	BIAS	0.065	0.047	0.045	0.029	0.033	0.025
	RMSE	0.237	0.284	0.198	0.209	0.149	0.155

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

as can be seen from Table 6.12, given a sample size of 200 and normally distributed error terms, the estimated biases of β_1 and β_2 are given by, respectively, -0.197 and -0.211 for 25% degree of censoring, and -0.505 and -0.550 for 50% degree of censoring. The biases of the OLS estimates of β_1 and β_2 increased further to -0.797 and -0.757, respectively, for 75% degree of censoring. Similar results are observed for low and large sample sizes.

It is not, however, surprising for the OLS estimator to perform badly, as mentioned above. What is surprising, from our results, is the poor performance of the estimators suggested by Wales and Woodland (1980), the H2SU and the WH2SU estimators. Table 6.10 depicts that the biases of these estimates ranges from 12 to 22 percent for H2SU estimates and deteriorates further to 21 to 47 percent for WH2SU estimates. More surprisingly, the bias of the estimates does not seem to decline with the increase in sample size or changes in distribution of the error structure of the model (see Tables 6.11 and 6.12). Also, as shown in Table 6.13, the estimates get worse for higher levels of censoring. To sum up, despite their large sample properties (i.e., consistency and asymptotic normality), the H2SU and the WH2SU estimators provide very poor results in all cases.

The nonlinear least squares estimator using all observations (NLSU) seems to perform well compared to the others, with the exception of the MLE estimator. As shown in Table 6.10, given a sample size of 100 and a 25% degree of censoring, the biases of the NLSU estimates range between 4 to 9 percent compared to about 20% or more for H2SU or WH2SU estimators over the three distributions. The biases of the NLSU estimates decline further with increases in sample size (see Tables 6.11-6.12). However, bias becomes a problem for higher degrees of censoring (see Tables 6.13 and

Table 6.13: The Effects of Degree of Censoring for Estimators using all observations, given N=200 and Normally distributed error terms.

		25%		50%		75%	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLS	EM*	0.803	0.789	0.495	0.450	0.203	0.243
	SE	0.050	0.102	0.041	0.075	0.030	0.053
	BIAS	-0.197	-0.211	-0.505	-0.550	-0.797	-0.757
	RMSE	0.204	0.235	0.507	0.555	0.797	0.758
MLE	EM	1.002	1.002	1.003	1.002	1.008	0.990
	SE	0.070	0.129	0.094	0.159	0.155	0.219
	BIAS	0.002	0.002	0.003	0.002	0.008	-0.010
	RMSE	0.070	0.129	0.094	0.159	0.156	0.219
H2SU	EM	0.817	0.870	0.473	0.505	0.311	0.377
	SE	0.096	0.151	0.037	0.166	0.079	0.264
	BIAS	-0.183	-0.130	-0.527	-0.595	-0.689	-0.623
	RMSE	0.206	0.199	0.528	0.522	0.693	0.677
WH2SU	EM	0.693	0.762	0.329	0.376	0.105	0.370
	SE	0.117	0.174	0.044	0.168	0.088	0.283
	BIAS	-0.361	-0.238	-0.671	-0.624	-0.895	-0.627
	RMSE	0.379	0.294	0.623	0.646	0.899	0.688
NLSU	EM	1.045	1.029	1.240	1.249	1.136	1.124
	SE	0.193	0.207	0.266	0.325	0.381	0.445
	BIAS	0.045	0.029	0.240	0.249	0.136	0.124
	RMSE	0.198	0.209	0.358	0.410	0.404	0.462

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

6.14). Furthermore, the NLSU estimates are very inefficient compared to the MLE estimates. For example, Table 6.10 depicts that the standard errors of the NLSU estimates of β_1 and β_2, given normally distributed error terms, are about 290 and 158 percent higher than their respective MLE estimates. The results are similar for medium and large sample sizes as well as for the non-normal distributions. Note that the nonlinear estimation is very slow compared to MLE estimation. For example, given a sample size of 400, normally distributed errors and a 25% degree of censoring, it takes about 5 minutes CPU time for NLSU, compared to 58 seconds for MLE estimates, to obtain results from 100 replications.

In general, our results indicate that the MLE estimator provides relatively better estimates under all circumstances, given the estimators which use all observations. Note that under normality conditions, it is widely claimed that the MLE provides consistent and more efficient estimates. In Table 6.10, the bias for the MLE of β_1 and β_2 is about 0.2 percent under normally distributed error terms. This is compared to about 2.5 and 4 percent bias under students'-t and chi-square distributions, respectively. The bias of the MLE estimates under the normal distribution disappears when sample size increases, but remains almost at the same level for the non-normal distributions (see Tables 6.11-6.12). Further, as shown in Table A.6 of Appendix A, the bias of the MLE estimator can be substantial under the non-normal distributions if the sample size is small and the degree of censoring high. These results may suggest that, although the bias is relatively small, the MLE is not robust to the distributional assumption of the error terms.

With regard to the efficiency of the MLE estimates, the MLE estimates appear to be most efficient under normal and students'-t distributions compared to the skewed

Table 6.14: The Effects of Degree of Censoring for Estimators using all observations, given N=200 and Chi-Square Distributed error terms.

		25%		50%		75%	
		β_1	β_2	β_1	β_2	β_1	β_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
True values		1.000	1.000	1.000	1.000	1.000	1.000
OLS	EM*	0.801	0.764	0.473	0.467	0.201	0.260
	SE	0.051	0.110	0.045	0.090	0.035	0.072
	BIAS	-0.199	-0.236	-0.257	-0.533	-0.799	-0.740
	RMSE	0.205	0.261	0.529	0.541	0.800	0.743
MLE	EM	1.049	1.033	1.048	1.033	1.032	1.092
	SE	0.080	0.146	0.119	0.193	0.200	0.283
	BIAS	0.049	0.033	0.048	0.033	0.032	0.092
	RMSE	0.094	0.149	0.128	0.196	0.202	0.298
H2SU	EM	0.775	0.828	0.442	0.560	0.325	0.337
	SE	0.077	0.150	0.041	0.192	0.099	0.318
	BIAS	-0.225	-0.172	-0.558	-0.440	-0.675	-0.663
	RMSE	0.237	0.228	0.559	0.480	0.682	0.735
WH2SU	EM	0.581	0.670	0.300	0.508	-	-
	SE	0.158	0.181	0.066	0.202	-	-
	BIAS	-0.419	-0.330	-0.700	-0.492	-	-
	RMSE	0.447	0.376	0.703	0.532	-	-
NLSU	EM	1.070	1.044	1.193	1.236	-	-
	SE	0.175	0.215	0.345	0.445	-	-
	BIAS	0.070	0.044	0.193	0.236	-	-
	RMSE	0.189	0.219	0.396	0.504	-	-

* EM=Estimated Mean, SE=Standard Error, RMSE=Root Mean Square Error.

(chi-square) distribution. Table 6.10 depicts that, given small sample size and low degrees of censoring, the standard errors of the MLE estimates when the errors have chi-square distribution are about 10 percent higher than the standard errors of the estimates for normal and students'-t distributions. Similar observations can be made for medium and large sample sizes. However, as shown in Table 6.14, the quality of the MLE estimates declines with increases in the level of censoring and non-normal distributed error terms (see also Table A.6, Appendix A). On the other hand, it is important to note that the MLE seems to be robust to symmetric but wide tailed distributions, as can be seen from the results for students'-t distributed error terms. That is, although there exists some bias under t-distributed error terms, the MLE estimates for normal and students'-t distribution are almost the same in most cases in terms of the RMSE; except for a sample size of 100 and 75% degree of censoring where the results for the t-distribution appears to be even inferior to those of the chi-square distributed error terms (see Table A.6, Appendix A). That is, using the RMSE criteria, the results for MLE estimates under students'-t distributed error terms are as good as those from normally distributed error terms. As to the effects of censoring, Tables 6.13-6.14 indicate that doubling the degree of censoring from 25% to 50% percent may cause a 25 to 30 percent increase in standard errors of estimates under normally distributed error terms (Table 6.13) and about 30 to 40 percent for chi-square distributed error terms (Table 6.14). A further increase in the degrees of censoring results in more inefficient estimates.

Finally, given the estimators that use all observations, one can make the following conclusions [see also Table 6.15 for some important notes]:

Table 6.15: Summary Notes on the Relative Performance of the Various Estimators.

Estimator	Degree of Censoring	Normal 100	Normal 200	Normal 400	Students'-t 100	Students'-t 200	Students'-t 400	Chi-square 100	Chi-square 200	Chi-square 400
MLE	25%	√	√	√	√	√	√	√	√	√
	50%	√	√	√	√	√	√	†	†	√
	75%	√	√	√	♠	√	√	♠	†	√
NLSU	25%	†	†	†	†	†	†	†	†	†
	50%	♠	♠	♠	♠	♠	♠	♠	♠	♠
	75%	♠	♠	♠	♠	♠	♠	♠	♠	♠
H2SU	25%	♠	♠	♠	♠	♠	♠	♠	♠	♠
	50%	♠	♠	♠	♠	♠	♠	♠	♠	♠
	75%	♠	♠	♠	♠	♠	♠	♠	♠	♠
WH2SU	25%	♠	♠	♠	♠	♠	♠	♠	♠	♠
	50%	♠	♠	♠	♠	♠	♠	♠	♠	♠
	75%	♠	♠	♠	♠	♠	♠	♠	♠	♠
OLS	Bias proportional to Degree of Censoring									

Note:
√ = performs well.
† = inefficiency appears to be a problem.
♠ = bias and inefficiency appear to be a problem.

Overall, the MLE performs better in all circumstances; i.e., for all sample sizes, distributions and degrees of censoring. In particular, the MLE performs much better under normal and students'-t distributions than it does under the chi-square distribution; except when the sample size is small (i.e., 100) and the degree of censoring high (i.e., 75%). A situation where the results under the t-distribution are no better than those under the chi-square distribution.

The H2SU and WH2SU estimators are no better than the simple OLS estimator and provide very poor results in all cases. Bias is a serious problem and gets worse with higher degrees of censoring.

The NLSU estimator is very inefficient compared to the MLE estimator in all cases. More over, bias becomes a problem as the degree of censoring increases.

6.4 Further Analysis of Selected Estimators

This Section presents a further comparison of estimators which are selected from Sections 6.2 and 6.3 above. Estimators which are either biased and/or relatively highly inefficient (or in general terms too poor to be candidates) are excluded from further discussion on the basis of the preceding discussions. Specifically, the ordinary least squares (OLSP) and the nonlinear least squares (NLSP) estimators are excluded from the estimators using only positive observations on y_i (i.e., from Section 6.2). Further, from the estimators using all observations (i.e., from Section 6.3), the simple ordinary least squares (OLS), the Heckman's two-step estimator based on the unconditional expectation of the model (H2SU) and its weighted version, the WH2SU estimator, are excluded due to their relative poor performances.

In total, six out of eleven estimators are discussed below. Given these estimators, a summary of relative root mean square errors (RMSE) of the estimators are provided in Tables 6.16 and 6.17 for 25% and 50% degrees of censoring, respectively, and for all sample sizes and distributions. These relative RMSE are obtained by dividing the RMSE of each estimator by the corresponding RMSE of the MLE estimator, for a given sample size, distribution and degree of censoring. For example, given a sample size of 100, normally distributed error terms and a 25% degree of censoring, the RMSE of β_1 for the H2S estimator is equal to 0.190 (see Table 6.1), and the corresponding RMSE for the MLE estimator is equal to 0.099 (see Table 6.10). Thus, the relative RMSE for β_1 using the H2S estimator is given by the ratio 0.190/0.099=1.919, which is shown at the top of Column 4 of Table 6.16. Others are calculated in a similar way.

As can be seen from Table 6.16, the NLSU estimator has the largest relative RMSE value for all sample sizes and distributions implying that the NLSU estimator is relatively poor (inefficient). For example, given a sample size of 100 and normally distributed error terms, the RMSE of the NLSU estimates of β_1 and β_2 are, respectively, 2.394 and 1.605 times greater than their respective MLE estimates. The relative RMSEs of the NLSU estimates of β_1 and β_2 further increase to 3.104 and 1.667, respectively, when the sample size becomes large (i.e., 400). This is because the relative efficiency of the MLE estimates increases at a higher speed with increases in sample size compared to that of the NLSU estimates. Further, the NLSU estimates deteriorate for higher levels of censoring as shown in Table 6.17. Note that the relative RMSE values of the NLSU estimates seem to decline for the non-normal

Table 6.16: Relative Root Mean Square Errors (RMSE) for all Sample Sizes and Distributions, given 25% Degree of Censoring.

Sample Size	Estimator	Para- -meter	Relative RMSE when the errors are Normal	Students'-t	Chi-Square
100	H2S	β_1	1.919	1.696	1.479
		β_2	1.322	1.261	1.172
	WH2S	β_1	1.879	1.892	1.564
		β_2	1.322	1.386	1.247
	3SE	β_1	1.151	1.147	1.179
		β_2	1.051	1.023	1.091
	W3SE	β_1	1.141	1.284	1.393
		β_2	1.056	1.131	1.257
	MLE	β_1	1.000	1.000	1.000
		β_2	1.000	1.000	1.000
	NLSU	β_1	2.394	2.333	2.179
		β_2	1.605	1.591	1.480
200	H2S	β_1	2.157	1.907	1.468
		β_2	1.434	1.344	1.215
	WH2S	β_1	2.086	2.333	1.532
		β_2	1.419	1.594	1.295
	3SE	β_1	1.143	1.133	1.159
		β_2	1.147	1.070	1.067
	W3SE	β_1	1.114	1.267	1.425
		β_2	1.147	1.234	1.262
	MLE	β_1	1.000	1.000	1.000
		β_2	1.000	1.000	1.000
	NLSU	β_1	2.829	2.320	2.011
		β_2	1.620	1.476	1.469
400	H2S	β_1	2.062	1.714	1.380
		β_2	1.333	1.322	1.169
	WH2S	β_1	1.979	2.286	1.422
		β_2	1.322	1.667	1.245
	3SE	β_1	1.167	1.089	1.197
		β_2	1.054	1.064	1.160
	W3SE	β_1	1.125	1.214	1.535
		β_2	1.053	1.269	1.462
	MLE	β_1	1.000	1.000	1.000
		β_2	1.000	1.000	1.000
	NLSU	β_1	3.104	2.161	1.915
		β_2	1.667	1.452	1.481

Table 6.17: Relative Root Mean Square Errors (RMSE) for all Sample Sizes and Distributions, given 50% Degree of Censoring.

Sample Size	Estimator	Para- -meter	Relative RMSE when the errors are Normal	Students'-t	Chi-Square
100	H2S	β_1	3.123	2.270	2.658
		β_2	2.005	1.843	1.697
	WH2S	β_1	2.938	2.083	2.848
		β_2	1.907	2.284	1.771
	3SE	β_1	1.192	1.084	1.408
		β_2	1.251	1.097	1.098
	W3SE	β_1	1.108	1.539	1.413
		β_2	1.214	1.458	1.185
	MLE	β_1	1.000	1.000	1.000
		β_2	1.000	1.000	1.000
	NLSU	β_1	3.331	3.225	2.837
		β_2	2.483	1.432	1.829
200	H2S	β_1	3.170	2.352	2.562
		β_2	1.962	1.731	1.694
	WH2S	β_1	2.925	3.374	2.687
		β_2	1.849	2.401	1.770
	3SE	β_1	1.212	1.014	1.484
		β_2	1.101	1.065	1.184
	W3SE	β_1	1.138	1.589	1.547
		β_2	1.063	1.681	1.347
	MLE	β_1	1.000	1.000	1.000
		β_2	1.000	1.000	1.000
	NLSU	β_1	3.808	3.388	3.094
		β_2	2.579	2.808	2.571
400	H2S	β_1	3.393	1.877	2.436
		β_2	1.945	1.703	1.655
	WH2S	β_1	3.147	3.109	2.574
		β_2	1.844	2.662	1.698
	3SE	β_1	1.295	0.740	1.713
		β_2	1.202	0.903	1.201
	W3SE	β_1	1.197	1.383	1.755
		β_2	1.174	1.703	1.403
	MLE	β_1	1.000	1.000	1.000
		β_2	1.000	1.000	1.000
	NLSU	β_1	3.492	2.055	2.064
		β_2	2.174	2.214	1.676

distributions; however, they still remain high compared to other estimators.

The NLSU estimator is followed by the H2S estimator and its weighted version, the WH2S estimator. As shown in Tables 6.16 and 6.17, both estimators exhibit very high ratios of relative RMSE's compared to, say, the 3SE or MLE estimators. Further, Table 6.17 depicts that, based on the relative RMSEs, the relative performance of the H2S and WH2S estimators deteriorates for higher levels of censoring. In general, the H2S and WH2S estimators are no closer to the MLE estimator or to the 3SE estimator. Note that, as discussed earlier in this Chapter, the WH2S estimator provides slightly more efficient estimates than the unweighted H2S estimator only when the errors are normal. This is also true for the W3SE and 3SE estimators.

On the other hand, it is important to note that the relative RMSE values indicate that the 3SE estimator provides results which are comparable to the MLE estimates. Given a low degree of censoring, the relative RMSE for 3SE estimates remains very close to the MLE estimates for normal and students'-t distributions. The relative performance of the 3SE estimator becomes better under students'-t distributed errors for all samples. More interestingly, or perhaps surprisingly, the evidence in Table 6.17 indicates that, as the degree of censoring increases, the relative performance of the 3SE estimator under the students'-t distribution, compared to the MLE estimator, improves significantly. For example, given 50% degree of censoring, a sample size of 400 and students'-t distributed error terms, the relative RMSE values for the 3SE estimates of β_1 and β_2, are, respectively, about 26 and 10 percent less than the corresponding relative RMSEs of MLE estimates. Results for 75% degree of censoring and large sample size also showed that the 3SE estimator performs better than the MLE estimator, if the errors have students'-t distribution. This is, however, not the case for the H2S estimator, which deteriorates for higher degrees of censoring, and its relative RMSEs remain significantly larger than both the MLE and 3SE estimators in all cases.

6.4.1 Comparison of Variances and Hypothesis Testing

In the discussions so far, we have considered the relative performance of the various estimators, mainly focusing on the unbiasedness and efficiency of estimates with respect to changes in sample size, distribution of the error term and degree of censoring. However, equally important in applied research is to see the performance of the estimators viz-a-viz statistical inference. That is, to investigate the performance of the estimators in terms of their relative reliability for hypothesis testing and/or construction of confidence intervals for the coefficients of the model. One of the important components in relation to this is the estimation of variances of the estimates for each estimator. Note that, in practice, the calculation of t-statistics for hypothesis testing and the construction of confidence intervals for the coefficients of the model involve the estimation of variances (and hence the standard errors) of the estimates for a given sample. These variances are obtained by substituting the sample estimates into the respective expressions of the asymptotic variance-covariance matrices of each estimator. For example, the variances of the coefficients using the MLE estimator are obtained as the sample estimates of the diagonal elements of the inverse of the information matrix (see equation (2.21) of Chapter 2). Similarly, the variances of the coefficients using other estimators are obtained based on their respective asymptotic variance-covariance expressions. It is, however, important to recall that the expressions for the asymptotic variance-covariance matrices of the estimators are derived on the assumption of normality of the error terms. Thus, given that the asymptotic variances are derived based on the assumption of normality of the error terms, it is important to ask how good the asymptotic variances approximate their respective Monte Carlo (true) variances under the different situations.

One of the main purposes of this Section is therefore to compare the asymptotic variances of the estimators with their respective Monte Carlo variances. That is, to see whether the asymptotic variances provide good (close) approximations of their respective true variances under a variety of error distributions, sample sizes and degrees of censoring. In general, asymptotic variances which are relatively larger than their corresponding true variances may imply confidence intervals of coefficients which are wider than the desired level, or alternatively, the probability of rejecting a true hypothesis becomes higher than it should be. On the other hand, asymptotic variances which are lower than their corresponding true variances imply confidence intervals which are relatively narrow compared to the desired level.

Tables 6.18 and 6.19 present the asymptotic as well as true variances of the various estimators for 25% and 50% degrees of censoring, respectively. Note that the asymptotic variances of the estimators are obtained by substituting actual values into the respective analytical (asymptotic) formulas of the variance-covariance matrices of the estimators, which are provided in Chapters 2 and 3 of this study. These are then compared with the Monte Carlo (true) variances of the estimators which are obtained based on the 3000 replications (samples) using equation (5.13) of Chapter 5 of this study.

As shown in Tables 6.18 and 6.19, just by considering the asymptotic variances of the estimators one can observe the vast difference in relative efficiency of the various

Table 6.18: Comparison of Variances of Estimators for all Sample Sizes and Distributions, given 25% Degree of Censoring.

Sample Size	Estimator	Para- -meter	Asymptotic Variance	Monte Carlo (true) variance under Normal	Students'-t	Chi-Square
100	H2S	β_1	0.0654	0.0361	0.0294	0.0298
		β_2	0.0781	0.0549	0.0490	0.0536
	WH2S	β_1	0.0631	0.0344	0.0357	0.0336
		β_2	0.0763	0.0544	0.0585	0.0605
	3SE	β_1	0.0135	0.0129	0.0136	0.0139
		β_2	0.0379	0.0346	0.0325	0.0407
	W3SE	β_1	0.0133	0.0126	0.0173	0.0167
		β_2	0.0375	0.0348	0.0394	0.0492
	MLE	β_1	0.0094	0.0094	0.0112	0.0136
		β_2	0.0328	0.0348	0.0343	0.0454
	NLSU	β_1	0.0476	0.0518	0.0537	0.0568
		β_2	0.0710	0.0787	0.0769	0.0829
200	H2S	β_1	0.0310	0.0225	0.0191	0.0182
		β_2	0.0375	0.0341	0.0286	0.0319
	WH2S	β_1	0.0301	0.0212	0.0263	0.0208
		β_2	0.0368	0.0331	0.0378	0.0372
	3SE	β_1	0.0071	0.0063	0.0070	0.0076
		β_2	0.0198	0.0216	0.0188	0.0212
	W3SE	β_1	0.0070	0.0060	0.0091	0.0093
		β_2	0.0196	0.0218	0.0250	0.0259
	MLE	β_1	0.0047	0.0048	0.0048	0.0061
		β_2	0.0170	0.0174	0.0164	0.0214
	NLSU	β_1	0.0308	0.0371	0.0249	0.0307
		β_2	0.0463	0.0429	0.0356	0.0463
400	H2S	β_1	0.0153	0.0098	0.0079	0.0086
		β_2	0.0193	0.0153	0.0139	0.0150
	WH2S	β_1	0.0148	0.0090	0.0122	0.0100
		β_2	0.0189	0.0149	0.0197	0.0176
	3SE	β_1	0.0034	0.0031	0.0033	0.0036
		β_2	0.0098	0.0096	0.0095	0.0108
	W3SE	β_1	0.0033	0.0030	0.0045	0.0042
		β_2	0.0097	0.0097	0.0139	0.0135
	MLE	β_1	0.0023	0.0024	0.0031	0.0032
		β_2	0.0086	0.0082	0.0084	0.0105
	NLSU	β_1	0.0141	0.0212	0.0144	0.0153
		β_2	0.0203	0.0234	0.0182	0.0224

Table 6.19: Comparison of Variances of Estimators for all Sample Sizes and Distributions, given 50% Degree of Censoring.

Sample Size	Estimator	Para- -meter	Asymptotic Variance	Monte Carlo (true) variance under Normal	Students'-t	Chi-Square
100	H2S	β_1	0.4929	0.1644	0.1489	0.2375
		β_2	0.5141	0.1850	0.1782	0.2513
	WH2S	β_1	0.4744	0.1456	0.2214	0.2717
		β_2	0.4971	0.1678	0.2671	0.2753
	3SE	β_1	0.0495	0.0237	0.0355	0.0372
		β_2	0.0720	0.0723	0.0655	0.0885
	W3SE	β_1	0.0476	0.0233	0.0677	0.0384
		β_2	0.0699	0.0682	0.1090	0.1072
	MLE	β_1	0.0167	0.0157	0.0327	0.0282
		β_2	0.0455	0.0449	0.0545	0.0644
	NLSU	β_1	0.2735	0.1409	0.2792	0.2349
		β_2	0.2560	0.2195	0.2766	0.2597
200	H2S	β_1	0.2092	0.0889	0.0838	0.1026
		β_2	0.1893	0.0972	0.0851	0.1053
	WH2S	β_1	0.2001	0.0756	0.1707	0.1187
		β_2	0.1819	0.0863	0.1604	0.1201
	3SE	β_1	0.0213	0.0130	0.0191	0.0169
		β_2	0.0327	0.0307	0.0364	0.0431
	W3SE	β_1	0.0204	0.0114	0.0409	0.0190
		β_2	0.0320	0.0285	0.0815	0.0534
	MLE	β_1	0.0083	0.0091	0.0176	0.0147
		β_2	0.0234	0.0266	0.0316	0.0414
	NLSU	β_1	0.1928	0.0695	0.1075	0.1192
		β_2	0.1935	0.1108	0.1398	0.1982
400	H2S	β_1	0.1164	0.0429	0.0512	0.0467
		β_2	0.1119	0.0449	0.0438	0.0463
	WH2S	β_1	0.1113	0.0367	0.1519	0.0584
		β_2	0.1073	0.0401	0.1101	0.0559
	3SE	β_1	0.0109	0.0062	0.0104	0.0082
		β_2	0.0170	0.0170	0.0166	0.0202
	W3SE	β_1	0.0105	0.0053	0.0307	0.0089
		β_2	0.0167	0.0164	0.0521	0.0256
	MLE	β_1	0.0040	0.0040	0.0135	0.0070
		β_2	0.0119	0.0112	0.0163	0.0167
	NLSU	β_1	0.0817	0.0413	0.0638	0.0296
		β_2	0.0798	0.0525	0.0772	0.0427

estimators. Specifically, the large values of the asymptotic variances of the H2S, WH2S and the NLSU estimates reveal their relative inefficiency compared to those of the MLE, 3SE and W3SE estimates in all cases, i.e., for all samples, distributions and degrees of censoring. For example, Table 6.18 depicts that, given a sample size of 100 and a 25% degree of censoring, the asymptotic variances of the MLE estimates of β_1 and β_2 are, respectively, 19.7 and 46.2 percent of the corresponding NLSU estimates. These values drop only slightly to 16.3 and 42.4 percent, respectively, if the sample size becomes large (400). Similarly, the asymptotic variances of the H2S and the WH2S estimates are substantially high compared to their respective MLE or 3SE estimates. These results are consistent with those discussed in the preceding sections. However, as stated earlier in this Section, the main interest here is not the efficiency of the estimators, but to compare the asymptotic variances viz-a-viz their corresponding true variances which has further implications for hypothesis testing and construction of confidence intervals for the coefficients of the model. The main points along this line are discussed as follows.

As can be seen from Tables 6.18-6.19, it is evident that the NLSU estimates are not only inefficient but the asymptotic variances of the NLSU estimates are also not good approximations of the corresponding true variances; and the difference between the asymptotic and the Monte Carlo variances worsens as the degree of censoring increases. Note that, as mentioned earlier in this Chapter, estimation using the NLSU estimator is very slow and convergence is not always guaranteed.

Regarding the MLE estimates, the asymptotic variances more or less provide very close (sometimes accurate) approximations of the true variances, given that the errors have a normal distribution. This is particularly true for larger samples. Further, it is also interesting to note that the MLE estimator seems to perform fairly well under the students'-t distribution, given that the degree of censoring is low. Whereas, for the chi-square distributed error terms, the true variances of the MLE estimates appear to be slightly greater than their respective asymptotic variances in all cases.

Tables 6.18-6.19 also reveal interesting results with regard to the variances of the H2S and 3SE estimators. The true variances of the H2S estimates are significantly lower than their respective asymptotic variances in all cases. For example, given a 25% degrees of censoring, a sample size of 400 and normally distributed error terms, the asymptotic variances of the H2S estimates of β_1 and β_2 are, respectively, about 56 and 26 percent larger than their respective true variances (see Table 6.18). The results for the non-normal distributions are also similar. Furthermore, the gap between the asymptotic variances and the true variances of the H2S estimates widens for higher levels of censoring (see Table 6.19). In general, this results indicate that confidence intervals based on the H2S estimates are likely to be wider (less precise) than they should be.

On the other hand, the situation is different for the 3SE estimator in which the asymptotic variances yield relatively good approximations of the true variances in most cases. It is also evident that the 3SE estimator appears to be robust under the non-normal distributions provided that the degree of censoring is low. In general, one can conclude the following important points based on the variances of the H2S and 3SE estimators. One, the results indicate that the asymptotic variances of the H2S estimates are not good approximations of their respective true variances. That is, the asymptotic variances generally overstate their corresponding true variances of the estimates in all cases. Two, the 3SE estimates, are not only more efficient than their corresponding H2S estimates, but also provide asymptotic variances which are closer to their corresponding true variances in most cases. This indicates that the 3SE appears to have better finite sample properties than the H2S estimator in almost all cases.

Note that the above comparisons provide important indications on the likely performances of the estimators in terms of their reliability for hypothesis testing and/or construction of confidence intervals for the coefficients of the model. But, a more accurate depiction of the implications of the variances can be obtained by conducting hypothesis tests (such as t-tests) and constructing confidence intervals for the coefficients. Thus, the remaining part of this Section provides a further investigation on the implications of the variances for hypothesis testing and/or construction of confidence intervals for the coefficients of the model. Specifically, we test the hypotheses:

$$H_0 : \beta_k = 1 \qquad (6.6)$$

$$H_1 : \beta_k \neq 1, \quad k = 1, 2. \qquad (6.7)$$

To test the hypotheses we use the test statistic:

$$t = \frac{\hat{\beta}_k - 1}{s.e.(\hat{\beta}_k)} \qquad (6.8)$$

where under the null hypothesis the statistic t is asymptotically distributed as a standard normal random variable, $\hat{\beta}_k$ is the sample estimate of β_k and $s.e.(\hat{\beta}_k)$ is the standard error of of $\hat{\beta}_k$, which is obtained as the square-root of the sample estimates of the diagonal elements of the variance-covariance matrix for each estimator in a given sample. A nominal 5% level of significance is considered so that the expected percentage of rejections whenever the null hypothesis is true is equal to 5%.

Or equivalently, a 95% confidence interval can be constructed such that:

$$P[\hat{\beta}_k - z \times s.e.(\hat{\beta}_k) < \beta_k < \hat{\beta}_k + z \times s.e.(\hat{\beta}_k)] = 0.95 \tag{6.9}$$

which is equivalent to

$$P(-1.96 < t < 1.96) = 0.95 \tag{6.10}$$

where t is defined by (6.8) and the standard z value at 5% significant level is approximately 1.96 for large N.

Given these circumstances, we obtained the percent of coefficients contained in the 95% confidence intervals for a few selected estimators. Note that the weighted versions of the H2S and 3SE estimators, i.e., the WH2S and W3SE are not included in the following analysis for two main reasons. (i) The computation of the variances (and hence the standard errors) of the estimates involve the inversion of an NxN matrix at each iteration in the experiment, which is quite slow and unattractive. This is also likely to be impractical for applied research especially if the sample size is large (e.g., using survey data). (ii) The WH2S and W3SE estimators are quite sensitive to violations of the assumptions about the error term, which is usually the case in applied research. Similarly, the NLSU estimator is excluded for its relatively poor performance. This leads us to concentrate on the comparison of the remaining three candidates, namely, the H2S, 3SE and the MLE estimators. Note that, as discussed in Chapters 2 and 3, the variances (standard errors) of the H2S and the 3SE estimators can be obtained using one of two alternatives. The first one is by substituting the elements of the respective covariance matrices by their consistent estimates. The second procedure is based on White's (1980b) idea. However, hypothesis tests and/or confidence intervals based on the first procedure were far from the desired precision, especially for the H2S estimator [see Tessema (1994)]. Thus, the standard errors for the H2S and 3SE estimators which are used to obtain the results in Tables 6.20-6.21 are estimated following White's (1980b) procedure.

Table 6.20 presents the percent of coefficients contained in the 95% confidence intervals for all sample sizes and distributions, given a 25% degree of censoring. These results reveal the following important points.

Under normality of the error terms, the MLE provides confidence intervals which are quite close (approximately the same) to the expected closure rate, which is 95%. For example, Table 6.20 depicts that, given a 25% degree of censoring, a sample size of 200 and normally distributed error terms, about 94.17 and 95.43 percent of the 3000 confidence intervals (samples) contain the true parameters, β_1 and β_2, respectively. The corresponding percentages for the 3SE estimator are, 92.73 and 81.73 for β_1 and β_2, respectively. The results for the H2S estimator are close to those of the 3SE. In general, as would be expected, the MLE appears to provide more reliable confidence intervals provided that the errors are normal. However, the situation appears to be relatively different under the non-normal distributions. For example, the evidence in Table 6.20 indicates that the H2S and the 3SE estimators can be as good as and sometimes even better than the MLE under the non-normal distributions (see results under the students'-t distribution).

Further, to examine the effects of censoring, Table 6.21 depicts the percent coefficients contained in the 95% confidence intervals for all samples and distributions, given a censoring level of 50%. As can be seen from the table, the results for both the H2S and the 3SE estimators appear to be larger than the desired level in most

Table 6.20: 95% Confidence Intervals of Estimators for all Sample Sizes and Distributions, given 25% Degree of Censoring.

Sample Size	Estimator	Para- -meter	% of Coefficients Contained in 95% C.I.*		
			Normal	Students'-t	Chi-Square
100	H2S	β_1	92.80	95.60	91.03
		β_2	81.97	88.87	82.97
	3SE	β_1	93.23	95.93	91.93
		β_2	81.17	88.93	83.60
	MLE	β_1	94.43	92.57	91.80
		β_2	93.93	93.50	93.93
200	H2S	β_1	92.23	95.67	91.00
		β_2	81.43	87.80	83.90
	3SE	β_1	92.73	96.03	92.03
		β_2	81.73	87.78	84.27
	MLE	β_1	94.17	93.47	89.50
		β_2	95.43	94.30	94.80
400	H2S	β_1	93.27	95.50	91.17
		β_2	83.13	90.10	85.03
	3SE	β_1	94.10	95.80	92.47
		β_2	83.43	90.27	85.97
	MLE	β_1	94.43	90.37	87.00
		β_2	95.87	92.20	94.10

—* C.I. stands for Confidence Interval.

Table 6.21: 95% Confidence Intervals of Estimators for all Sample Sizes and Distributions, given 50% Degree of Censoring.

Sample Size	Estimator	Para- -meter	% of Coefficients Contained in 95% C.I.*		
			Normal	Students'-t	Chi-Square
100	H2S	β_1	90.60	93.60	87.90
		β_2	88.63	92.83	84.20
	3SE	β_1	99.66	99.00	98.30
		β_2	98.43	97.43	96.67
	MLE	β_1	95.13	91.60	92.73
		β_2	94.10	93.80	92.50
200	H2S	β_1	99.43	97.17	97.67
		β_2	98.10	96.83	95.93
	3SE	β_1	99.63	99.10	98.83
		β_2	98.27	97.57	96.50
	MLE	β_1	94.80	87.70	91.47
		β_2	94.03	93.50	93.93
400	H2S	β_1	99.73	97.93	97.77
		β_2	98.23	98.07	96.17
	3SE	β_1	99.87	98.90	98.86
		β_2	98.33	98.50	96.57
	MLE	β_1	94.93	82.77	88.97
		β_2	94.47	92.73	94.50

—* C.I. stands for Confidence Interval.

cases. This is in line with the earlier comparison of the variances especially for the H2S estimates.

As before, the MLE performs quite well under the normal distribution but not as good as under the non-normal distributions. However, it should be noted that when using the MLE estimator one can obtain the variances (and hence the standard errors) of the estimates using any one of four alternative, but asymptotically equivalent, variance-covariance estimators, of which the inverse of the information matrix used in the above analysis is among these alternatives. It is therefore interesting to see whether the performance of the MLE estimator improves or not by making use of other variance-covariance matrix estimators. A further analysis along this line will be provided in the next Chapter.

6.5 Summary and Conclusions

In this Chapter we examined the small sample properties of some of the estimators of the tobit model. These estimators include, two ordinary least squares estimators: one using only the positive (non-limit) observations on the dependent variable, y_i, (OLSP) and the other using all limit and non-limit observations on y_i (OLS), the Heckman's two-step estimator (H2S) and its weighted version, the weighted Heckman's two-step estimator (WH2S), Heckman's two-step estimator based on the unconditional expectation of the model (H2SU) and its weighted version, the WH2SU, the maximum likelihood estimator (MLE) and two nonlinear least squares estimators. Further, a three-step estimation procedure which is referred to as the three-step estimator (3SE) and its weighted version, the weighted three-step estimator (W3SE), are also suggested and investigated. The effects of sample size, degree (level) of censoring and distributional assumptions of the error structure of the model are investigated. The main conclusions, among others, are the following.

The least squares estimators are seriously affected by the degree of censoring and provide biased estimates; the bias being an increasing function of the degree of censoring. Similarly, the H2SU and the WH2SU estimators are no better than the least squares estimators. In other words, the H2SU and the WH2SU estimators are biased and inefficient and get worse with increases in the degree of censoring.

The nonlinear least squares estimators are generally less efficient and computationally very slow compared to the MLE or 3SE estimators. Most importantly, the nonlinear least squares estimators are sensitive to the degree of censoring, and convergence is not always guaranteed.

Under normality conditions, the MLE estimator gives the best results followed by the 3SE estimator. The loss in efficiency of the 3SE estimator compared to the MLE estimator is marginal. However, both the MLE and 3SE estimators appear to be sensitive for the skewed (chi-square) distributed error terms in terms of efficiency. On the other hand, given low levels of censoring, the MLE estimator performs well under the students'-t distribution. If the degree of censoring is high, the MLE estimates under students'-t distribution can be less efficient than the 3SE estimator. It is also important to note that the t-tests and confidence intervals based upon the MLE estimates are quite good under the normal compared to those of the non-normal distributions.

Further, our results indicate that the H2S estimator, although less efficient compared to the 3SE or MLE estimators in all cases, seems to perform well in terms bias, given low levels of censoring. However, it can be highly inefficient for degrees of censoring as high as 50%. This result is in contrast to that of the 3SE estimator which yields results which are very close to the MLE estimates.

Finally, recall that, as discussed in Chapter 2, previous studies related to tobit estimators have indicated that the MLE can be biased if the assumption of normality about the error term does not hold. However, our results do not generally support this conclusion. Specifically, the evidence in this Chapter shows that bias is not a serious problem for the MLE if we assume normality when the errors are generated from the students'-t and chi-square distributions; except when the sample size is small (100), coupled with high degree of censoring, in which case there appears to be some bias under the non-normal distributions, particularly under the students'-t distribution (e.g., see Table A.6, Appendix A).

Given that the MLE estimator is widely used in applied research, a further investigation of its consistency as well as the use of the alternative but asymptotically equivalent variance-covariance matrices and their effects on the performance of the MLE estimator will be provided in the next Chapter.