Chapter 7

The MLE: Consistency, Variance Estimation and Hypothesis Testing

7.1 Introduction

The maximum likelihood estimator is widely used in estimation and inference for many applied studies of tobit models [see Deagan and White (1976), Jarque (1987), Addesina and Zinnah (1993)]. On the other hand, some studies have indicated that, unlike the regular regression models, the MLE of the tobit model may have serious consequences such as inconsistency if we assume normality of the error term and this assumption is incorrect.

The purpose of this Chapter is to investigate the properties of the MLE estimator further so that, combined with the results presented in Chapter 6, we may be able to make some specific conclusions. Specifically, the main objectives of this Chapter can be summarised as follows:

1. To provide more information on the consistency (or inconsistency) of the MLE under various error distributions based upon the asymptotic results of the MLE estimator. It is important to note that, throughout the Chapter, the asymptotic results that will be used in this analysis are those derived under the correct specification of the model, i.e., normality of the error terms. These results will then be evaluated empirically using observations generated from various distributions. In other words, the purpose of this Chapter is to investigate the effects of misspecification by making use of the asymptotic results of the estimator.

2. The second major objective is to examine the effects of the alternative, but asymptotically equivalent, variance-covariance matrix estimators suggested in the literature for estimating the variances of the coefficients of the model. In practice, one can use any one of (four) alternative variance-covariance matrix estimators, of which three of them are usually associated with particular algorithms used for maximization to obtain the ML estimates. However, the choice of one of these variance-covariance matrix estimators is unlikely to be neutral for estimation of the variances if the assumption of normality does not hold. Thus, it is important to examine the unbiasedness (or robustness) of the alternative variance-covariance matrix estimators in the presence of misspecification.

3. Another main purpose which is closely related to the variance-covariance matrix estimators is their implications for hypothesis testing and/or construction of

confidence intervals for the coefficients of the model. That is, the difference in magnitude between the true and the estimated variances from the alternative variance-covariance estimators may not necessary lead to different conclusions in hypothesis testing unless the differences are substantial (significant).

It is therefore the purpose of this Chapter to examine whether the different variance-covariance estimators lead to the same conclusions in hypothesis testing and/or construction of confidence intervals for the coefficients of the model.

But first, we summarize some of the findings of the preceding Chapter regarding the maximum likelihood estimator.

As we have discussed in Chapter 6, the Monte Carlo results indicate, among others, three major points regarding the MLE of the model. These points may be summarized as follows:

(i) Under normality of the error terms, the MLE performs well in terms of bias. Further, although there seems to be some bias for the students'-t and chi-square distributions when the sample size is small and the degree of censoring high, the MLE appears to do well otherwise.

(ii) The asymptotic variances of the MLE provide good approximations of the true (Monte Carlo) variances of the estimators, given that the errors have a normal distribution. This is consistent with our expectations since the expressions for the asymptotic variances are obtained under the assumption of normality of the error terms. But, interestingly, it is also evident that the MLE performs fairly well if we assume normality when in fact the errors have the students'-t distribution. However, this is not true for the chi-square distributed errors. That is, the Monte Carlo variances under chi-square distributed errors overestimate their corresponding asymptotic variances in almost all cases.

(iii) Not surprisingly, the t-tests and 95% confidence intervals for the coefficients of the model are quite good under the normal distribution. However, results for the non-normal distributions are relatively inferior than those for the normally distributed errors.

The above points raise some important questions regarding the performance of the MLE estimator. (i) Given that the biases reported for the students'-t and chi-square distributions decline as the sample size increases, can one conclude, on the basis of this information, that the MLE is sensitive to the violations of the assumptions about the error term of model. (ii) The evidence on the variances, t-tests and confidence intervals suggests that the results for the MLE under the non-normal distribution does not seem to be as good as those of the normal distribution. Once again, does this imply that the MLE is not robust under the non-normality of the disturbances of the model. The remaining sections of this Chapter present further investigation on the properties of the MLE which may be helpful to answer these and other related questions.

Specifically, Section 7.2. investigates the consistency (or inconsistency) of the MLE under the various distributional assumptions of the model. Section 7.3. presents a detailed analysis on the use of the alternative variance-covariance matrix estimators of the MLE and their implications for hypothesis testing and/or construction of confidence intervals for the coefficients. Finally, Section 7.4. summarises the results.

7.2 Consistency of the ML Estimator

This Section evaluates whether the MLE is consistent or not under various distributional assumptions. In order to do this, we need to discuss briefly the asymptotic results of the MLE first. Note that, as stated above, the asymptotic results are derived on the assumption of normality of the error terms of the model. These results are then evaluated empirically, assuming normality, while the errors are generated from the different distributions, in order to examine the effects of misspecification, which is a common problem in applied research.

Recall the likelihood function of the tobit model which is given by

$$L = \prod_0 [1 - F_i] \prod_1 \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-(y_i - x_i'\beta)/2\sigma^2\} \tag{7.1}$$

where the first product is evaluated over the N_0 observations for which $y_i = 0$ and the second product is evaluated over N_1 observations for which $y_i > 0$. Note that f_i and F_i are, as defined before, the density and distribution functions of the standard normal distribution, respectively.

The log-likelihood function is given as

$$\log L = \sum\nolimits_0 \log(1 - F_i) + \sum\nolimits_1 \log \frac{1}{(2\pi\sigma^2)^{1/2}} - \sum\nolimits_1 \frac{1}{2\sigma^2}(y_i - x_i'\beta)^2 \tag{7.2}$$

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

Note that the term 'normal-based' will be used frequently in this Chapter to indicate that the asymptotic results (expressions) are derived assuming normality of the error terms.

Given this, the maximum-likelihood estimators of the parameters of the tobit model are obtained as a solution of the first partial derivatives. That is, if $\tilde{\theta}$ is a MLE of $\theta = (\beta', \sigma^2)$ then it should satisfy the first order condition[1]

$$\left.\frac{\partial \log L}{\partial \theta}\right|_{\tilde{\theta}} = 0 \tag{7.3}$$

In general, solving equation (7.3) provides equations which are non-linear in θ. This implies that an iterative procedure should be used to generate a value for the ML estimator $\tilde{\theta}$. It also implies that since we do not have explicit expressions for $\tilde{\theta}$, we rely on the properties of $\tilde{\theta}$ of which equation (7.3) plays an important role. Some of the important results are presented below in a general context.

Under certain regularity conditions, it is straightforward to show that [see Dhrymes (1970, p. 114), Stewart (1991, p. 123), Davidson and Mackinnon (1993, p. 255)],

$$E\left[\frac{\partial \log L}{\partial \theta}\right] = 0 \tag{7.4}$$

and

$$V\left[\frac{\partial \log L}{\partial \theta}\right] = E\left[-\frac{\partial^2 \log L}{\partial \theta \partial \theta'}\right] = \mathcal{I}(\theta) \tag{7.5}$$

That is, (7.3) is a random variable with mean zero and variance given by the information matrix, $\mathcal{I}(\theta)$. This is usually denoted as:

$$\frac{\partial \log L}{\partial \theta} \sim [0,\ \mathcal{I}(\theta)] \tag{7.6}$$

Further, equation (7.5) implies that

[1]The full expressions of first and second derivatives of the normal-based log-likelihood function are given in (2.16)-(2.20) in Chapter 2 of this study.

$$E\left[-N^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right] = N^{-1}\mathcal{I}(\theta) \tag{7.7}$$

Note that, under certain regularity conditions, it is assumed that the expression in (7.7) converges in probability to a nonsingular matrix Q such that

$$\text{plim}\left[-N^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right] = \lim_{N\to\infty} N^{-1}\mathcal{I}(\theta) = Q \tag{7.8}$$

Now, given the above results, expanding (7.3) in a Taylor series around the value θ gives

$$\left.\frac{\partial \log L}{\partial\theta}\right|_{\tilde{\theta}} = 0 = \frac{\partial \log L}{\partial\theta} + \frac{\partial^2 \log L}{\partial\theta\partial\theta'}(\tilde{\theta}-\theta) + R \tag{7.9}$$

where R is the remainder of the series.

Assuming that the remainder, R, is relatively small and the second derivative exists and is non-singular, we have

$$(\tilde{\theta}-\theta) \simeq -\left[\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1}\frac{\partial \log L}{\partial\theta} \tag{7.10}$$

Equation (7.10) is very important and shows how one can overcome the difficulties of not having an explicit expression for the MLE. Further, given the arguments in (7.7) and (7.9) and using equation (7.10) we have

$$(\tilde{\theta}-\theta) = \left[-N^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1} N^{-1}\frac{\partial \log L}{\partial\theta} \tag{7.11}$$

which implies that for $\tilde{\theta}$ to be consistent

$$\begin{aligned}\text{plim}(\tilde{\theta}-\theta) &= \left[\text{plim} - N^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1}\text{plim}\, N^{-1}\,\frac{\partial \log L}{\partial\theta} \\ &= Q^{-1}\times 0 = 0 \end{aligned}\tag{7.12}$$

The relationship in (7.12) holds for ML estimators in general if the model is correctly specified. Similarly, Amemiya (1973) has shown that, under the assumptions of the model, it is also true for the tobit MLE. The main objective of this Section is, however, to investigate whether (7.12) holds when it is derived from (7.2), but, in fact the errors are non-normal. Specifically, the main interest is to evaluate (7.12) empirically using observations generated from several distributions when in fact the first and second derivatives are derived from the normal-based log-likelihood function.

To discuss (7.12) further, let $I_N(\theta)$ indicate the empirical inconsistency obtained for a given distribution of the error term and a sample size, given that θ is known. That is, define

$$I_N(\theta) = \frac{1}{M}\sum_{m=1}^{M} - \frac{1}{N}\left[\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1} N^{-1}\frac{\partial \log L}{\partial\theta}, \qquad M = 1, 2, ..., 3000. \qquad (7.13)$$

given the derivatives come from the normal-based likelihood. It is also important to note that (7.13) uses the actual parameters but the $y_i's$ are generated based on the various distributions.

Given this, there are two important points that can be deduced from the relationships in (7.12)-(7.13) above.

(i) Note that from (7.12) $I_N(\theta)$ must be equal to zero for sufficiently large sample size, given that the model is correctly specified. That is, if we evaluate (7.13) using observations generated from the normal distribution we expect the value of $I_N(\theta)$ to be zero or sufficiently close to zero (i.e., $I_N(\theta) \simeq 0$). This is because, as discussed above, the results in (7.12)-(7.13) are derived from the normal-based log-likelihood function given in (7.2).

(ii) However, if (7.12)- (7.13) are derived from the normal-based log-likelihood function and if we evaluate (7.13) using observations which are generated from the non-normal distributions, then the empirical inconsistency $I_N(\theta)$ may not be equal to zero and its magnitude gives an indication of the inconsistency under misspecification of the log-likelihood.

The advantage of using (7.13) as an indicator in the analysis of consistency is that it provides for the effects of the error distribution explicitly by exploiting the asymptotic properties of the MLE estimator. Further, it can be computed without having to obtain values for the MLE during the simulation run.

Given this, we compared the consistency (inconsistency) of the MLE under various distributions, sample sizes and degrees of censoring. The results of the Monte Carlo comparison are discussed in the following section.

7.2.1 Monte Carlo Comparisons for Consistency

Below, we obtain results for the three distributions; namely, the normal, students'-t and chi-square distributions. For each distribution, empirical inconsistencies of the coefficients are estimated using (7.13) over $M = 3000$ replications for the small, medium and large samples, and varying degrees of censoring based on the experimental design discussed in Chapter 5. The results are summarized in Table 7.1. Note that, although our discussion focuses mainly on the results related to β_1 and β_2, we have reported results for the constant term, β_0, as well.

As can be seen from Table 7.1, the empirical inconsistencies of all coefficients under normality of error the terms are as expected. That is, the values of the empirical

Table 7.1: Empirical Inconsistency ($I_N(\theta)$) of the MLE Estimator for the three distributions and Sample Sizes.

Sample Size	Para- -meter	25% Degree of Censoring			50% Degree of Censoring		
		Normal	Students'-t	Chi-square	Normal	Students'-t	Chi-square
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
100	β_0	0.0040	-0.0714	-0.1220	-0.0000	-0.2150	-0.2107
	β_1	0.0000	0.0290	0.0362	0.0010	0.0636	0.0389
	β_2	0.0043	0.0332	0.0216	-0.0091	0.0536	0.0292
200	β_0	-0.0058	-0.0597	-0.1106	-0.0033	-0.2171	-0.1919
	β_1	0.0021	0.0249	0.0324	0.0018	0.0767	0.0382
	β_2	0.0009	0.0228	0.0241	0.0004	0.0675	0.0343
400	β_0	-0.0005	-0.0584	-0.1419	0.0022	-0.2282	-0.2136
	β_1	0.0001	0.0242	0.0431	-0.0003	0.0686	0.0429
	β_2	-0.0004	0.0264	0.0283	0.0021	0.0498	0.0392

inconsistencies for the coefficients are equivalent (or almost equivalent) to zero, i.e., $I_N(\theta) \simeq 0$ for all sample sizes. This implies no inconsistency, given that the assumption of normality is correct. However, the results for the non-normal distributions appear to be slightly different from those of the normal distribution.

Under the students'-t distribution, the empirical inconsistency for both β_1 and β_2 is about 3 percent when the sample size is small, and about 2 percent for both medium and large samples, provided that the degree of censoring is low, [see Column (4) of Table 7.1]. These values, although relatively larger than their corresponding estimates under the normal distribution, may be considered sufficiently close to zero. This implies that, given a low level of censoring, there seems to be little (or no) inconsistency of coefficients for the t-distributed error terms. Whereas, the empirical inconsistency in β_0 is relatively higher and ranges from -0.0714 for the small sample size to -0.0584 for the large sample size.

The results under the skewed distribution (chi-square) are similar to that of the students'-t distribution for β_1 and β_2 for all sample sizes, given a 25% degree of censoring. That is, the average inconsistencies of β_1 and β_2 are between 2 to 4 percent for all sample sizes [see Column (5) of Table 7.1]. These results are comparable to the biases reported in Chapter 6. However, given a 25% degree of censoring, the inconsistency in β_0 for the chi-square distribution appears to be substantially larger. Further, the inconsistency in β_0 does not seem to decline even for larger sample sizes. In fact, in order to see whether the inconsistency of β_0 declines or not for very large sample sizes, we obtained results by considering a sample size of 1000 and a 25% degree of censoring; it was found that the inconsistency in β_0 under the skewed distribution remains as high as for the small sample sizes. Thus, except for β_0, the results in Table 7.1 indicate that inconsistency may not be a serious problem under the chi-square distribution as well, given that the degree of censoring is low.

Further, higher degrees of censoring were considered to examine whether or not the inconsistency of the estimators is affected by the proportion of limit observations on the dependent variable, y_i. The results for the three distributions and sample sizes, given a 50% degree of censoring are shown in Columns (6) to (8) of Table 7.1.

As before, the values for the empirical inconsistency of the coefficients under the normal distribution are equivalent (or approximately equivalent) to zero as would be expected. That is, the MLE estimator provides consistent estimates even for small sample sizes and high degrees of censoring, provided that the assumption of normality of the error of the model is actually correct. Again, the results for the non-normal distributions are slightly different.

Given the students'-t distribution, the empirical inconsistency of β_1 and β_2, although still small, almost doubled for the 50% degree of censoring compared to that for the 25% degree of censoring [see Column (7) of Table 7.1]. But most notably, the results also show that the inconsistency of β_0 has increased substantially with increases in the degree of censoring. For example, given a sample size of 100 and t-distributed error terms, the empirical inconsistency of β_0 is given by -0.2150 for the 50% degree of censoring compared to -0.0714 for the 25% degree of censoring. The results are similar for the medium and large samples implying that the inconsistency in β_0 can be substantial under the students'-t distribution and becomes as high as that of the chi-square distribution, if the degree of censoring is not small. On the other hand, the empirical inconsistencies of β_1 and β_2 under the skewed distribution does not appear to be affected much by the increase in the degree of censoring [see

Column (8) of Table 7.1].

Given these results, a few points can be generalized with regard to the consistency of the MLE estimators under the various distributions of the error terms.

> Obviously, no inconsistency is observed even for the small sample sizes and high degrees of censoring, provided that the errors are normally distributed. More importantly, the inconsistencies in β_1 and β_2 under the non-normal distributions, although they appear to be relatively larger than those of the normal distribution, are not generally substantial. In other words, it can be said that there appears to be little inconsistency of the MLE estimates under the students'-t and chi-square distributions except for the constant term of the model. As to the constant term of the model, it can be inconsistent under the skewed distribution even for larger samples and lower degrees of censoring. It is also clear that the constant term can be inconsistent under the students'-t distribution, if the the degree of censoring is not small.
>
> Note that the constant term, as shown in Chapter 2, is useful in estimating the probability and cumulative distributions of the normal random variable which in turn are important to estimate results such as the response of dependent variable, y, for a unit change in an explanatory variable, say x_1. Thus, in general, the constant term can be important in applied research, depending on the research objectives. If, however, one assumes that the constant term is not important, then the results discussed above indicate that the MLE estimator of the tobit model appears to be robust in terms of consistency of the coefficients to changes in distributional assumptions of the error term. The significance of

these results is that they are obtained by making use of the asymptotic theory and provide more information as to the effects of the error distribution without the need for obtaining the values for the MLE.

These results are also in agreement with those in Chapter 6 in which it was concluded that bias is not a serious problem in MLE for the three distributions, except for small samples and high degrees of censoring where the bias increases for the non-normal distributions.

Note that, as discussed in Chapter 2, Arabmazar and Schmidt (1981, 1982) indicated that the MLE estimator of the tobit model can be sensitive to the distribution of the error term. That is, if the assumption of normality is violated the MLE may lead to inconsistent estimates. This, in general, appears to be in contrast to the results discussed above. However, it should also be noted that the findings given in Arabmazar and Schmidt (1981, 1982) are based on analysis of a special case of the tobit model which contains only the constant term. We would have a similar finding if one had to make conclusions on the consistency of the coefficients in our model based only on the results reported for the constant term under the non-normal distributions. In other words, also noted by Arabmazar and Schmidt (1982, p.1055), the relevance of the conclusions from a model with only a constant term to other coefficients in a more general model can be questionable.

Finally, it should be noted that many applied researchers are more interested in making inferences about the population characteristics than in just the point estimates obtained based on a particular sample size. Under these circumstances, whether the normal-based MLE performs well or not under non-normal error terms depends on the reliability of inference procedures such as hypothesis tests and/or confidence intervals for the coefficients of the model. If it is true that the normal-based MLE provides reasonably reliable inferences from hypothesis testing and/or confidence intervals when the errors are non-normal, then there is very little need for other, probably more complicated, estimators of the model. That is, if normal-based procedures are robust, then there is not much gain from using more complicated estimators such as the non-parametric and/or semi-parametric estimators of the model.

The next Section presents a thorough examination of the performance of the MLE estimator in statistical inference under the three distributions and various sample sizes and degrees of censoring. More specifically, it evaluates the different variance-covariance estimators of the model and their implications for hypothesis testing and/or confidence intervals for the coefficients of the model. But first we present a brief review of the asymptotic distribution of the MLE estimator.

7.3 Asymptotic Distribution of the MLE

The asymptotic distribution of the MLE estimator follows from the results in Section 7.2 above. Once again, we summarize the useful normal-based asymptotic results below. Technical details and proofs will be excluded. However, appropriate references will be cited when necessary.

Following equation (7.11) and using an alternative scaling we have

$$\sqrt{N}\,(\tilde{\theta}-\theta)=\left[-N^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1} N^{-1/2}\,\frac{\partial \log L}{\partial\theta} \tag{7.14}$$

We know from (7.12) that the first part of the right hand expression of equation (7.14) converges asymptotically to a non-stochastic matrix. The only stochastic element in the right hand side expression of (7.14) is the later part, i.e., $N^{-1/2}\,\frac{\partial \log L}{\partial\theta}$. But from the results in (7.6) and by applying a central limit theorem (CLT), it can be shown that $N^{-1/2}\,\frac{\partial \log L}{\partial\theta}$ is asymptotically normal, [see Dhrymes (1970. p.123), Davidson and Mackinnon (1993, p.262)]. Therefore it follows that $\sqrt{N}\,(\tilde{\theta}-\theta)$ is also asymptotically normal. That is[2]

$$\sqrt{N}\,(\tilde{\theta}-\theta)\overset{a}{=}\left[-N^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1} N^{-1/2}\,\frac{\partial \log L}{\partial\theta} \tag{7.15}$$

It follows that, under certain regularity conditions, that the random variable $\sqrt{N}\,(\tilde{\theta}-\theta)$ is normally distributed with zero mean and asymptotic variance-covariance matrix given by

$$\begin{aligned} V\left(\sqrt{N}\,(\tilde{\theta}-\theta)\right) &= \operatorname{plim}\frac{1}{N}[\mathcal{H}(\theta)]^{-1}\operatorname{plim}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \log L_i}{\partial\theta}\frac{\partial \log L_i}{\partial\theta'}\right] \\ &\quad \operatorname{plim}\frac{1}{N}[\mathcal{H}(\theta)]^{-1} \end{aligned} \tag{7.16}$$

where $\mathcal{H}(\theta)$ is referred to as the *Hessian* matrix; and is defined as

$$\mathcal{H}(\theta)=\frac{\partial^2 \log L}{\partial\theta\partial\theta'}$$

and the *outer product of the gradient vector*, $\mathcal{G}(\theta)$, is defined by

[2]The symbol $\overset{a}{=}$ indicates that, asymptotically, both the left and the right hand side expressions have the same limiting distribution

$$\mathcal{G}(\theta) = \sum_{i=1}^{N} \frac{\partial \log L_i}{\partial \theta} \frac{\partial \log L_i}{\partial \theta'}$$

where $\partial \log L_i / \partial \theta$ is the contribution of the i^{th} observation to the gradient vector. The expression given by (7.16) can be simplified further. That is, assuming that the model is correctly specified and given suitable regularity conditions, the *Hessian*, $\mathcal{H}(\theta)$, and the *outer product*, $\mathcal{G}(\theta)$, matrices are asymptotically equivalent forms of the *Fisher's information matrix*, $-\mathcal{I}(\theta)$, [see Kendall and Stuart (1967 p.53-55), Davidson and Mackinnon (1993, p.263)]. An immediate consequence of this is that

$$\begin{aligned} -\text{plim}\frac{1}{N}[\mathcal{H}(\theta)] &= \text{plim}\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \log L_i}{\partial \theta} \frac{\partial \log L_i}{\partial \theta'}\right] \\ &= \lim \frac{1}{N}[\mathcal{I}(\theta)] \end{aligned} \tag{7.17}$$

where $\mathcal{I}(\theta)$ is as defined in (7.5).

Now, equation (7.17) implies that the variance-covariance matrix of the MLE can be estimated based on a number of asymptotically equivalent alternatives, taking the advantage of the *information matrix* equivalence. That is, in practice the variance-covariance matrix of the MLE can be calculated using any one of the following alternative estimators, and by evaluating these estimators at the maximum likelihood estimates.

Let $\tilde{\theta}$ be, as defined earlier, the maximum likelihood estimator of $\theta = (\beta', \sigma^2)$. Then, the variance-covariance matrix estimators of $\tilde{\theta}$ are defined as follows:

(i) The variance-covariance matrix estimator based on the *Hessian matrix* which is defined as

$$V_{Hes.}(\tilde{\theta}) = \left[- \left. \frac{\partial^2 \log L}{\partial\theta\partial\theta'} \right|_{\theta=\tilde{\theta}} \right]^{-1} \tag{7.18}$$

(ii) The variance-covariance matrix estimator based on the *information* matrix, given as

$$V_{Inf.}(\tilde{\theta}) = \left\{ E \left[- \left. \frac{\partial^2 \log L}{\partial\theta\partial\theta'} \right|_{\theta=\tilde{\theta}} \right] \right\}^{-1} \tag{7.19}$$

where the full expression of the above expectation is given by equation (2.22) in Chapter 2 of this study.

(iii) The third alternative variance-covariance matrix estimator is based on the matrix of *outer products* of the first derivatives of the log-likelihoods. This is defined as

$$V_{O.P.}(\tilde{\theta}) = \left[\sum_{i=1}^{N} \frac{\partial \log L_i}{\partial\theta} \left. \frac{\partial \log L_i}{\partial\theta'} \right|_{\theta=\tilde{\theta}} \right]^{-1} \tag{7.20}$$

where $\partial \log L_i / \partial\theta$ is, as defined before, the contribution of the i^{th} observation to the gradient vector.

These different covariance matrix estimators are usually associated with the computer algorithms employed for maximization. The estimators in (i) and (ii) may be used for Newton-type and method of scoring algorithms, respectively. While the use of the outer product matrix given in (iii) is closely associated with the method suggested by Berndt, Hall, Hall and Hausman (1974), which is usually referred to as the BHHH algorithm.

However, although convenient, this does not necessarily imply that one has to use the same procedure for both the estimation of parameters and the covariance matrix. In other words, it is possible to use one of the algorithms to estimate the parameters and another for the covariance matrix estimator.

Another alternative covariance matrix estimator which is not associated with any of the procedures discussed above is one which is based on the general variance-covariance matrix given by (7.16). This estimator is usually referred to as the *robust* (White-type) covariance matrix estimator and is defined by

$$V_{Rob.}(\tilde{\theta}) = V_{Hes.}(\tilde{\theta})\, V_{O.P.}(\tilde{\theta})^{-1}\, V_{Hes.}(\tilde{\theta}) \tag{7.21}$$

where $V_{Hes.}(\tilde{\theta})$ and $V_{O.P.}(\tilde{\theta})$ are as defined in (7.18) and (7.20), respectively.

This estimator can be viewed as analogous to the heteroscedastic consistent covariance matrix estimator of the regular regression model which was suggested by White (1980b). The estimator was further discussed by MacKinnon and White (1985) who proposed some finite sample corrections for the covariance matrix estimator in a linear regression context. White (1982, 1983) generalized the estimator further to accommodate more general models and misspecifications. Specifically, White (1982) showed that in the presence of misspecification the variance-covariance matrix estimator in (7.21) provides a robust covariance matrix estimator for the coefficients of the MLE estimator. Similarly, tests such as Wald and Lagrange multiplier can be robustified by the use of this covariance matrix estimator. He also showed that if the model is correctly specified the robust covariance matrix estimator, $V_{Rob.}(\theta)$, is asymptotically equivalent to the inverse of *Fisher's information* matrix.

Note that what is important here is that all the covariance matrix estimators should give approximately the same results for large samples, given that the model is correctly specified. Otherwise, according to White (1982), any significant differences between these covariance matrix estimators are usually considered as an indication of model misspecification.

This, however, does not seem to be true in finite samples. Studies which are related to the finite sample properties of these covariance matrix estimators indicate that even when the model is correctly specified the covariance matrix estimators may yield significantly different results [Griffiths, Hill and Pope (1987), Calzolari and Fiorentini (1990)].

In particular, Calzolari and Fiorentini (1990) studied the finite sample properties of the variance-covariance matrix estimators of the standard tobit model. They noted that the variances obtained based on both the *Hessian* and the *information* matrices yield almost identical results but can be substantially different from those of the other estimators. This conclusion is similar to that of Griffiths, Hill and Pope (1987) who investigated the properties of the variance-covariance estimators of the probit model. In general, these studies concluded that the choice of a particular variance-covariance estimator is not neutral in the estimation of variances of the coefficients of the model.

However, as stated at the outset of this Chapter, the interest of this Section focusses on the following two points.

> Unlike the regular regression model, the tobit model is likely to be more sensitive to misspecification. That is, the effects of incorrectly assuming normality of the error terms is likely to have some effect on the estimation of the variance-covariance matrix and this effect may vary from one distribution to the other.

It is therefore worth investigating the robustness of the alternative variance-covariance matrix estimators to the violations of the assumptions about the distributions of the error term of the model.

Moreover, an important aspect of the variance-covariance matrix estimators is their implication for statistical inference about the coefficients of the model. Thus, this Chapter further investigates whether the different variance-covariance matrix estimators lead to the same (or sufficiently close) conclusions for hypothesis testing and/or construction of confidence intervals for the coefficients of the model.

In general, the following Section investigates the finite sample properties of the alternative variance-covariance matrix estimators in the estimation of variances as well as for hypothesis testing and/or confidence intervals of the coefficients of the model under a variety of error distributions, sample sizes and degrees of censoring.

7.3.1 Monte Carlo Comparison of Variance Estimators

This Section presents a Monte Carlo comparison of the alternative variance-covariance matrix estimators discussed in the preceding section. The four alternative estimators discussed above are considered in this comparison. These include the variance-covariance matrix estimators based on

(i) The information matrix, $V_{Inf.}$.

(ii) The Hessian matrix, $V_{Hes.}$.

(iii) The outer product of the gradient vector, $V_{O.P.}$.

(iv) The robust (White-type) variance estimator, $V_{Rob.}$.

Further, since the interest is mainly in the variances of the coefficients, only the diagonal elements of the variance-covariance matrices are considered in the analysis. Thus, average variances are computed based on 3000 replications for each coefficient using the four different variance estimators. The estimated variances are then compared to see how close they estimate the true variances of the coefficients. Recall that the true (Monte Carlo) variances of the coefficients are computed using equation (5.13), in Chapter 5.

The summary statistics are given in Tables 7.2 to 7.4 for all sample sizes, degrees of censoring and distributional assumptions of the error term. For instance, Table 7.2 presents the true as well as the average estimated variances of the coefficients of the model for the three sample sizes and distributions, given that the degree of censoring is 25%. As can be seen from Table 7.2, Columns (1) and (2) present, respectively, the sample size and the error distribution of the model. The true (Monte Carlo) variances of the coefficients are listed in Column (3). The corresponding average estimated variances are given in Columns (4), (5), (6) and (7) which are obtained using the estimators, $V_{Inf.}$, $V_{Hes.}$, $V_{O.P.}$ and $V_{Rob.}$, respectively. Finally, the last two Columns (8) and (9) of the Table depict, respectively, the percentage (%) of times where the estimated variances obtained using $V_{O.P.}$ are greater than those of $V_{Inf.}$ and $V_{Hes.}$, respectively. These are computed to indicate whether there exists any systematic difference between the three variance estimators. Tables 7.3 and 7.4 present similar statistics for the medium (50%) and high (75%) degrees of censoring, respectively.

Given this, the following discussion concentrates on the main differences and/or similarities that may be observed within a particular variance estimator, compared to

Table 7.2: Comparison of Variance Estimators of the MLE Estimator for all Sample Sizes and Distributions, Given 25% Degree of Censoring.

Sample Size	Distrib- -ution		True Variance	Estimated Variances using				% $V_{O.P.}$ > than	
				$V_{Inf.}$	$V_{Hes.}$	$V_{O.P.}$	$V_{Rob.}$	$V_{Inf.}$	$V_{Hes.}$
(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)
100	Normal	β_0	0.0569	0.0554	0.0556	0.0594	0.0552	66.3	67.1
		β_1	0.0094	0.0091	0.0091	0.0099	0.0090	69.1	69.3
		β_2	0.0348	0.0326	0.0327	0.0355	0.0322	70.9	71.1
	Students'-t	β_0	0.0647	0.0503	0.0502	0.0719	0.0519	83.1	83.4
		β_1	0.0112	0.0083	0.0084	0.0138	0.0094	80.7	80.7
		β_2	0.0343	0.0309	0.0309	0.0524	0.0315	82.2	82.2
	Chi-square	β_0	0.0790	0.0707	0.0712	0.0762	0.0776	61.7	60.8
		β_1	0.0136	0.0116	0.0116	0.0128	0.0131	63.5	62.3
		β_2	0.0453	0.0450	0.0449	0.0530	0.0453	75.2	75.6
200	Normal	β_0	0.0282	0.0278	0.0278	0.0287	0.0276	62.6	62.8
		β_1	0.0048	0.0046	0.0046	0.0048	0.0046	67.3	67.6
		β_2	0.0174	0.0177	0.0177	0.0185	0.0176	67.2	67.8
	Students'-t	β_0	0.0278	0.0242	0.0242	0.0306	0.0259	82.5	83.0
		β_1	0.0048	0.0040	0.0040	0.0052	0.0046	77.4	77.0
		β_2	0.0164	0.0150	0.0150	0.0201	0.0158	78.7	78.9
	Chi-square	β_0	0.0367	0.0326	0.0330	0.0325	0.0367	46.4	41.2
		β_1	0.0061	0.0053	0.0054	0.0054	0.0061	51.8	49.1
		β_2	0.0214	0.0211	0.0211	0.0226	0.0218	66.0	65.7
400	Normal	β_0	0.0147	0.0139	0.0139	0.0141	0.0139	59.0	59.0
		β_1	0.0024	0.0023	0.0023	0.0024	0.0023	60.9	62.0
		β_2	0.0082	0.0085	0.0085	0.0087	0.0085	61.1	61.1
	Students'-t	β_0	0.0187	0.0124	0.0124	0.0148	0.0156	80.4	80.4
		β_1	0.0031	0.0020	0.0021	0.0025	0.0027	69.7	69.6
		β_2	0.0084	0.0071	0.0071	0.0089	0.0078	77.2	76.9
	Chi-square	β_0	0.0180	0.0166	0.0167	0.0161	0.0185	34.6	29.6
		β_1	0.0032	0.0027	0.0028	0.0027	0.0032	38.0	35.0
		β_2	0.0105	0.0103	0.0103	0.0107	0.0105	63.4	64.0

the true variances, as well as between the estimators as a result of changes in sample size, distribution of the error term and degree of censoring.

As can be seen from Table 7.2, one can make the following general comments on the different variance estimators. In almost all cases, the estimated variances based on the information matrix, $V_{Inf.}$ and the Hessian matrix, $V_{Hes.}$, provide identical results. This observation is similar to previous related studies [see Calzolari and Fiorentini (1990) and Griffiths, Hill and Pope (1987)]. But, what is more interesting about this result is that the estimators yield almost identical results under the non-normal distributions as well. Further, the variance estimators, $V_{Inf.}$ and $V_{Hes.}$, provide results which are lower than their corresponding true variances in almost all cases, except for the normal distribution where the difference between the true and the estimated variances is negligible (sometimes even identical). However, it is important to note that the gap between the true and the estimated variances gets bigger under the students'-t distribution. For example, as shown in Table 7.2, given a 25% degree of censoring and a sample size of 100, the variances of the coefficients obtained using $V_{Inf.}$ and $V_{Hes.}$ are about 10 to 35 percent lower than their corresponding true variances under the students'-t distribution. This is compared to less than 6 and 8 percent for the normal and chi-square distributions, respectively. Further, the difference between the true and the estimated variances using both $V_{Inf.}$ and $V_{Hes.}$ under the students'-t distribution increases to about 20 to 50 percent for the large sample size (400) while the differences between the true and the estimated variances for the other two distributions decline with increases in sample size.

This implies that the variance estimators $V_{Inf.}$ and $V_{Hes.}$ provide results which appear to be biased downwards and this bias becomes substantial under the students'-t distribution. The biases under the chi-square distribution are not as big as those under the students'-t distribution and decline as the sample size increases.

On the other hand, the variance estimator based on the outer product matrix, $V_{O.P.}$, provides average estimated variances which are larger than those of the $V_{Inf.}$ and $V_{Hes.}$ estimators in almost all cases. The $V_{Rob.}$ estimator, although it appears to provide intermediate results, does not follow the same pattern in all cases. Again, the difference between the variance estimators is quite small (negligible) under the normal distribution where all estimators yield results which are quite close (sometimes the same) to the true variances of the coefficients.

In general, the results indicate that the variances of the coefficients which are obtained using both $V_{Inf.}$ and $V_{Hes.}$ estimators are smaller than their respective true variances especially under the students'-t distribution. These results, perhaps, may be considered an indication that both $V_{Inf.}$ and $V_{Hes.}$ lead to underestimation of the true variances under the symmetric but fat tailed distribution. Whereas the variances obtained using the $V_{O.P.}$ and $V_{Rob.}$ do not seem to follow the same pattern.

One important question that could be raised in the comparison of the variance estimators is whether the four variance estimators yield the same (or sufficiently close) results under the correct specification of the model, i.e., when the errors are normally distributed. As shown in Table 7.2, the variances of the coefficients using both the $V_{Inf.}$, $V_{Hes.}$ and $V_{Rob.}$ are almost identical especially for the medium and large sample sizes. As indicated earlier, the former two yield the same results in almost all cases. The variances obtained using the $V_{O.P.}$, although slightly different, are also quite close to the others. Specifically, given a 25% degree of censoring and normally distributed error terms, the variances obtained from $V_{O.P.}$ are larger than those of $V_{Inf.}$ (and/or $V_{Hes.}$) by approximately 8 percent for the small sample size, and less than 4 percent for the medium sample size. These differences decline further for the large sample size. Therefore, these results indicate that, under the normality of the error terms, the variance estimator, $V_{O.P.}$, may yield results which are quite close to those of others, provided that the degree of censoring is low. Thus, given that the errors are normally distributed, the four variance estimators yield results which are quite close (sometimes the same) to each other as well as to the true variances of the coefficients of the model. The implication of this is, as will be discussed later in this Chapter, that hypothesis tests and confidence intervals for the coefficients of the model will lead to the same conclusion, irrespective of the variance-covariance matrix estimators. In other words, the choice of the variance estimator appears to be neutral, given that the errors are normally distributed.

Further, the situation appears to be similar for the 50% degrees of censoring and normally distributed error terms, except for the small sample. Table 7.3 depicts results for the variance estimators for all sample sizes and distributions, given that the degree of censoring is 50%. The estimated variances under the normal distribution depict that the variances of the coefficients obtained using $V_{O.P.}$ are about 12 percent larger than their respective variances estimated using $V_{Inf.}$ for the small sample size. This percentage drops to about 3 percent when the sample becomes large. It is also evident that the average variances from $V_{Inf.}$ and $V_{Hes.}$ remain very close to the true variances. However, the difference between the $V_{Inf.}$ and $V_{O.P.}$ can be as big as about 50 percent for the small sample size, and about 20 percent for the large sample size,

Table 7.3: Comparison of Variance Estimators of the MLE Estimator for all Sample Sizes and Distributions, Given 50% Degree of Censoring.

Sample Size	Distrib- -ution		True Variance	Estimated Variances using				% $V_{O.P.}$ > than	
				$V_{Inf.}$	$V_{Hes.}$	$V_{O.P.}$	$V_{Rob.}$	$V_{Inf.}$	$V_{Hes.}$
(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)
100	Normal	β_0	0.1239	0.1223	0.1230	0.1398	0.1187	70.9	73.6
		β_1	0.0157	0.0156	0.0157	0.0181	0.0151	72.5	74.7
		β_2	0.0449	0.0421	0.0421	0.0472	0.0415	71.7	72.7
	Students'-t	β_0	0.2968	0.1336	0.1345	0.1809	0.2009	72.4	73.4
		β_1	0.0327	0.0161	0.0162	0.0252	0.0235	73.9	74.4
		β_2	0.0545	0.0434	0.0434	0.0679	0.0487	79.8	80.9
	Chi-square	β_0	0.2237	0.1789	0.1807	0.1897	0.2116	53.5	53.2
		β_1	0.0282	0.0222	0.0224	0.0245	0.0265	57.1	56.4
		β_2	0.0644	0.0548	0.0550	0.0627	0.0594	68.3	68.8
200	Normal	β_0	0.0622	0.0618	0.0620	0.0662	0.0609	66.0	69.0
		β_1	0.0079	0.0080	0.0081	0.0087	0.0079	67.1	69.1
		β_2	0.0266	0.0252	0.0252	0.0269	0.0251	66.5	67.1
	Students'-t	β_0	0.1553	0.0683	0.0688	0.0783	0.1137	59.5	58.8
		β_1	0.0176	0.0086	0.0088	0.0109	0.0137	60.7	59.8
		β_2	0.0316	0.0231	0.0229	0.0322	0.0277	75.5	76.5
	Chi-square	β_0	0.1143	0.0902	0.0907	0.0859	0.1097	37.1	34.1
		β_1	0.0147	0.0115	0.0116	0.0116	0.0139	47.1	45.3
		β_2	0.0414	0.0373	0.0374	0.0403	0.0395	64.1	64.5
400	Normal	β_0	0.0308	0.0307	0.0308	0.0318	0.0305	62.2	64.2
		β_1	0.0040	0.0039	0.0039	0.0041	0.0039	63.5	65.2
		β_2	0.0112	0.0112	0.0112	0.0116	0.0112	64.0	64.6
	Students'-t	β_0	0.1318	0.0345	0.0345	0.0361	0.0854	46.6	44.8
		β_1	0.0135	0.0043	0.0043	0.0050	0.0094	49.6	47.7
		β_2	0.0163	0.0121	0.0120	0.0157	0.0146	72.6	73.8
	Chi-square	β_0	0.0547	0.0437	0.0439	0.0390	0.0543	19.0	14.9
		β_1	0.0070	0.0056	0.0056	0.0052	0.0069	29.9	26.0
		β_2	0.0167	0.0166	0.0165	0.0169	0.0176	51.8	53.0

if the degree of censoring is high. This can be seen from Table 7.4 below.

In general, as would be expected asymptotically, all three variance estimators may yield the same (or sufficiently close) results under the normal distribution, given that the degree of censoring is low. As the degree of censoring increases the variances estimated using $V_{O.P.}$ appear to be larger than their corresponding true variances especially for the small and medium sample sizes. It is also evident that, as the degree of censoring increases and if the sample size is small, the variances obtained using the $V_{Rob.}$ tend to underestimate their corresponding true variances (see Tables 7.3-7.4). Chesher and Jewitt (1987) made similar observations for the robust (White-type) estimator in a linear regression framework.

However, unlike the results from the normal distribution, the four variance estimators appear to be different under the non-normal distributions. In particular, the difference between the true and the estimated variances becomes quite substantial under the students'-t distribution and varies from one variance estimator to another. For instance, given a degree of censoring of 25% and a medium (200) sample size, the variances of β_1 and β_2 using $V_{O.P.}$ are given by 0.0052 and 0.0201, respectively, under the students'-t distribution. These results are about 8 to 22 percent larger than their respective true variances and over 30 percent larger than the average variances estimated by using the $V_{Inf.}$ or $V_{Hes.}$. The relative difference between the estimated and true variances as well as among the variance estimators under the students'-t distribution increases dramatically for higher degrees of censoring (see Table 7.4). Note that the large differences between the variances under the students'-t distribution is aggravated due to the fact that the variances obtained using both $V_{Inf.}$ and $V_{Hes.}$, as discussed earlier, are below their corresponding true variances and hence much more

Table 7.4: Comparison of Variance Estimators of the MLE Estimator for all Sample Sizes and Distributions, given 75% Degree of Censoring.

Sample Size	Distrib- -ution		True Variance	Estimated Variances using				% $V_{O.P.}$ > than	
				$V_{Inf.}$	$V_{Hes.}$	$V_{O.P.}$	$V_{Rob.}$	$V_{Inf.}$	$V_{Hes.}$
(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)
100	Normal	β_0	0.5542	0.5191	0.5294	0.7640	0.4643	83.0	85.4
		β_1	0.0488	0.0458	0.0467	0.0672	0.0406	82.0	84.4
		β_2	0.0911	0.0826	0.0834	0.1066	0.0779	80.4	82.7
	Students'-t	β_0	2.1815	0.7486	0.7569	1.4313	1.0922	76.9	78.5
		β_1	0.1492	0.0627	0.0645	0.1437	0.0817	78.6	80.5
		β_2	0.2225	0.1292	0.1301	0.2728	0.1522	84.5	85.9
	Chi-square	β_0	0.9886	0.7407	0.7514	1.0226	0.7927	70.6	72.6
		β_1	0.0876	0.0654	0.0664	0.0968	0.0675	73.1	74.9
		β_2	0.1653	0.1460	0.1478	0.2045	0.1449	78.2	79.8
200	Normal	β_0	0.2735	0.2599	0.2625	0.3186	0.2445	74.4	77.0
		β_1	0.0244	0.0232	0.0234	0.0287	0.0218	74.4	76.3
		β_2	0.0516	0.0491	0.0494	0.0559	0.0478	72.2	74.4
	Students'-t	β_0	1.7907	0.4164	0.4107	0.5638	0.9202	59.9	60.7
		β_1	0.1211	0.0356	0.0358	0.0565	0.0677	62.9	63.3
		β_2	0.1429	0.0767	0.0765	0.1272	0.1002	75.8	77.7
	Chi-square	β_0	0.4411	0.3564	0.3566	0.3714	0.4258	51.7	51.5
		β_1	0.0388	0.0323	0.0323	0.0361	0.0366	59.1	59.0
		β_2	0.0837	0.0758	0.0760	0.0876	0.0793	67.3	69.1
400	Normal	β_0	0.1289	0.1289	0.1293	0.1426	0.1254	67.9	70.3
		β_1	0.0115	0.0116	0.0117	0.0129	0.0113	67.4	69.4
		β_2	0.0229	0.0230	0.0230	0.0246	0.0226	64.8	67.4
	Students'-t	β_0	1.9360	0.2447	0.2382	0.2601	0.8431	39.7	38.9
		β_1	0.1110	0.0207	0.0208	0.0273	0.0531	46.7	46.2
		β_2	0.1225	0.0445	0.0441	0.0722	0.0707	70.0	71.4
	Chi-square	β_0	0.2154	0.1715	0.1708	0.1564	0.2160	32.5	30.5
		β_1	0.0189	0.0156	0.0156	0.0151	0.0186	40.7	39.9
		β_2	0.0426	0.0392	0.0392	0.0412	0.0420	57.0	58.1

lower than those of $V_{O.P.}$. It is also clear that the quality of the various estimators generally declines for the high degree of censoring coupled with the students'-t distribution for all sample sizes. It is not, however, clear why the relative performance of the various estimators appeared to be inferior under the students'-t distribution compared to the skewed distribution.

On the other hand, the gap between the true and the estimated variances from the four variance estimators under the chi-square distribution does not seem to be as high as that of the students'-t distribution. In fact, it is interesting to note that the four variance estimators provide results which are very close (sometimes even the same) to the true variances for the medium and large samples, given that the degree of censoring is not high (see Tables 7.2 and 7.3). If the degree of censoring is high (75%), the difference between the true and the estimated variances under the chi-square distribution starts to increase but still remains much lower than the relative differences observed under the students'-t distribution (see Table 7.4).

Further, it is also important to note that, not only does $V_{O.P.}$ provide variances which are larger that those of $V_{Inf.}$ and $V_{Hes.}$, the difference between the estimators appears to be systematic. As shown in Columns (8) and (9) of Table 7.2, $V_{O.P.}$ provides larger variances than $V_{Inf.}$ and $V_{Hes.}$ in most cases for all sample sizes and distributions, even when the model is correctly specified. For example, given a sample size of 100 and a 25% degree of censoring, the variances obtained using $V_{O.P.}$ are about 70 percent of the time larger than the variances obtained using $V_{Inf.}$ or $V_{Hes.}$ under the normal distribution. Similarly, $V_{O.P}$ provides larger variances than both $V_{Inf.}$ and $V_{Hes.}$ about 80 percent of the time under the students'-t distribution. Similar observations can be made for the medium and large samples as well as for higher degrees of censoring (see tables 7.3-7.4). However, there does not appear to be a similar relationship between the estimators under the chi-square distribution except for the small sample size.

Finally, given the discussions above, the following points can be concluded regarding the variance-covariance matrix estimators of the MLE of the tobit model.

> In general, the variance estimators based on the information matrix, $V_{Inf.}$, and the Hessian matrix, $V_{Hes.}$, provide identical results in almost all cases. Further, the variance estimator based on the outer product matrix, $V_{O.P.}$, yields results which are larger than those of $V_{Inf.}$ and $V_{Hes.}$ in almost all cases and these differences appear to be systematic for all but the chi-square distribution.

> Under normally distributed error terms, the three estimators, namely, $V_{Inf.}$, $V_{Hes.}$ and $V_{Rob.}$ provide average variances which are quite close (if not identical) to the true variances. Further, $V_{O.P.}$ provides similar results, provided that the degree of censoring is not high. In general, as would be expected, there appears to be little difference between the true and estimated variances for all four estimators under the normal distribution.

> Whereas, under non-normality of the error terms, the performance of the variance estimators varies depending on the sample size, degree of censoring and type of distribution. In particular, the two variance estimators, $V_{Inf.}$ and $V_{Hes.}$ yield variances which are relatively lower than their corresponding true variances under the students'-t distribution. This can be considered as an indication that the two estimators understate the true variances of the coefficients. On the other hand, the evidence shows that the robust $V_{Rob.}$ estimator appears

to be relatively better under the non-normal distributions especially under the skewed distribution. $V_{O.P.}$ appears to be the second best under the non-normal distributions in general.

Finally, the most important point here is that the choice of a particular variance-covariance matrix estimator appears to be neutral under the normal distribution except if the sample size is small and a high degree of censoring. However, the relative performance of the variance estimators is likely to be substantially different if one assumes normality of the error terms when in fact they are not normal. That is, when the model is not correctly specified. This is particularly clear from the results obtained under the students'-t distribution.

However, our results are not conclusive as to whether one variance estimator is superior (or relatively robust) than the other in terms of their reliability for hypothesis testing and/or confidence intervals for the coefficients of the model. In other words, although hypothesis tests and/or confidence intervals of the coefficients are likely to be the same or close under the normal distribution irrespective of the variance estimators, this may not be the case under the non-normal distributions. Thus, the implications of the different variance estimators for hypothesis testing and/or construction of confidence intervals needs to be investigated. This is discussed thoroughly in the following Section.

7.3.2 Implications of the Variance Estimators for Hypothesis Testing

In the preceding section we compared the alternative variance-covariance matrix estimators in the maximum likelihood framework of the tobit model. These estimators included, the variance-covariance matrix estimator based on the information matrix, $V_{Inf.}$, the variance-covariance matrix estimator based on the Hessian matrix, $V_{Hes.}$, the variance-covariance matrix estimator based on the outer product of the gradient vector, $V_{O.P.}$, and the robust (White-type), $V_{Rob.}$, variance-covariance matrix estimator.

The main conclusions, among others, included that there exists very little variation among the four variance estimators, provided that the errors are normally distributed and the degree of censoring is not high. However, under non-normality of the error terms, the variances obtained using the different estimators can be substantially different from the true variances as well as from each other. In other words, the choice of a particular variance-covariance estimator does not appear to be neutral in the maximum likelihood framework of the tobit model if the model is misspecified.

However, just comparing the average variances, while useful in indicating the differences between the variance estimators, does not necessarily imply the superiority of one estimator over the other in terms of their reliability for hypothesis testing of the coefficients of the model. It is quite possible that, unless the differences are substantial, different variance estimators may lead to the same conclusion if used for testing hypothesis about the parameters of the model. This is most likely to be the case under the normal distribution where the average variances of the coefficients based on the alternative variance estimators appeared to be closer to the true variances in most cases.

Thus, the performance of the estimators depends on their reliability in hypothesis testing and/or confidence intervals of the parameters of the model, which is the main purpose of the discussion below.

Specifically, we test the hypothesis:

$$\begin{aligned} H_0 : \beta_k &= 1 \\ H_1 : \beta_k &\neq 1, \quad k = 1, 2. \end{aligned} \tag{7.22}$$

To test the hypothesis we use the test statistic given by

$$t = \frac{\tilde{\beta}_k - 1}{\sqrt{Var(\tilde{\beta}_k)}} \tag{7.23}$$

where $\tilde{\beta}_k$ is the sample estimate of β_k and $Var(\tilde{\beta}_k)$ is the variance of $\tilde{\beta}_k$ which is obtained using the four alternative variance-covariance matrix estimators discussed above.

Under the null hypothesis, the statistic t is asymptotically distributed as a normal random variable. 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%.

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

$$P[\tilde{\beta}_k - z \times s.e(\tilde{\beta}_k) < \beta_k < \tilde{\beta}_k + z \times s.e.(\tilde{\beta}_k)] = 0.95 \tag{7.24}$$

which is equivalent to

$$P(-1.96 < t < 1.96) = 0.95 \tag{7.25}$$

where t is defined by (7.23) and a standard z value at a 5% significance level is 1.96.

Given this, we obtained the percent of coefficients contained in the 95% confidence intervals using the four variance-covariance matrix estimators in the estimation of the variances of the coefficients. The summary statistics for the 95% confidence intervals of β_1 and β_2, based on the three variance estimators, are given in Tables 7.5 to 7.7 for all sample sizes, degrees of censoring and distributional assumptions of the error term. The following discussion concentrates on whether these results are the same or close enough to the desired (expected) level, i.e. 95%.

Table 7.5 depicts results for the 95% confidence intervals using the four variance estimators, given a 25% degree of censoring. These results show that the four variance estimators perform quite well even for the small sample size, provided that the errors are normally distributed. For example, given a sample size of 100 and normally distributed error terms, the confidence intervals for β_1 and β_2 are, respectively, 94.43% and 93.93% for $V_{Inf.}$, and 94.50% and 94.03% for $V_{Hes.}$. The corresponding values of β_1 and β_2, respectively, are 95.10% and 95.07% for $V_{O.P.}$, and 94.20% and 93.57% for $V_{Rob.}$. These results are quite close, especially for the $V_{O.P.}$, to the 95% closure rate. The results improve further for the medium and large samples under the normal distribution. This is also in line with variance comparisons discussed in the preceding section.

However, as can be seen from Table 7.5, the four variance estimators may yield slightly different results for the confidence intervals under non-normal error terms. For instance, given a small sample size and students'-t distribution, 92.57 and 93.50 percent of the confidence intervals contain β_1 and β_2, respectively, using $V_{Inf.}$ as compared to 94.97 and 96.70 percent for $V_{O.P.}$. The values for the $V_{Rob.}$ using the same sample size and distribution are 94.63% and 93.90%, respectively, for β_1 and

Table 7.5: 95% Confidence Intervals using the four Variance Estimators of the MLE for all Sample Sizes and Distributions, Given 25% Degree of Censoring.

Sample Size	Distrib- -ution		95% Confidence Intervals using $V_{Inf.}$	$V_{Hes.}$	$V_{O.P.}$	$V_{Rob.}$
(1)	(2)		(3)	(4)	(5)	(6)
100	Normal	β_1	94.43	94.50	95.10	94.20
		β_2	93.93	94.03	95.07	93.57
	Students'-t	β_1	92.57	92.70	94.97	94.63
		β_2	93.50	93.70	96.27	93.90
	Chi-square	β_1	91.80	91.96	92.37	93.80
		β_2	93.93	93.93	95.03	94.27
200	Normal	β_1	94.17	94.20	94.77	93.96
		β_2	95.43	95.40	95.70	95.23
	Students'-t	β_1	93.47	93.47	95.13	95.43
		β_2	94.30	94.20	95.83	94.73
	Chi-square	β_1	89.50	89.73	89.17	92.43
		β_2	94.80	94.80	95.03	95.07
400	Normal	β_1	94.43	94.43	94.67	94.17
		β_2	95.87	95.77	95.73	95.70
	Students'-t	β_1	90.37	90.37	92.23	93.90
		β_2	92.90	92.90	94.60	94.10
	Chi-square	β_1	87.00	87.23	86.07	89.73
		β_2	94.10	94.20	94.43	94.03

β_2. These results, although still not that far from the desired level, indicate that the $V_{Inf.}$ tends to be slightly lower than the desired 95% confidence level. More notably, the confidence intervals of the coefficients obtained using $V_{Inf.}$ do not seem to improve when the sample size increases. For example, given a large sample size and t-distributed errors, the confidence intervals for β_1 and β_2 using $V_{Inf.}$ are given by 90.37 and 92.90, respectively, which are even lower than their corresponding values for the small sample size. These results reflect underestimation of the true variances of the coefficients as discussed in the preceding section. On the other hand, it is interesting to note that the results for $V_{Rob.}$ under the students'-t distribution appeared to be as good as under the normal distribution. The results for the $V_{O.P.}$ are also close to those of the $V_{Rob.}$ for the students'-t distribution (see Table 7.5 for N=200). These results indicate that, relative to $V_{Inf.}$ and $V_{Hes.}$, $V_{Rob.}$ appears to be robust for the symmetric but fat tailed distribution.

Further, the confidence intervals of the coefficients under the chi-square distribution appear to be better under $V_{Rob.}$ and $V_{O.P.}$ as compared to those of $V_{Inf.}$ and $V_{Hes.}$, for all sample sizes. In general, the results for the $V_{Inf.}$ and $V_{Hes.}$, although as good as the those of the other two estimators under the normal distribution, are relatively inferior under the non-normality of the error terms.

The difference between the variance estimators for hypothesis testing of the coefficients, similar to that of the variance comparisons, gets more visible when the degree of censoring increases. Table 7.6 presents results for the 95% confidence intervals for all sample sizes and distributions, given a 50% degree of censoring. As can be seen from the table, the results for $V_{Inf.}$ and $V_{Hes.}$ under the non-normal distributions appear to be well below the 95% closure rate. For example, as shown in

Table 7.6: 95% Confidence Intervals using the four Variance Estimators of the MLE for all Sample Sizes and Distributions, Given 50% Degree of Censoring.

Sample Size	Distrib- -ution		95% Confidence Intervals using $V_{Inf.}$	$V_{Hes.}$	$V_{O.P.}$	$V_{Rob.}$
(1)	(2)		(3)	(4)	(5)	(6)
100	Normal	β_1	95.13	95.00	96.00	93.83
		β_2	94.10	94.20	95.03	93.53
	Students'-t	β_1	91.60	91.60	93.53	94.13
		β_2	93.80	93.77	96.13	94.53
	Chi-square	β_1	92.73	92.80	92.90	94.23
		β_2	92.50	92.47	93.76	93.27
200	Normal	β_1	94.80	94.90	95.57	93.83
		β_2	94.03	94.07	94.63	94.00
	Students'-t	β_1	87.70	87.77	89.30	93.73
		β_2	93.50	93.47	95.57	94.97
	Chi-square	β_1	91.47	91.40	90.43	94.27
		β_2	93.93	93.93	94.53	94.60
400	Normal	β_1	94.93	95.13	95.30	94.50
		β_2	94.47	94.40	94.57	94.23
	Students'-t	β_1	82.77	83.00	82.43	93.60
		β_2	92.73	92.60	94.33	94.57
	Chi-square	β_1	88.97	89.13	87.13	93.13
		β_2	94.50	94.47	94.40	95.00

Table 7.6, given a large sample size (400) and the students'-t distribution, only 82.77 and 92.73 percent of the confidence intervals contain the true parameters, β_1 and β_2, respectively, when using $V_{Inf.}$. The respective values using $V_{Rob.}$ are given by 93.60 and 94.57 percent. These values, especially those of β_1, indicate that the hypothesis tests based on both $V_{Inf.}$ and $V_{Hes.}$ can be sometimes misleading if the assumption of normality of the error term is not correct. Similar observations can be made for the chi-square distribution where $V_{Rob.}$ appears to do well compared to others.

The results for the high degree of censoring (i.e., 75%) are summarized in Table 7.7 for all sample sizes and distributions. As can be seen from Table 7.7, all the estimators perform fairly well under the normal distribution, except for $V_{Rob.}$ where the confidence intervals appear to understate the desired 95% level for the small and medium sample sizes. Further, it is evident that the results obtained based on $V_{Inf.}$ can be relatively inferior to those of $V_{O.P.}$ and $V_{Rob.}$ under the non-normal distributions. It is also clear that because the quality of the estimates generally declines for all estimators for the high degree of censoring the distinction between the estimators becomes less obvious.

Note that, as stated earlier in this Chapter, the main interest is to see whether one or more of the estimators are more robust than others for hypothesis testing. Thus, given the discussions above and in an attempt to establish a relative ranking on the robustness of the estimators, the results in Tables 7.5-7.7 are evaluated further as follows.

The values of the 95% confidence intervals can be considered as outcomes from a Bernoulli experiment with N=3000 and with a probability of success equal to $p = 0.95$. Now, given a significance level, say $\alpha = 0.05$, one can construct a confidence interval

Table 7.7: 95% Confidence Intervals using the four Variance Estimators of the MLE for all Sample Sizes and Distributions, Given 75% Degree of Censoring.

Sample Size	Distrib- -ution		95% Confidence Intervals using			
			$V_{Inf.}$	$V_{Hes.}$	$V_{O.P.}$	$V_{Rob.}$
(1)	(2)		(3)	(4)	(5)	(6)
100	Normal	β_1	93.47	93.87	96.17	91.33
		β_2	93.30	93.70	95.87	91.83
	Students'-t	β_1	88.10	88.56	92.80	89.37
		β_2	92.70	92.77	95.63	92.63
	Chi-square	β_1	92.10	92.13	94.03	90.10
		β_2	94.40	94.37	95.57	93.50
200	Normal	β_1	94.63	94.87	96.03	93.00
		β_2	94.33	94.20	95.33	93.60
	Students'-t	β_1	86.17	86.47	88.30	91.43
		β_2	91.33	91.23	94.10	93.27
	Chi-square	β_1	93.30	93.40	93.00	94.50
		β_2	93.57	93.40	94.57	94.07
400	Normal	β_1	95.23	95.27	95.43	94.33
		β_2	94.87	94.93	95.33	94.93
	Students'-t	β_1	82.20	82.30	80.50	92.77
		β_2	87.50	87.47	90.03	93.53
	Chi-square	β_1	92.67	92.77	91.23	94.47
		β_2	93.10	93.13	93.07	94.43

for the 'true' proportion of successful interval estimates as

$$P(\hat{p} - 1.96 \times s.e.(p) < p < \hat{p} + 1.96 \times s.e.(p)) = 0.95 \tag{7.26}$$

where the standard error of p, $s.e.(p)$, is given by

$$s.e.(p) = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{N}} = \sqrt{\frac{0.95 \times 0.05}{3000}} = 0.004$$

Thus, the 95% confidence interval for the true proportion p is

$$(94.22 < p < 95.78) \tag{7.27}$$

Thus, the reliability or robustness of the alternative variance estimators for hypothesis testing can be examined by comparing whether or not the values for the confidence intervals reported in Tables 7.5 to 7.7 lie within the desired (expected) limits given by (7.27).

Following this procedure, Table 7.8 below summarizes the number and percentage of times that the estimated values of the confidence intervals lie within the interval (7.27) for all sample sizes and degrees of censoring. These results are compared accross the three distributions. Note that, as can be seen from Tables 7.5-7.7, each variance estimator is examined using a total number of 27 different setups in the experiment (i.e, 3 distributions $\times$ 3 sample sizes $\times$ 3 degrees of censoring $=$ 27). So, for each distribution we have 9 different setups (i.e., 3 sample sizes $\times$ 3 degrees of censoring $=$ 9 setups).

For example, as shown in Table 7.8, Column (1) lists the four variance estimators. The first row of the table presents the three distributions and underneath are the corresponding values. For instance, using $V_{Inf.}$ and given a normal distribution, the number (No.) of times the proportion of successful interval estimates was not

Table 7.8: Ranking on the Robustness of the Estimators.

Estimator		Normal		Students'-t		Chi-square		Total	
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$V_{Inf.}$	No.	7	5	0	1	0	3	7	9
	%	78	56	0	11	0	33	26	33
$V_{Hes.}$	No.	7	4	0	0	0	3	7	7
	%	78	44	0	0	0	33	26	26
$V_{O.P.}$	No.	6	8	2	4	0	7	8	19
	%	67	89	22	44	0	78	30	70
$V_{Rob.}$	No.	2	4	2	4	4	7	8	15
	%	22	44	22	44	44	78	30	56

significantly different from 0.95 are 7 and 5 for β_1 and β_2, respectively. The respective percentages, each of which are calculated out of 9 different combinations (setups), are given by 78 and 56 percent. Results for the students'-t and the chi-square distributions are presented in Columns (4) to (7) and can be interpreted in a similar way. The interpretation of the last Columns (8) and (9) are also similar, except the total number (No.) is obtained by adding the corresponding values of β_1 and β_2 over the three distributions and the percentage (%) of these numbers are calculated out of the total number of 27 setups (instead of 9 setups).

Given this, Table 7.8 reveals that $V_{Inf.}$ and $V_{Hes.}$, while performing quite well under the normal distribution, are generally poor under the non-normal distributions. This can be seen from the low overall performance ranging between 26 to 33 percent (see

Columns (8) and (9) of Table 7.8). Further, $V_{O.P.}$ appears to be as good as $V_{Rob.}$ under the students'-t distribution while the later is relatively superior under the chi-square distribution. The overall performance of the estimators indicates that $V_{O.P.}$ appears to be the best providing the desired level of confidence intervals up to 70 percent of the time. This is followed by the $V_{Rob.}$ with an overall performance of about 30 to 60 percent for all distributions.

However, the results in Table 7.8 do not appear to reflect the real picture regarding the robustness of the estimators that we discussed earlier based on the results provided in Tables 7.5-7.7. The main reason for this is that the interval in (7.27) is relatively narrow (more precise). That is, although most of the values of the confidence intervals are quite close to the desired level, they are not close enough to lie within the desired limits given by (7.27) making the comparison less realistic as far as the robustness of the estimators is concerned. Thus, in order to obtain a more realistic comparison between the various estimators one needs to consider a relatively wider confidence interval which is discussed below.

Suppose, we are prepared to accept a margin of error, say, plus or minus 2 of the desired level so that the confidence limits will be between 93.00 to 97.00 percent. In other words, an estimator is considered robust in relative terms if it provides values of the confidence intervals between 93.00 to 97.00. Given this, we compared the robustness of the various estimators and the summary statistics is given in Table 7.9. Note that the interpretation of Table 7.9 is the same as before except this time we have used a relatively wider interval (i.e., $93.00 < p < 97.00$).

As can be seen from Table 7.9, all estimators perform quite well under the correct specification of the model. In particular, $V_{Inf.}$, $V_{Hes.}$ and $V_{O.P.}$ provide the desired

Table 7.9: Further Ranking of Robustness of the Estimators.

Estimator		Normal		Students'-t		Chi-square		Total	
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$V_{Inf.}$	No.	9	9	1	4	1	8	11	21
	%	100	100	11	44	11	89	41	78
$V_{Hes.}$	No.	9	9	1	4	1	8	11	21
	%	100	100	11	44	11	89	41	78
$V_{O.P.}$	No.	9	9	3	8	2	9	14	26
	%	100	100	33	89	22	100	52	96
$V_{Rob.}$	No.	8	8	6	8	6	9	20	25
	%	89	89	67	89	67	100	74	93

results in all cases, provided that the errors are normally distributed. This is quite in agreement with the results discussed above both in terms of variance estimation and hypothesis testing. However, $V_{Rob.}$ and $V_{O.P.}$ appear to perform better under the non-normal distributions. Specifically, the results for V_{Rob} under the non-normal distributions appear to be as good as those of the normal distribution. Further, the overall results indicated by Columns (8) and (9) of Table 7.9 depict that $V_{Rob.}$ is relatively robust and provides the desired levels for hypothesis testing in most cases, i.e., about 75 to 95 percent of the time. However, as stated earlier, the $V_{Rob.}$ estimator tends to underestimate the desired level (i.e., both in terms of variance and hypothesis testing) under the normal distribution. The $V_{O.P}$ appears to the next best with an overall performance ranging between 50 to 95.

In general, the following points can be concluded with respect to the implications of the three variance-covariance matrix estimators for hypothesis testing and/or confidence intervals of the coefficients in the ML framework of the tobit model.

Under normality of the error terms of the model, the four variance estimators yield approximately the same (close) results in hypothesis testing leading to the same conclusions about the parameters of the model. Note that, as discussed in the preceding section, although there appears to be minor differences between the true and average estimated variances in some cases, these differences do not, however, imply the superiority of one estimator over the other in terms of hypothesis testing, provided that the errors are normally distributed. Thus, as would be expected asymptotically [see White (1982, 1983)], the choice of any particular estimator appears to be neutral under the correct specification of the model. Note that, if the degree of censoring is high, a sample size of at least 200 may be required to obtain relatively reliable results under the normal distribution.

Under non-normality of the errors, it is evident that the variance estimators based on the information and Hessian matrices (i.e., $V_{Inf.}$ and $V_{Hes.}$) appear to be generally inferior as compared to the other two estimators for all sample sizes and degrees of censoring. Specifically, $V_{Inf.}$ and $V_{Hes.}$ provide confidence intervals which are relatively narrower than they should be, and hence the probability of rejecting a true hypothesis can be substantial as the degree of censoring increases. In other words, $V_{Inf.}$ and $V_{Hes.}$ are relatively sensitive to violations of the normality assumption of the model. Note that $V_{Inf.}$ is widely used in applied research in the estimation of the variance-covariance matrix of tobit models.

Interestingly, the hypothesis testing results obtained using $V_{Rob.}$ appear to be relatively better than others under non-normality of the error terms. In particular, the performance of the $V_{Rob.}$ estimator under the chi-square distribution is as good as under the normal distribution. The results for $V_{O.P.}$ under the chi-square distribution are relatively closer to those of $V_{Rob.}$. These results are consistent with the variance comparisons discussed in the preceding Section.

In general, $V_{Rob.}$ appears to be relatively robust to the violations of normality of the error terms followed by $V_{O.P.}$. These results may imply that, as argued by White (1982), any significance difference between the three variance estimators may be considered an indication of misspecification of the model. It is, perhaps, important to note that this conclusion may be considered relatively strong as there are not huge differences between the estimators. On the other hand, the three variance estimators may lead to the same conclusions in hypothesis testing if the model is correctly specified.

Finally, recall that in Chapter 6, in terms of hypothesis testing, we concluded that the MLE did not perform well under the non-normal distributions relative to its performance under the normal distribution. Thus, as can be seen in the present Chapter, the result would improve slightly by using the robust variance-covariance matrix estimator, $V_{Rob.}$ (or $V_{O.P.}$).

7.4 Summary and Conclusions

The method of maximum likelihood estimation is popular because of its desirable properties such as consistency and asymptotic normality, under certain regularity conditions. However, some studies have indicated that, unlike the regular regression model, the MLE of the censored regression (tobit) model may have undesirable consequences such as inconsistency if the assumption of normality of the errors is violated. In other words, the MLE of the tobit model is sensitive to misspecification of the model. This, however, does not seem to be generally true.

In this Chapter, we examined the consistency of the MLE under a variety of distributions for the error term, namely, the normal, the students'-t and the chi-square distribution. The effects of each distribution on the consistency of the MLE is investigated based on the asymptotic properties of the MLE. The main conclusions, among others, are that there seems to be very little (or no) inconsistency in β_1 and β_2 for both the students'-t and chi-square distributions. On the other hand, the constant term (β_0) of the model can be inconsistent under the skewed distribution even when the sample size becomes large. The evidence also shows that, if the degree of censoring is high, the inconsistency in β_0 can be substantial under the students'-t distribution. Note that, as discussed earlier in this Chapter, the constant term can be important to obtain results related to the tobit model depending on the particular objectives of the research.

However, assuming that the constant term is not very important, inconsistency does not appear to be a serious problem for the MLE estimator if we assume normality when in fact the errors are not (i.e., for the students'-t and chi-square distributions).

These results are also consistent with the results discussed in Chapter 6 in which it was concluded that the MLE performs quite well under the non-normal distributions in terms of bias unless the sample is very small (100) and the degree of censoring is high (75%). Clearly, the results under the normal distribution are as anticipated. That is, no inconsistency is observed for all sample sizes and degrees of censoring, provided that the model is correctly specified.

Note that when estimates are obtained using the MLE estimator, the variance-covariance matrix for the coefficients can be obtained using a number of alternative covariance matrix estimators. These estimators are based on the following (i) the inverse of the information matrix, (ii) the inverse of the Hessian matrix (iii) the inverse of the outer product of the gradient vector, and (iv) the robust (White-type) covariance matrix estimator. Each one of the first three of these estimators is usually associated with a particular algorithm employed in the estimation procedure. These alternative estimators, although asymptotically equivalent under the correct specification of the model, are not generally the same in finite samples, and are more likely to be different if the assumption of normality does not hold in practice.

Therefore the performance of these variance-covariance matrix estimators in the estimation of the variances of coefficients of the model is investigated under the different distributions, sample sizes and degrees of censoring. The results indicate that, under normality of the error terms, the variances obtained using the information matrix, Hessian matrix and the robust (White-type) estimators yield almost identical results in almost all cases. Furthermore, given that the degree of censoring is low, all four variance estimators may provide results which are quite close under the normal distribution. It is also evident that $V_{Inf.}$ and $V_{Hes.}$ yield almost identical results in all cases.

On the other hand, the alternative variance estimators may yield substantially different variances of the coefficients, if the assumption of normality of the error term is violated, that is, under the students'-t and chi-square distributions. Our results also reveal that the variances obtained from the outer product matrix are almost systematically larger than the variances obtained using the information and Hessian matrices. In general, these results simply imply that the choice of a particular covariance matrix estimator is not neutral in the estimation of variances of coefficients in the ML framework especially in the presence of misspecification.

Thus, given that the different covariance matrix estimators yield different results under non-normal distributions, we examined further the implications of the alternative covariance matrix estimators in hypothesis testing and/or confidence interval construction for the coefficients of the model. Not surprisingly, the four covariance matrix estimators lead to the same conclusions in hypothesis testing and/or confidence interval construction, under normality of the error terms. However, the $V_{Rob.}$ estimator tends to be biased downwards if the sample size is small. However, under non-normality of the error terms, hypothesis tests based on the robust (White-type) estimator appear to be relatively superior to the others. In particular, hypothesis tests based on the information and Hessian matrices are relatively more sensitive to violations of the assumptions about the error term of the model and provide confidence intervals which are relatively narrower than they should be.

Finally, the following points can be deduced from the discussions in this Chapter:

(i) Although some studies have indicated that the MLE is inconsistent if the assumption of normality is violated, the validity of the models considered in those studies to more general models is questionable. This study suggests that inconsistency is not a serious problem for the coefficients of the model, except for the constant term which may or may not have serious consequences in applied research. The effects of the constant term in the estimation of responses will be discussed later in Chapter 9.

(ii) As far as statistical inference is concerned, the MLE estimator performs quite well regardless of the variance-covariance matrix estimator used to estimate the variances of the coefficients, provided that the assumption of normality of the errors holds. Under the non-normal distributions, the results vary slightly depending on the variance-covariance matrix estimator employed to estimate the variances of the coefficients of the model. This study suggests that the robust (White-type) estimator is relatively superior under the non-normal distributions, and is as good as others under the normal distribution except for the small sample size. The covariance matrix estimators based on the information and Hessian matrices stand as the less preferable under the non-normal error terms. These conclusions, of course, need to be tested in practice using a large variety of econometric and/or economic applications.

Chapter 8

The 3SE Vs H2S Estimator: The Effects of Correlation

8.1 Introduction

The Heckman's two-step estimator (H2S) is often used in the estimation of tobit models because of its computational ease. However, as discussed in Chapter 3, collinearity between the explanatory variables, x's, and the estimated inverse of Mill's ratio, $\hat{\lambda}(x_i'\hat{\alpha})$, is unavoidable and often strong in the second step of the H2S procedure.

This is not the case for the three-step estimator (3SE), which is constructed to avoid the multicollinearity problem. Thus, in finite samples, the H2S estimator is likely to be less precise depending on the degree of correlation that exists between the explanatory variables and the inverse of the estimated Mill's ratio.

The purpose of this Chapter is to investigate the effects of this correlation on the performance of the estimators. Section 8.2 presents a brief review of the H2S and the

3SE estimators. The experimental design which is used in this particular experiment is presented in Section 8.3. The results are discussed in Section 8.4 and Section 8.5 presents the main conclusions.

8.2 An Overview of the H2S and 3S Estimators

In order to understand the design of the experiment in the next section, it is important to summarize the main results and relationships of the H2S and 3SE estimators. These results are extracted from the detailed discussions provided in Chapters 2 and 3 of this study.

We recall the model given by

$$y_i = x_i'\beta + \sigma\lambda(x_i'\alpha) + \varepsilon_i \tag{8.1}$$

where the various components of the model are defined in Chapter 2, Section 2.4.

Heckman's two-step estimator (H2S) involves the estimation of α and hence $\lambda(x_i'\alpha)$ using the probit maximum likelihood estimator in the first step of the procedure. Then the method of ordinary least squares is applied to (8.1) after replacing α and $\lambda(x_i'\alpha)$ by their consistent estimates, say, $\hat{\alpha}$ and $\hat{\lambda}(x_i'\hat{\alpha})$, respectively, based only on the N_1 observations for which $y_i > 0$.

The H2S estimator is straightforward and is often preferred because of its computational ease. However, some studies have indicated that the H2S estimator performs relatively poorly in finite samples [see Wales and Woodland (1980), Nelson (1984), Paarsch (1984), Manning, Duan and Rogers (1987), Hay, Leu and Rohrer (1987), Hartman (1991), Nawata (1993, 1994)]. The major reason is that even if no correlation exists between the x's, there is always some correlation (often strong) between the x_i' and $\hat{\lambda}(x_i'\hat{\alpha})$. Thus this correlation is inherent to the particular form of the model. The correlation is usually strong because $\lambda(x_i'\alpha)$ is approximately linear in the index $x_i'\alpha$ for a wide range of observations on x_i' and hence strongly correlated with x_i'. The reliability of the estimates from the H2S procedure depends on the degree of correlation that exists between the explanatory variables in general and between $\hat{\lambda}(x_i'\hat{\alpha})$ and x_i' in particular.

Further, contrary to theoretical expectations the estimated value of σ obtained by directly regressing (8.1) is not guaranteed to be positive. This problem is usually ignored because it has very little practical importance in applied research. However, it can be considered as an indication of the unreliability of the H2S estimator.

On the other hand, the three-step estimator discussed in Chapter 3, provides a substantial improvement over the H2S estimator in that the above problems do not occur. Computationally, the 3SE is also very simple.

In order to discuss the 3SE estimator further, let us summarize the main steps involved in the estimation procedure. Note that the first step of the procedure is the same as that of the H2S in which the probit maximum likelihood estimator is used to obtain consistent estimates of α and hence $\lambda(x_i'\alpha)$. The last two steps of the 3SE are as follows.

In step two, because $\alpha = \beta/\sigma$, equation (8.1) can be written as

$$y_i = \sigma[x_i'\alpha + \lambda(x_i'\alpha)] + \varepsilon_i \tag{8.2}$$

Equation (8.2) is a simple regression model with no constant term. Hence, σ can be estimated by regressing y_i on $[x_i'\hat{\alpha} + \hat{\lambda}(x_i'\hat{\alpha})]$, using only those N_1 observations for which $y_i > 0$. Further, since both the right and left hand side of the model (8.2) are positive, by contrast with the H2S, the estimated value of σ will be always positive. More importantly, σ can be estimated without any problem even if $x_i'\hat{\alpha}$ and $\hat{\lambda}(x_i'\hat{\alpha})$ are not distinguishable (i.e., with out multicollinearity problem).

Finally, the β coefficients of the model can be estimated by one additional step, as follows.

Let $\hat{\sigma}_{3S}$ be the estimate of σ obtained by regressing y_i on $[x_i'\hat{\alpha} + \hat{\lambda}(x_i'\hat{\alpha})]$. Then, substituting $\hat{\sigma}_{3S}$ in equation (8.1) and rearranging the model, gives

$$y_i - \hat{\sigma}_{3S}\hat{\lambda}(x_i'\hat{\alpha}) = x_i'\beta + \varepsilon_i + \eta_i \tag{8.3}$$

which can be written as

$$\tilde{y}_i = x_i'\beta + \varepsilon_i + \eta_i \tag{8.4}$$

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

Equation (8.4) is a simple linear model and hence one can estimate the coefficients of the model by regressing $\tilde{y}_i$ on the x's using the N_1 observations. There are two important differences between the H2S and the 3SE estimators. (i) Equation (8.4) does not involve $\hat{\lambda}(x_i'\hat{\alpha})$ which is the main source of the multicollinearity problem in the H2S procedure; (ii) The estimated value of σ, $\hat{\sigma}_{3S}$, is guaranteed to be positive and can be estimated with more precision in the second step of the three-step procedure. Other details including the asymptotic properties of the estimators are provided in Chapters 2 and 3.

Given the above results, the main objective of this Chapter is to investigate the effects of collinearity between the explanatory variables and the inverse of the estimated Mill's ratio on the performance of the estimators, for different sample sizes and degrees of censoring. Collinearity may arise in the model due to several reasons. Some of these are: (i) Economic variables are usually correlated and this may increase the degree of correlation between x_i' and $\lambda(x_i'\alpha)$. That is, if the x's are correlated with each other this may then lead to a stronger correlation between x_i' and $\hat{\lambda}(x_i'\hat{\alpha})$. (ii) We know that $\lambda(x_i'\alpha)$ can be approximated by a linear function of the index $x_i'\alpha$ in a wide range of observations on the x's. Thus the correlation between x_i' and $\lambda(x_i'\alpha)$ can be high. (iii) The level of correlation between x_i' and $\hat{\lambda}(x_i'\hat{\alpha})$ can be high for high degrees of censoring. This is because as the degree of censoring increases the observations in $\lambda(x_i'\alpha)$ may shrink to a certain range where $\lambda(x_i'\alpha)$ is approximately linear in $x_i'\alpha$ and hence x_i'.

Given this, it is necessary to modify the design of the experiment provided in Chapter 5 in such a way that the objectives of this Chapter can be met. These changes are presented in the following Section.

8.3 The Design of the Experiment

8.3.1 The Model

The specific form of the model to be investigated is similar to the model discussed in Chapter 5. The model is given by

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \quad i = 1, ..., N \tag{8.5}$$

$$\begin{aligned} y_i &= y_i^* \text{ if } y_i^* > 0 \\ &= 0 \text{ if } y_i^* \leq 0 \end{aligned} \tag{8.6}$$

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 when 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, and

u is an Nx1 vector of identically and independently distributed normal random errors with zero mean and variance σ^2.

The objectives of the Monte Carlo experiment in this particular Chapter are to investigate:

1. the effects of collinearity between the x's and $\hat{\lambda}(x_i'\hat{\alpha})$ on the performance of the estimators;
2. the effects of the degree of censoring; and
3. the effects of sample size.

Below is the data generation process which is used to achieve these objectives.

8.3.2 The Data Generation Process

The data generation process of this experiment is based on the model defined in (8.5) which is given by

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \qquad i = 1, ..., N \tag{8.7}$$

where the y_i^*'s, the x's, the β's and the u_i's are as defined in (8.5)-(8.6).

Given (8.7), the various components of the model are determined as follows:

(i) The explanatory variables, x's, of the model are generated in the same way as in Chapter 5. That is, 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. The observations on the second explanatory variable, x_{2i}, $i = 1, ..., N$, are generated uniformly from the interval [-1,1] and independently of x_{1i}.

(ii) The disturbance term of the model, u_i, is assumed to be normally distributed with mean zero and variance equal to one. That is, the error term is generated from the standard normal distribution such that $u_i \sim N(0, 1)$. Note that since the objective is to investigate the effects of correlation combined with different sample sizes and degrees of censoring, it is considered sufficient to examine these effects under the correct specification of the model. That is, when the assumption of normality about the error term holds.

(iii) The main departure in the design of this experiment, compared to that of Chapter 5, is the determination of the coefficients of the model, particularly β_1 and β_2. Note that, in Chapter 5, the parameters β_1 and β_2 are set to be equal to one, (i.e., $\beta_1 = \beta_2 = 1$). However, in this case β_1 and β_2 are determined in such a way that different levels of correlation can be obtained between $\hat{\lambda}(x_i'\hat{\alpha})$ and the index $\hat{z}_i = x_i'\hat{\alpha}$, and

β_0 is used to determine the degree of censoring and hence takes different values depending on the particular level of censoring.

It is important to note at this stage that the choice of the parameter values in the data generation process is not neutral. This is because different sets of parameter values lead to different levels of correlation and consequently affect the experimental results. In other words, the objective of this experiment can also be viewed as investigating the effects of changing parameter values (or the effects of the data generation process); which indirectly is the same as investigating the effects of degree of correlation. Note that most studies which are related to tobit (or sample selection) models have concentrated on the effects of the error distribution, sample size and degree of censoring. Thus the role of the data generation process such as the choice of the parameters has been ignored.

Given this, let's define the following:

Let $\hat{\rho}_{\hat{z}\hat{\lambda}}$ be the correlation between the index $z_i = x_i'\hat{\alpha}$ and $\hat{\lambda}(x_i'\hat{\alpha})$ using the observations for which y_i is positive. Clearly, $\hat{\rho}_{\hat{z}\hat{\lambda}}$ can be used to indicate the level of correlation that exists between the x's and $\hat{\lambda}(x_i'\hat{\alpha})$. Hence, while the 3SE is unlikely to be affected, the quality of the estimates from the H2S estimator depends on the level of this correlation. The aim is therefore to investigate the relative performance of the H2S and the 3SE estimators under various levels of the correlation, $\hat{\rho}_{\hat{z}\hat{\lambda}}$. To achieve this we proceed as follows.

As discussed above, $\hat{\lambda}(x_i'\hat{\alpha})$ can be approximated as a linear function of the index

$\hat{z}_i = x_i'\hat{\alpha}$, where $\alpha = \beta/\sigma$, in a wide range of observations on x_i'. Assuming that σ^2 is fixed (constant), which is equal to one in this case, one way of controlling the range of observations in the x's is by changing the parameter values of the β's so that different levels of $\hat{\rho}_{\hat{z}\hat{\lambda}}$ can be obtained.

For example, given (i) and (ii) above, the values of $\beta_0 = -4.000$, $\beta_1 = 4.000$ and $\beta_2 = 0.500$ yield an average correlation of $\hat{\rho}_{\hat{z}\hat{\lambda}} = -0.50$ and a degree of censoring of 25%. If we want to increase the level of correlation, say, to $\hat{\rho}_{\hat{z}\hat{\lambda}} = -0.90$ but have the same degree of censoring (i.e., 25%), it can be obtained by changing the parameter values of β_0, β_1 and β_2 to -0.950, 1.000 and 0.100, respectively. Various levels of $\hat{\rho}_{\hat{z}\hat{\lambda}}$ ranging from -0.50 to -0.95 are obtained using a similar procedure for all sample sizes and degrees of censoring. These values are determined based on preliminary experiments.

Further, similar to the experimental design in Chapter 5, we considered three levels of censoring, namely, low (25%), medium (50%) and high (75%). The effects of sample size are also investigated by considering the sample sizes of 100, 200 and 400, respectively, for small, medium and large sample sizes. Finally, the results of the experiment are computed based on 3000 replications (samples). The same seed was used in all experiments. The results are discussed in the following section.

8.4 Comparison of Results

As noted earlier, the main purpose of this Chapter is to investigate the relative performance of the Heckman's two-step (H2S) and the three-step (3S) estimators under various levels of correlation between the explanatory variables and the estimated inverse of Mill's ratio. The ordinary least squares estimator based on the observations for which y_i is positive (OLSP) and the maximum likelihood estimator (MLE) are also included for comparison purposes.

Given this, consider Table 8.1 which presents results for a sample size of 100 and a 25% degree of censoring. Column (1) of the table lists different correlation levels ranging from -0.50 (for low) to -0.95 (for high). Column (2) provides the list of estimators and the parameters to be estimated. In Column (3) are the true parameter values. Columns (4) and (5), respectively, present the estimated mean and the standard errors of the estimates. Finally, Columns (6) and (7) depict the Bias and the RMSE of the estimates, respectively. Other tables may be interpreted in a similar way.

As can be seen from Table 8.1, when $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.50 the RMSEs of β_0, β_1 and β_2 are 0.412, 0.149 and 0.207, respectively, using the H2S estimator. The corresponding estimates using the 3SE are given by 0.374, 0.138 and 0.206, respectively. These results are close, implying that both the H2S and the 3SE estimators may provide similar results provided that the correlation is small. Similarly, the performances of the H2S and the 3S estimators remain fairly close to each other given that the correlation is not large. This can be seen from the results corresponding to correlation levels as high as -0.70 in Table 8.1. Note that bias is not a problem for all the estimators except for the OLSP, as would be expected.

However, using the RMSE criteria, the gap between the H2S and the 3SE starts to increase as the level of correlation increases and gets wider for high levels of correlation. For instance, the RMSEs of the H2S estimates of β_1 and β_2 are about 26 and 67

Table 8.1: Comparison of Estimators under Various levels of Correlation, Given N=100 and 25% Degree of Censoring.

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$)	Estimator		True Value	Estimated Mean	Standard Error	Bias	RMSE
(1)	(2)		(3)	(4)	(5)	(6)	(7)
-0.50	H2S	β_0	-4.000	-3.994	0.408	0.056	0.412
		β_1	4.000	3.983	0.148	-0.017	0.149
		β_2	0.500	0.492	0.207	-0.008	0.207
	3SE	β_0	-4.000	-3.925	0.366	-0.075	0.374
		β_1	4.000	3.976	0.137	-0.024	0.138
		β_2	0.500	0.492	0.206	-0.008	0.206
	OLSP	β_0	-4.000	-3.552	0.331	0.448	0.557
		β_1	4.000	3.853	0.127	-0.147	0.194
		β_2	0.500	0.498	0.205	-0.002	0.205
	MLE	β_0	-4.000	-4.015	0.309	-0.015	0.310
		β_1	4.000	4.006	0.120	0.006	0.120
		β_2	0.500	0.492	0.204	-0.008	0.204
-0.60	H2S	β_0	-3.000	-2.941	0.427	0.059	0.431
		β_1	3.000	2.980	0.154	-0.020	0.155
		β_2	1.000	0.998	0.0216	-0.002	0.216
	3SE	β_0	-3.000	-2.925	0.363	0.075	0.371
		β_1	3.000	2.974	0.134	-0.026	0.136
		β_2	1.000	0.997	0.215	-0.003	0.215
	OLSP	β_0	-3.000	-2.419	0.315	0.581	0.661
		β_1	3.000	2.810	0.121	-0.190	0.226
		β_2	1.000	0.953	0.214	-0.047	0.219
	MLE	β_0	-3.000	-3.012	0.291	-0.012	0.291
		β_1	3.000	3.003	0.113	0.003	0.113
		β_2	1.000	1.004	0.208	0.004	0.208
-0.70	H2S	β_0	-2.000	-1.939	0.512	0.061	0.456
		β_1	2.000	1.980	0.179	-0.020	0.180
		β_2	0.500	0.499	0.194	-0.001	0.194
	3SE	β_0	-2.000	-1.941	0.355	0.059	0.360
		β_1	2.000	1.981	0.132	-0.019	0.134
		β_2	0.500	0.499	0.192	-0.001	0.192
	OLSP	β_0	-2.000	-1.186	0.278	0.814	0.860
		β_1	2.000	1.737	0.112	-0.263	0.285
		β_2	0.500	0.504	0.187	0.004	0.187
	MLE	β_0	-2.000	-2.004	0.269	-0.004	0.269
		β_1	2.000	2.001	0.107	0.001	0.107
		β_2	0.500	0.501	0.186	0.001	0.186

Table 8.1 continued next page

Cont'd from Table 8.1.

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$)	Estimator		True Value	Estimated Mean	Standard Error	Bias	RMSE
(1)	(2)		(3)	(4)	(5)	(6)	(7)
-0.85	H2S	β_0	-0.750	-0.713	0.611	0.037	0.612
		β_1	1.000	0.989	0.190	-0.011	0.190
		β_2	1.000	0.991	0.234	-0.009	0.234
	3SE	β_0	-0.750	-0.703	0.303	0.047	0.307
		β_1	1.000	0.985	0.113	-0.015	0.114
		β_2	1.000	0.989	0.186	-0.011	0.186
	OLSP	β_0	-0.750	0.080	0.220	0.830	0.858
		β_1	1.000	0.760	0.093	-0.240	0.258
		β_2	1.000	0.788	0.167	-0.212	0.270
	MLE	β_0	-0.750	-0.755	0.241	-0.005	0.241
		β_1	1.000	1.002	0.099	0.002	0.099
		β_2	1.000	1.002	0.177	0.002	0.177
-0.90	H2S	β_0	-0.950	-0.888	0.875	0.062	0.877
		β_1	1.000	0.978	0.269	-0.022	0.270
		β_2	0.100	0.099	0.187	-0.001	0.187
	3SE	β_0	-0.950	-0.899	0.322	0.061	0.328
		β_1	1.000	0.979	0.119	-0.021	0.121
		β_2	0.100	0.101	0.173	0.001	0.173
	OLSP	β_0	-0.950	0.057	0.215	1.007	1.029
		β_1	1.000	0.606	0.092	-0.304	0.318
		β_2	0.100	0.068	0.160	-0.032	0.163
	MLE	β_0	-0.950	-0.954	0.250	-0.004	0.250
		β_1	1.000	1.001	0.100	0.001	0.100
		β_2	0.100	0.104	0.168	0.004	0.167
-0.95	H2S	β_0	-0.150	-0.141	1.385	0.009	1.385
		β_1	0.500	0.488	0.320	-0.012	0.321
		β_2	0.500	0.494	0.395	-0.006	0.395
	3SE	β_0	-0.150	-0.121	0.253	0.029	0.255
		β_1	0.500	0.491	0.099	-0.009	0.099
		β_2	0.500	0.494	0.184	-0.006	0.184
	OLSP	β_0	-0.150	0.636	0.182	0.786	0.807
		β_1	0.500	0.303	0.081	-0.192	0.208
		β_2	0.500	0.313	0.156	-0.187	0.243
	MLE	β_0	-0.150	-0.155	0.222	-0.005	0.222
		β_1	0.500	0.500	0.092	0.000	0.092
		β_2	0.500	0.504	0.178	0.004	0.178

percent larger than their corresponding 3S estimates, respectively, when $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.85. The difference between the two estimators is much bigger for the constant term in which the RMSE for the H2S estimator is almost twice that of the 3S estimator (see Table 8.1). These relative differences are quite large compared to, say, less than 10 percent when $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.50 or at most 30 percent when $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.70.

The relative performance of the H2S estimator, compared to the 3SE or MLE, deteriorates further for higher levels of correlation. More specifically, Table 8.1 depicts that when $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.95, the RMSEs of β_1 and β_2 under the H2S estimator are over 2 to 3 times that of the 3SE. The difference is even much bigger for the constant term in which the RMSE of β_0 for the H2S estimator is about 5 times that of the 3SE estimator. Moreover, it is surprising to note that if the correlation is high the H2S estimator performs badly even compared to the OLSP estimator in terms of the RMSE criteria.

On the other hand, the 3SE provides results which are much better than the H2S estimator and remain quite close to the MLE estimator in all cases. For example, given a high correlation (i.e., $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.95) the difference in RMSEs of the coefficients between the MLE and 3SE is at most 15 percent as compared to as large as six times for the H2S estimator.

The results of the experiment also reveal similar conclusions for the medium sample size. Table 8.2 below depicts results under various levels of correlation, given a sample size of 200 and a 25% degree of censoring. As can be seen from Table 8.2, the difference in RMSEs between the H2S and the 3S estimators ranges from zero to about 10 percent for the three coefficients, provided that the correlation is low, i.e., -0.50. Again, although the relative performance of the H2S and the 3S estimators

Table 8.2: Comparison of Estimators under Various levels of Correlation, Given N=200 and 25% Degree of Censoring.

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$)	Estimator		True Value	Estimated Mean	Standard Error	Bias	RMSE
(1)	(2)		(3)	(4)	(5)	(6)	(7)
-0.50	H2S	β_0	-4.000	-3.969	0.295	0.031	0.296
		β_1	4.000	3.990	0.107	-0.010	0.108
		β_2	0.500	0.497	0.137	-0.003	0.137
	3SE	β_0	-4.000	-3.956	0.259	0.044	0.263
		β_1	4.000	3.986	0.097	-0.014	0.098
		β_2	0.500	0.497	0.136	-0.003	0.136
	OLSP	β_0	-4.000	-3.545	0.243	0.455	0.516
		β_1	4.000	3.850	0.090	-0.150	0.175
		β_2	0.500	0.488	0.137	-0.012	0.137
	MLE	β_0	-4.000	-4.006	0.217	-0.006	0.217
		β_1	4.000	4.003	0.084	0.003	0.084
		β_2	0.500	0.498	0.134	-0.002	0.134
-0.60	H2S	β_0	-3.000	-2.962	0.323	0.038	0.325
		β_1	3.000	2.987	0.116	0.013	0.117
		β_2	1.000	0.996	0.147	-0.004	0.147
	3SE	β_0	-3.000	-2.963	0.268	0.037	0.271
		β_1	3.000	2.988	0.100	-0.012	0.101
		β_2	1.000	0.996	0.140	-0.004	0.140
	OLSP	β_0	-3.000	-2.341	0.222	0.658	0.695
		β_1	3.000	2.783	0.087	-0.217	0.233
		β_2	1.000	0.941	0.138	-0.059	0.150
	MLE	β_0	-3.000	-3.008	0.208	-0.008	0.208
		β_1	3.000	3.002	0.082	0.002	0.082
		β_2	1.000	0.999	0.137	-0.001	0.137
-0.70	H2S	β_0	-2.000	-1.972	0.359	0.028	0.360
		β_1	2.000	1.991	0.126	-0.009	0.127
		β_2	0.500	0.495	0.136	-0.005	0.137
	3SE	β_0	-2.000	-1.970	0.261	0.030	0.263
		β_1	2.000	1.990	0.097	-0.010	0.098
		β_2	0.500	0.495	0.135	-0.005	0.135
	OLSP	β_0	-2.000	-1.216	0.205	0.784	0.810
		β_1	2.000	1.745	0.081	-0.255	0.267
		β_2	0.500	0.450	0.132	-0.050	0.142
	MLE	β_0	-2.000	-2.005	0.195	-0.005	0.195
		β_1	2.000	2.002	0.077	0.002	0.007
		β_2	0.500	0.496	0.131	-0.004	0.131

Table 8.2 continued next page

Cont'd from Table 8.2

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$) (1)	Estimator (2)		True Value (3)	Estimated Mean (4)	Standard Error (5)	Bias (6)	RMSE (7)
-0.85	H2S	β_0	-0.750	-0.740	0.441	0.010	0.441
		β_1	1.000	0.996	0.139	-0.004	0.139
		β_2	1.000	0.991	0.175	-0.009	0.175
	3SE	β_0	-0.750	-0.730	0.207	0.019	0.208
		β_1	1.000	0.994	0.078	-0.006	0.078
		β_2	1.000	0.990	0.141	-0.010	0.141
	OLSP	β_0	-0.750	0.078	0.145	0.828	0.841
		β_1	1.000	0.755	0.063	-0.245	0.253
		β_2	1.000	0.783	0.125	-0.217	0.250
	MLE	β_0	-0.750	-0.752	0.162	-0.002	0.162
		β_1	1.000	1.001	0.067	0.001	0.067
		β_2	1.000	0.997	0.133	-0.003	0.133
-0.90	H2S	β_0	-0.950	-0.943	0.618	0.007	0.618
		β_1	1.000	0.997	0.193	-0.003	0.193
		β_2	0.100	0.095	0.139	-0.005	0.139
	3SE	β_0	-0.950	-0.930	0.231	0.020	0.232
		β_1	1.000	0.994	0.087	-0.006	0.087
		β_2	0.100	0.095	0.135	-0.005	0.135
	OLSP	β_0	-0.950	0.062	0.156	1.012	1.024
		β_1	1.000	0.695	0.067	-0.305	0.313
		β_2	0.100	0.071	0.123	-0.029	0.126
	MLE	β_0	-0.950	-0.958	0.178	-0.008	0.178
		β_1	1.000	1.002	0.072	0.002	0.072
		β_2	0.100	0.095	0.132	-0.005	0.132
-0.95	H2S	β_0	-0.150	-0.167	1.011	-0.017	1.011
		β_1	0.500	0.499	0.242	-0.001	0.242
		β_2	0.500	0.498	0.281	-0.002	0.281
	3SE	β_0	-0.150	-0.133	0.176	0.016	0.178
		β_1	0.500	0.496	0.069	-0.004	0.069
		β_2	0.500	0.494	0.136	-0.006	0.136
	OLSP	β_0	-0.150	0.624	0.126	0.774	0.784
		β_1	0.500	0.313	0.056	-0.187	0.195
		β_2	0.500	0.318	0.115	-0.182	0.215
	MLE	β_0	-0.150	-0.148	0.156	-0.002	0.156
		β_1	0.500	0.500	0.065	0.000	0.065
		β_2	0.500	0.498	0.132	-0.002	0.132

starts to depart slowly as the level of correlation increases, the two estimators remain fairly close for levels of correlation up to -0.70. For example, the RMSEs of the H2S estimates are about 16 and 26 percent larger than their corresponding 3S estimates, respectively, for the correlation levels of -0.60 and -0.70. But, this difference widens for higher levels of correlation and the quality of the H2S estimator deteriorates rapidly. This is evident from the results corresponding to $\hat{\rho}_{\hat{z}\hat{\lambda}} = -0.85$ and -0.95. Specifically, when $\hat{\rho}_{\hat{z}\hat{\lambda}} = -0.95$ the RMSEs of β_1, β_2 and β_0 under the H2S are, respectively, about 2, 3 and 5 times larger than their corresponding 3S estimates. Whereas, the RMSEs of the 3S estimates, as compared to the ML estimates, are only 2 to 3 percent larger for β_1 and β_2 and about 15 percent for β_0.

As to the effects of sample size, the results in Tables 8.1-8.2 also suggest that, as the sample size increases and provided that the correlation level is not large, both the H2S and the 3S estimators may provide similar results. This evidence is particularly clear from Table 8.3 which depicts results for various levels of correlation, given a large sample size (400) and a 25% degree of censoring. Specifically, given low correlation (i.e., $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.50), the RMSEs for β_0, β_1 and β_2, respectively, are given by 0.209, 0.077 and 0.095 for H2S compared to 0.186, 0.070 and 0.095 for the 3SE estimator. These results are very close (some times the same) and imply that, given a low level of correlation and a small degree of censoring, one needs a sample size of at least 400 to obtain similar (or sufficiently close) results from both the H2S and 3SE estimators. However, as the correlation increases, even slightly, the relative performance of the H2S estimator starts to deteriorate and gets worse when the correlation is very high even if the sample size is large. In other words, the H2S estimator, compared to the 3SE, is highly sensitive to the increases in correlation irrespective of the sample size.

Table 8.3: Comparison of Estimators under Various levels of Correlation, Given N=400 and 25% Degree of Censoring.

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$) (1)	Estimator (2)		True Value (3)	Estimated Mean (4)	Standard Error (5)	Bias (6)	RMSE (7)
-0.50	H2S	β_0	-4.000	-3.989	0.209	0.011	0.209
		β_1	4.000	3.996	0.077	-0.004	0.077
		β_2	0.500	0.497	0.095	-0.003	0.095
	3SE	β_0	-4.000	-3.982	0.185	0.018	0.186
		β_1	4.000	3.994	0.070	-0.006	0.070
		β_2	0.500	0.496	0.095	-0.004	0.095
	OLSP	β_0	-4.000	-3.528	0.161	0.472	0.499
		β_1	4.000	3.843	0.063	-0.157	0.169
		β_2	0.500	0.467	0.094	-0.033	0.100
	MLE	β_0	-4.000	-4.005	0.152	-0.005	0.152
		β_1	4.000	4.002	0.060	0.002	0.060
		β_2	0.500	0.498	0.093	-0.002	0.093
-0.60	H2S	β_0	-3.000	-2.984	0.214	0.016	0.215
		β_1	3.000	2.994	0.078	-0.006	0.078
		β_2	1.000	0.998	0.098	-0.002	0.098
	3SE	β_0	-3.000	-2.982	0.177	0.018	0.177
		β_1	3.000	2.994	0.066	-0.006	0.066
		β_2	1.000	0.998	0.097	-0.002	0.097
	OLSP	β_0	-3.000	-2.408	0.151	0.592	0.611
		β_1	3.000	2.805	0.059	-0.195	0.204
		β_2	1.000	0.972	0.095	-0.028	0.099
	MLE	β_0	-3.000	-3.002	0.141	-0.002	0.141
		β_1	3.000	3.000	0.056	0.000	0.056
		β_2	1.000	0.999	0.096	-0.001	0.096
-0.70	H2S	β_0	-2.000	-1.986	0.251	0.014	0.252
		β_1	2.000	1.996	0.088	-0.004	0.088
		β_2	0.500	0.499	0.097	-0.001	0.097
	3SE	β_0	-2.000	-1.983	0.183	0.017	0.184
		β_1	2.000	1.994	0.068	-0.006	0.068
		β_2	0.500	0.499	0.096	-0.001	0.096
	OLSP	β_0	-2.000	-1.191	0.139	0.809	0.821
		β_1	2.000	1.738	0.056	-0.262	0.268
		β_2	0.500	0.449	0.094	-0.051	0.107
	MLE	β_0	-2.000	-2.002	0.135	-0.002	0.135
		β_1	2.000	2.001	0.054	0.001	0.054
		β_2	0.500	0.500	0.093	0.000	0.093

Table 8.3 continued next page

Cont'd from Table 8.3

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$)	Estimator		True Value	Estimated Mean	Standard Error	Bias	RMSE
(1)	(2)		(3)	(4)	(5)	(6)	(7)
-0.85	H2S	β_0	-0.750	-0.745	0.320	0.005	0.320
		β_1	1.000	0.998	0.102	-0.002	0.102
		β_2	1.000	0.995	0.130	-0.005	0.130
	3SE	β_0	-0.750	-0.739	0.148	0.011	0.149
		β_1	1.000	0.997	0.056	-0.003	0.056
		β_2	1.000	0.994	0.100	-0.006	0.100
	OLSP	β_0	-0.750	0.101	0.104	0.851	0.857
		β_1	1.000	0.748	0.045	-0.252	0.256
		β_2	1.000	0.755	0.089	-0.245	0.261
	MLE	β_0	-0.750	-0.752	0.115	-0.002	0.115
		β_1	1.000	1.001	0.047	0.001	0.047
		β_2	1.000	0.998	0.093	-0.002	0.093
-0.90	H2S	β_0	-0.950	-0.926	0.442	0.024	0.443
		β_1	1.000	0.992	0.139	-0.008	0.140
		β_2	0.100	0.100	0.098	0.000	0.098
	3SE	β_0	-0.950	-0.936	0.160	0.014	0.160
		β_1	1.000	0.995	0.060	-0.005	0.060
		β_2	0.100	0.100	0.096	0.000	0.096
	OLSP	β_0	-0.950	0.065	0.108	1.015	1.021
		β_1	1.000	0.693	0.047	-0.307	0.310
		β_2	0.100	0.066	0.088	-0.034	0.094
	MLE	β_0	-0.950	-0.951	0.122	-0.001	0.122
		β_1	1.000	1.000	0.049	0.000	0.049
		β_2	0.100	0.101	0.093	0.001	0.093
-0.95	H2S	β_0	-0.150	-0.158	0.682	-0.008	0.682
		β_1	0.500	0.500	0.171	0.000	0.171
		β_2	0.500	0.496	0.186	-0.004	0.186
	3SE	β_0	-0.150	-0.145	0.124	0.005	0.124
		β_1	0.500	0.498	0.049	-0.002	0.049
		β_2	0.500	0.496	0.095	-0.004	0.095
	OLSP	β_0	-0.150	0.636	0.088	0.786	0.791
		β_1	0.500	0.306	0.039	-0.194	0.198
		β_2	0.500	0.315	0.083	-0.185	0.203
	MLE	β_0	-0.150	-0.152	0.109	-0.002	0.109
		β_1	0.500	0.501	0.045	0.001	0.045
		β_2	0.500	0.497	0.091	-0.003	0.091

Interestingly, the difference between the 3SE and the MLE estimators remains marginal especially in β_1 and β_2. For example, as shown in Table 8.3, given a high correlation (i.e., $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.95), the RMSEs of β_1 and β_2 under the 3SE are given by 0.049 and 0.095, respectively. The corresponding estimates under the MLE are given by 0.045 and 0.091, respectively. These results imply that the difference between the 3S and the ML estimators is only marginal. Note that these results are obtained under normality of the error term a situation where the ML estimator is expected to do well. Further, as shown in Chapter 6, the 3S estimator can be as good as (or sometimes even better), if the assumption of the error term is violated, (e.g., under the t-distribution).

Further, we examined the effects of the degree of censoring combined with various levels of correlation. The summary statistics are given in Table 8.4 (and Tables A.7-A.8 of Appendix A). For example, Table 8.4 below depicts results for a sample size of 200 and a 50% degree of censoring. Before discussing the results in Table 8.4, it is important to note that, as mentioned earlier in this Chapter, an increase in the degree of censoring has an adverse effect on the level of correlation. That is, a high level of censoring causes an increase in the level of correlation. To explain this point a little further consider the following example. Recall that, given the data generation process discussed in the preceding Section, a correlation level of, say, -0.70 and a 25% degree of censoring are obtained by setting the parameter values of β_0, β_1 and β_2 to be -2.000, 2.000 and 0.500, respectively, (see Tables 8.1-8.3). But, in order to increase the degree of censoring, say, to 50%, one needs to change only the value of β_0 to -4.000 while β_1 and β_2 remain the same. However, as a result of increasing the degree of censoring the level of correlation consequently increases to -0.85 and hence should

Table 8.4: Comparison of Estimators under Various levels of Correlation, Given N=200 and 50% Degree of Censoring.

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$)	Estimator		True Value	Estimated Mean	Standard Error	Bias	RMSE
(1)	(2)		(3)	(4)	(5)	(6)	(7)
-0.60	H2S	β_0	-8.000	-7.943	0.671	0.057	0.673
		β_1	4.000	3.984	0.212	-0.016	0.212
		β_2	0.500	0.491	0.174	-0.009	0.175
	3SE	β_0	-8.000	-7.909	0.561	0.091	0.568
		β_1	4.000	3.974	0.182	-0.026	0.184
		β_2	0.500	0.488	0.171	-0.012	0.172
	OLSP	β_0	-8.000	-6.909	0.475	1.091	1.190
		β_1	4.000	3.677	0.159	-0.323	0.360
		β_2	0.500	0.428	0.169	-0.072	0.184
	MLE	β_0	-8.000	-8.021	0.431	-0.021	0.431
		β_1	4.000	4.007	0.145	0.007	0.145
		β_2	0.500	0.495	0.165	-0.005	0.165
-0.85	H2S	β_0	-4.000	-3.924	0.845	0.076	0.849
		β_1	2.000	1.977	0.256	-0.023	0.257
		β_2	0.500	0.494	0.178	-0.006	0.178
	3SE	β_0	-4.000	-3.936	0.486	0.064	0.490
		β_1	2.000	1.980	0.158	-0.020	0.160
		β_2	0.500	0.494	0.172	-0.006	0.173
	OLSP	β_0	-4.000	-2.287	0.362	1.713	1.751
		β_1	2.000	1.501	0.127	-0.499	0.515
		β_2	0.500	0.415	0.166	-0.085	0.186
	MLE	β_0	-4.000	-4.020	0.345	-0.020	0.345
		β_1	2.000	2.006	0.119	0.006	0.119
		β_2	0.500	0.498	0.163	-0.002	0.163
-0.95	H2S	β_0	-2.000	-1.970	1.144	0.030	1.144
		β_1	1.000	0.990	0.301	-0.010	0.302
		β_2	1.000	0.991	0.309	-0.009	0.309
	3SE	β_0	-2.000	-1.975	0.347	0.025	0.347
		β_1	1.000	0.993	0.112	-0.007	0.113
		β_2	1.000	0.994	0.173	-0.006	0.173
	OLSP	β_0	-2.000	-0.259	0.223	1.741	1.755
		β_1	1.000	0.552	0.083	-0.448	0.456
		β_2	1.000	0.601	0.144	-0.399	0.424
	MLE	β_0	-2.000	-2.012	0.250	-0.012	0.250
		β_1	1.000	1.003	0.090	0.003	0.090
		β_2	1.000	1.001	0.153	0.001	0.153

be adjusted accordingly (see Table 8.4). Other values for the level of correlation are obtained in a similar way.

As shown in Table 8.4, the H2S and the 3SE are fairly close, provided that the level of correlation is not large. In particular, when $\hat{\rho}_{\hat{z}\hat{\lambda}}$=-0.60, the RMSEs of the H2S estimator are about 2 to 18 percent larger than their corresponding 3S estimates for all coefficients. However, the performance of the H2S, as compared to the 3S or ML estimators, deteriorates as the correlation increases and the RMSEs of the H2S gets incomparably large for higher levels of correlation. A further increase in the degree of censoring implies much higher levels of correlation and hence a further deterioration in the quality of the H2S estimates relative to those of the 3S or ML estimates (see also Tables A.7 and A.8, Appendix A). It is important to note that the relatively poor performance of the H2S estimator is almost entirely related to the level of correlation. This is also consistent with the conclusions made in Chapter 6 regarding the H2S estimator. That is, the main reason why the H2S performed badly for higher degrees of censoring is because an increase in the degree of censoring means an increase in the level of correlation.

Finally, in order to have a clear indication on the relative performance of the estimators, Table 8.5 below provides more information on the reliability (efficiency) of the estimators relative to the MLE estimator. That is, as shown in Table 8.5, we obtained the root mean square errors of the H2S, 3SE and OLSP estimates relative to those of the MLE estimates; for low, moderately high and high levels of correlation as well as for different sample sizes and degrees of censoring. As can be seen from the table, the results reveal that the 3SE estimator remains quite close to the MLE in all cases. Whereas, the H2S estimator performs badly sometimes even worse than

Table 8.5: Finite Sample Root Mean Square Errors Relative to those of the MLE estimator.

Correlation	Esti-	100			200			400		
$(\hat{\rho}_{\hat{z}\hat{\lambda}})$	-mator	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
					Degree of Censoring=25%					
-0.50	H2S	1.329	1.242	1.015	1.364	1.286	1.022	1.375	1.283	1.021
	3SE	1.206	1.150	1.009	1.212	1.167	1.015	1.224	1.167	1.021
	OLSP	1.796	1.617	1.005	2.373	2.083	1.022	3.283	2.817	1.075
-0.85	H2S	2.539	1.919	1.322	2.722	2.075	1.316	2.783	2.170	1.398
	3SE	1.274	1.151	1.051	1.284	1.164	1.060	1.296	1.191	1.075
	OLSP	3.560	2.606	1.525	5.191	3.776	1.879	7.452	5.445	2.806
-0.95	H2S	6.238	3.489	2.219	6.481	3.723	2.129	6.257	3.800	2.044
	3SE	1.148	1.076	1.033	1.141	1.061	1.030	1.138	1.089	1.044
	OLSP	3.653	2.261	1.365	5.026	3.000	1.629	7.256	4.400	2.231
					Degree of Censoring=50%					
-0.60	H2S	1.599	1.487	1.052	1.561	1.462	1.061	1.536	1.447	1.018
	3SE	1.309	1.246	1.036	1.318	1.269	1.042	1.279	1.233	1.018
	OLSP	2.030	1.854	1.032	2.761	2.483	1.115	3.795	3.408	1.080
-0.85	H2S	2.609	2.266	1.086	2.461	2.160	1.092	2.537	2.212	1.093
	3SE	1.464	1.367	1.057	1.420	1.344	1.061	1.430	1.329	1.047
	OLSP	3.713	3.172	1.021	5.075	4.328	1.141	7.307	6.141	1.299
-0.95	H2S	4.564	3.378	2.037	4.576	3.356	2.019	5.153	3.844	2.193
	3SE	1.367	1.251	1.139	1.388	1.255	1.131	1.392	1.281	1.131
	OLSP	4.601	3.392	2.028	7.020	5.066	2.771	10.170	7.422	3.877

the OLSP estimator except when the correlation is small.

In general, the 3S estimator outperforms the H2S estimator in almost all cases and the difference in relative performance between the two estimators becomes substantially high even for moderately high levels of correlation, irrespective of the sample size. If the correlation is high, say, about -0.95, then the H2S estimator can be even less preferrable than the biased OLSP estimator using the RMSE criteria. Note that all estimators perform quite well in terms of bias except the OLSP. They also have the correct signs.

Another important aspect in the comparison of the estimators is to examine their likely performance in hypothesis testing and/or construction of confidence intervals under various levels of correlation, sample sizes and degrees of censoring. One way of doing this is to compare the asymptotic standard errors of the estimators with their corresponding Monte Carlo standard errors. As discussed in Chapter 6, this comparison has very important implications for applied research. That is, if the asymptotic standard errors of an estimator are lower than their corresponding Monte Carlo standard errors then the confidence intervals for the coefficients are likely to be narrower (over precise) than they should be. On the other hand, asymptotic variances which are larger than their corresponding true variances may imply confidence intervals which are wider than the desired level. Note that the asymptotic standard errors are computed as the square root of the diagonal elements of the variance-covariance matrices of the estimators. The variance-covariance matrices of the estimators are provided in Chapters 2 and 3 of this study. For example, the asymptotic standard errors of the MLE are obtained as the square roots of the diagonal elements of equation (2.22), Chapter 2, which is the inverse of the information matrix. Others are obtained in a similar way.

Given this, Table 8.6 provides the finite sample (Monte Carlo) standard errors relative to asymptotic standard errors of the estimators for different levels of correlation, sample sizes and degrees of censoring. As can be seen for the table, it is clear that the asymptotic standard errors of the MLE are good approximations of their corresponding Monte Carlo standard errors. Similarly, the 3SE provides Monte Carlo standard errors which are quite close to their corresponding asymptotic standard errors in almost all cases. On the other hand, the H2S estimator provides finite sample standard errors which are smaller than their corresponding asymptotic values in all cases except when the correlation is small. These results indicate that confidence intervals for the coefficients are likely to be wider than they should be if one uses the H2S estimator, unless the correlation is small. In general, hypothesis tests and/or confidence intervals based on the H2S estimator can be misleading if the correlation between the explanatory variables and the inverse of Mill's ratio is high.

As a final remark, this Chapter also demonstrates the effects of the data generation process on the performance of the estimators, a case usually overlooked in most Monte Carlo studies related to the model. That is, most studies have concentrated on the effects of the error distribution, sample size and degree of censoring on the performance of the estimators (e.g., Paarsch (1984), Moon (1989)). The effects of the data generation process such as the β's and or X's is ignored. However, as shown in this Chapter, different sets of parameter values may lead to different levels of correlation and consequently implying different conclusions about the estimators, particularly the H2S estimator which is very sensitive to the degree of correlation. In other words, it is not surprising if the H2S estimator sometimes performs very poorly;

Table 8.6: Finite Sample Standard Errors Relative to Asymptotic Standard Errors.

Correlation ($\hat{\rho}_{\hat{z}\hat{\lambda}}$)	Esti--mator	100			200			400		
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
					Degree of Censoring=25%					
-0.50	H2S	0.861	0.881	1.095	0.949	0.955	1.015	0.946	0.963	0.960
	3SE	1.034	1.030	1.095	1.061	1.043	1.007	1.063	1.061	0.960
	OLSP	0.925	0.948	1.068	0.984	0.968	0.986	0.920	0.951	0.931
	MLE	1.013	1.017	1.096	1.009	1.000	0.993	1.000	1.017	0.958
-0.85	H2S	0.741	0.742	0.839	0.752	0.780	0.902	0.794	0.823	0.935
	3SE	0.984	0.974	0.954	0.932	0.918	1.000	0.961	0.966	1.010
	OLSP	0.738	0.808	0.831	0.659	0.741	0.862	0.693	0.776	0.873
	MLE	1.008	1.021	0.978	0.964	0.971	1.023	0.975	0.979	1.000
-0.95	H2S	0.858	0.788	0.940	0.940	0.890	1.004	0.868	0.868	0.912
	3SE	1.024	1.021	1.039	1.011	1.000	1.071	1.016	1.021	1.055
	OLSP	0.714	0.786	0.834	0.708	0.767	0.858	0.709	0.765	0.874
	MLE	1.000	1.000	1.011	1.000	1.000	1.039	0.991	0.978	1.000
					Degree of Censoring=50%					
-0.60	H2S	0.921	0.918	1.144	0.895	0.910	1.017	0.920	0.925	0.934
	3SE	0.997	0.980	1.136	1.033	1.028	1.000	1.018	1.016	0.942
	OLSP	0.896	0.903	1.095	0.908	1.082	0.966	0.883	0.901	0.888
	MLE	1.002	0.985	1.127	1.005	1.000	0.994	1.006	1.000	0.949
-0.85	H2S	0.830	0.839	1.128	0.743	0.762	1.017	0.749	0.777	0.967
	3SE	0.963	0.958	1.157	0.915	0.913	1.029	0.895	0.911	0.926
	OLSP	0.717	0.763	1.034	0.693	0.743	0.943	0.646	0.698	0.822
	MLE	0.998	0.988	1.025	1.000	0.992	1.032	0.996	1.000	0.955
-0.95	H2S	0.610	0.650	0.608	0.631	0.661	0.710	0.669	0.718	0.743
	3SE	0.734	0.761	0.910	0.766	0.772	0.956	0.747	0.779	0.985
	OLSP	0.451	0.529	0.708	0.454	0.515	0.795	0.439	0.517	0.759
	MLE	1.044	1.039	1.009	0.992	0.989	1.000	0.994	1.016	1.046

it may be due to the bias against the estimator resulting from the experimental design. Thus, the validity of the conclusions made based on a particular data generation process can be misleading and hence should be treated with caution. Some useful discussion along this line is given by Leung and Yu (1994).

8.5 Summary and Conclusions

It is well known that the Heckman's two-step (H2S) estimator of the tobit model has poor finite sample properties. This is because of the unavoidable and often strong multicollinearity between the explanatory variables and the estimated inverse of Mill's ratio. On the other hand, this is not the case for the three-step estimator (3SE) which is suggested in this study.

In this Chapter, we examined the finite sample properties of the H2S and 3SE estimators along with other estimators of the tobit model, namely, the ordinary least squares based on those observations for which the dependent variable, y_i, is positive (OLSP) and the maximum likelihood estimator (MLE) of the model. The effects of collinearity on the performance of the estimators is investigated under different sample sizes and degrees of censoring. The main conclusions, among others, include the following points.

Both the H2S and the 3S estimators may provide similar results, provided that the correlation is not large. More specifically, our results suggest that the two estimators may perform well for correlation levels as large as -0.70. However, as the correlation increases the relative performance of the H2S estimator starts to decline and deteriorates very quickly for higher levels of correlation. In particular, if the correlation is very high, say, -0.90 or more, the H2S estimator performs badly, even compared to the biased OLSP estimator.

As to the effects of sample size, one needs a sample size of at least 400 to obtain sufficiently close (or the same) results using both the H2S and the 3S estimators, given that the correlation is small, i.e., -0.50. However, the performance of the H2S, compared to the 3SE, declines with increases in the level of correlation even if the sample size is large. It is also evident that an increase in the degree of censoring implies higher levels of correlation, hence leading to poor performance of the H2S as compared to the 3S or MLE estimators.

On the other hand, the results depict that the 3S estimator, as would be expected theoretically, is quite robust to the level of correlation and provides results which are quite close to the ML estimates in almost all cases. It is also important to note that, as shown in Chapter 6, the 3SE appears to be robust if the assumption of normality about the error term does not hold (e.g., see results under the students'-t distribution).

In general, the 3S estimator outperforms the H2S estimator and the gap between the two estimators can be substantial even for moderate levels of correlation. If the correlation is very high, the H2S can be even less preferable than the biased OLSP estimator in terms of the RMSE criteria. The evidence also indicates that hypothesis tests and confidence intervals based on the H2S estimates can be misleading if the correlation between the explanatory variables and the estimated inverse of Mill's ratio is large. Specifically, the experimental results reveal that the confidence intervals based on the H2S estimates are likely to be wider (less precise) than they should be if the correlation is not small. Clearly, all the estimators perform quite well in terms of bias except for the OLSP estimator. Not surprisingly, the OLSP estimator performs poorly in general.

Finally, it is quite clear that the 3S estimator outperforms the H2S estimator in almost all cases. However, the most important point that needs to be stressed here is that whether the three-step estimator could be extended to more general models, where estimation by using MLE is not straightforward. Such models include the two-equation tobit models discussed in Manning, Duan and Rogers (1987), Leung and Yu (1994) or Type-II tobit models in general (see Amemiya (1985), Maddala (1983)).