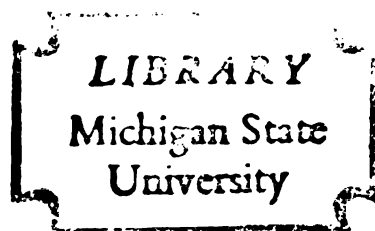


THESIS



This is to certify that the

dissertation entitled

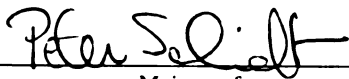
ROBUSTNESS OF THE TOBIT ESTIMATORS TO
HETEROSKEDASTICITY AND NON-NORMALITY

presented by

Abbas Arabmazar

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics


Major professor

Date November, 1981

You who were
a strong follower of
Islam and Khomeini, why
you did not write/choose a subject
in your "Islamic Economic
You know Why?"



OVERDUE FINES:

25¢ per day per item

RETURNING LIBRARY MATERIALS:

Place in book return to remove
charge from circulation records

ROBUSTNESS OF THE TOBIT ESTIMATORS TO
HETEROSKEDASTICITY AND NON-NORMALITY

By

Abbas Arabmazar

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1981

ABSTRACT

ROBUSTNESS OF THE TOBIT ESTIMATORS TO HETEROSKEDASTICITY AND NON-NORMALITY

By

Abbas Arabmazar

9/1/73

It is known that the estimates of the parameters in the Tobit Model and other limited, Truncated and Censored dependent variable models are not robust against the misspecification of the model. The inconsistency of the censored Tobit estimator when the errors are heteroskedastic is shown. To investigate the severity of this inconsistency, a simple model of a constant-term-only regression is utilized, and the value of the asymptotic bias (inconsistency) is calculated for a variety of parameter values. Using the Lagrangian multiplier test principle, a test of heteroskedasticity is derived for both the truncated and censored Tobit models. It is also shown that the usual Tobit MLE which assumes normality is inconsistent when the disturbances are in fact non-normal. For the simple Tobit model with a constant term as the only regressor, the asymptotic bias of the normal MLE is calculated for a variety of non-normal errors.

To My Parents,
and my brothers,
Rasool, Ali, and Amir

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my committee chairperson Professor Peter Schmidt. It was his constant attention, helpful comments, editorial skill, and general encouragement that made completion of this study possible.

I also wish to thank other members of my dissertation committee, William Quinn, John Goddeeris, and James Johannes.

In addition, I would like to acknowledge the generous moral and financial support of members of my family throughout these years. Finally, I owe thanks to my typist, Mrs. Nancy Heath.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
 Chapter	
I. INTRODUCTION	1
1.1 Review of Literature and Statement of the Problem	1
II. ROBUSTNESS TO HETEROSKEDASTICITY	10
2.1 Introduction	10
2.2 Derivation of the Inconsistency	13
2.3 Calculation of the Inconsistency	18
2.4 Conclusions	21
III. TEST FOR HETEROSKEDASTICITY	25
3.1 Introduction	25
3.2 The Model and the Test Statistic	28
3.3 The Statistic for the Truncated Case	30
3.4 The Statistic for the Censored Case	34
3.5 Conclusions	37
IV. ROBUSTNESS TO NON-NORMALITY	38
4.1 Introduction	38
4.2 The Model and Its Estimators	41
4.3 Derivation of the Inconsistency	44
4.4 Calculations	47
4.5 Extension to the Regression Case	50
4.6 Derivation of the Inconsistency and Calculations	52
4.7 Conclusions	56

Chapter		Page
Appendix A	73
Appendix B	86
V. CONCLUSIONS	93
BIBLIOGRAPHY	97

LIST OF TABLES

Table		Page
2.1	Probability limits of estimates of μ and σ $r_1 = r_2 = 0.5$	22
2.2	Probability limits of estimates of μ and σ , $r_1 = .2$, $r_2 = .8$	23
2.3	Probability limits of estimates of μ and σ , $r_1 = .8$, $r_2 = .2$ (variance known) .	24
4.1	Truncated-Moments of selected zero-mean (Symmetric) distributions (variance known).	58
4.2	Asymptotic biases-- t_5	59
4.3	Asymptotic biases-- t_{10}	60
4.4	Asymptotic biases-- t_{20} (variance known) .	61
4.5	Asymptotic biases--Laplace	62
4.6	Asymptotic biases--Logistic	63
4.7	Asymptotic biases of regression coeffi- cients, Censored - t_5 ($p = .5$)	64
4.8	Asymptotic biases of regression coeffi- cients, Censored - t_5 ($p = .5$)	65
4.9	Asymptotic biases of regression coeffi- cients, Censored - t_5	66

LIST OF FIGURES

Figure	Page
4.1 Asymptotic Bias of Estimates of μ , t_{10} Distribution	67
4.2 Asymptotic Bias of Estimates of μ , t_{10} Distribution	68
4.3 Asymptotic Bias of Estimates of μ , Censored Sample	69
4.4 Asymptotic Bias of Estimates of μ , Truncated Sample	70
4.5 Asymptotic Bias of Estimates of μ , Binary Sample	71
4.6 Asymptotic Bias of Estimates of μ , Logistic Distribution	72

CHAPTER I

INTRODUCTION

1.1 Reivew of Literature and Statement of the Problem

The estimation of models with qualitative and limited dependent variables has received considerable theoretical and empirical attention. The appealing feature of these models is that they allow one to answer a variety of questions based on an incomplete sample, which could not previously be answered. Although our study is primarily concerned with what is known in the econometric literature as the Tobit model, it should be noted that similar considerations apply for a wide range of models with qualitative and limited dependent variables which might loosely be grouped under the heading of "sample selection models."

In the case of the ordinary regression model, there exists a substantial literature on the violation of basic assumptions, such as heteroskedasticity, non-normality, autocorrelation, etc., whereas there is relatively little corresponding analysis for the case of limited dependent variable models. A selection of papers which apply the Tobit model to economic problems includes

Tobin (1958), Cragg (1971), Gronau (1973, 1974), Heckman (1974, 1976), Lee and Trost (1978), Hausman and Wise (1977), Nelson (1977). Other related papers are MacFadden (1974) and Schmidt and Strauss (1975).

Before I formally define the model, it is necessary to define two statistical terms, which are used throughout the study. The term censored is applied to a sample in which some observations are recorded only as below (or above) some threshold, the exact value in such a case not being observed (having been censored). The term truncated is applied to the samples in which such observations, i.e., below some threshold, are excluded entirely. Note that in the econometric literature, however, the term truncated is often applied to the censored sample case, apparently in reference to the variable rather than sample.

Censored Tobit model.--Consider the regression model defined by

$$\begin{aligned} y_i &= x_i \beta + \epsilon_i && \text{if RHS} > 0 \\ &= 0 && \text{if RHS} \leq 0 \quad (i = 1, 2, \dots, T) \end{aligned} \tag{1.1.1}$$

where β is a k -component column vector of unknown parameters, x_i is a k -component row vector of known constants, and the ϵ_i 's are independent with distribution $N(0, \sigma^2)$. This model was first suggested by Tobin (1958). Amemiya

(1973) points out that any non-zero known constant can be considered as a censoring point of the model, without much extra complication, so the zero censoring point is not as restrictive as it may seem.

Truncated Tobit model.--Consider the regression model defined by

$$y_i = x_i \beta + \varepsilon_i \quad \text{if and only if } \text{RHS} > 0 \quad (1.1.2)$$

where all the variables are defined as above. Note that the observations on y are truncated to the left of zero.

The Tobit model is a special case of the sample selection model of Heckman (1976), while the Probit model (binary dependent variable) is a special case of the Tobit model. See Goldberger (1980).

It is well known that ordinary least squares will produce biased and inconsistent estimates of the regression parameters in the Tobit model. Maximum likelihood estimation is being used with increasing frequency to avoid this inconsistency. There are other procedures suggested by Heckman (1976, 1979) and Amemiya (1973). The strong consistency and asymptotic normality of the maximum likelihood estimators of the regression parameters in the Tobit model has been proved by Amemiya (1973).

In this study we investigate the robustness of Tobit Normal maximum likelihood estimators to two types of misspecification; heteroskedasticity and non-normality.

From what is known so far, it seems that the assumptions required of these models are quite strong, and any violation, such as heteroskedasticity or non-normality, may result in an asymptotic bias as severe as in the naive ordinary least square (OLS) formulation. This is in contrast with OLS estimators in ordinary regression models, which are consistent when disturbances are heteroskedastic or the assumption of normality of the error terms is violated.

Maddala and Nelson (1975) show that the Tobit estimators are inconsistent when the model is misspecified to be homoskedastic. This is in unfavorable contrast with complete sample least square estimators which are consistent under heteroskedasticity. Hurd (1979) proves that the MLE's of the truncated Tobit model are inconsistent under heteroskedasticity in the simple case in which the only right-hand variable is the constant; i.e., estimating the mean of a truncated normal random variable. He introduces heteroskedasticity into the model by assuming that one subset of the observations has a disturbance variance different from the rest. In other words, the sample comes from two normal populations, with equal location parameters

and different scale parameters, with equal probabilities. Taking the probability limits of the first order conditions, he is able to get an implicit form of the asymptotic bias or inconsistency. By numerically solving for the bias, he comes up with answers to the obvious questions of the direction and severity of the bias. Using different sets of parameter values, he concludes that the bias is substantial even when the heteroskedasticity is in the range to be expected in empirical work. He also finds that the sign of the asymptotic bias is not generally known; and he states that "it would be surprising" if these results did not generally hold for the censored Tobit model. The robustness of MLE's of the Tobit model is also considered by Nelson (1979) and Maddala (1979).

Chapter II extends Hurd's analysis to the censored Tobit model. The implicit form of the inconsistency of the MLE's in the censored case is derived, and it is calculated for a variety of parameter values. This turns out to make a surprisingly large difference. The robustness of the MLE's to heteroskedasticity is much greater in the censored case than in the truncated case.

Chapter III contains a simple test for heteroskedasticity based on the Lagrange multiplier (LM) test, which is also known as Rao's efficient score test. In the Tobit model defined in (1.1.1) or (1.1.2), we specify the nature

of heteroskedasticity as $\sigma_i^2 = z_i \alpha$, where α is a p -component column vector unrelated to the coefficients β , and z_i is a p -component row vector (with first element unity). In other words, the variance of disturbances is a linear function of a set of exogenous variables (elements of z_i). This allows the null hypothesis of homoskedasticity to be

$$H_0: \alpha_2 = \dots = \alpha_p = 0$$

for then $z_i \alpha = \alpha_1 = \sigma^2$ is constant. The LM test statistic is obtained from the result of maximizing the likelihood subject to the parameter constraints implied by the null hypothesis and can be computed from the Lagrangian multiplier corresponding to the constraints or from the first order conditions, as in Rao (1973). Silvey (1959) showed that the LM test has the same asymptotic power as the likelihood ratio (LR) test. For more on LM tests see, for example, Silvey (1959), Rao (1973), and Breusch and Pagan (1979, 1980). Breusch and Pagan pointed out several applications of LM test to model specification in Econometrics and found that in many instances the LM statistic can be computed by a regression using the residuals of the fitted model. Lee (1981b) also suggests a LM test for misspecifications in sample selection models, based on the Pearson family of distributions.

The robustness of the Tobit estimator to non-normality is also a potentially important point since

there is typically not any compelling reason to believe that the disturbances are normal. The MLE's based on the normality assumption are inconsistent when the disturbances are non-normal; see Goldberger (1980) or Nelson (1981). This is not true for least square estimators in the standard linear regression model, which are unbiased and consistent under violation of the distributional assumptions. An obvious question here is the severity of the inconsistency under different conditions.

Goldberger (1980) considers a simple case of estimation of the mean of a truncated random variable, and analyzes the asymptotic bias caused by violation of the normality assumption. He considers some symmetric distributions (Student, Logistic, Laplace) to be the true distributions that were misspecified as normal, and numerically calculates the inconsistency or asymptotic bias of the estimator of the mean of each distribution, given that their variances are known. He finds, not surprisingly, that the bias is negligible when the truncation is mild, and it is substantial when the truncation is extreme (less than 15 percent of population is retained in selected population). He concludes that the size of the bias for moderate degrees of truncation is unexpected.

An obvious extension of Goldberger's work is to consider the non-normality bias of the censored Tobit model. It is also interesting to know what would be the

effect of estimating the variance on the size of the bias of the estimate of the mean, both in the truncated and the censored case; that is, it is worth knowing whether the assumption of known variance matters.

Chapter IV considers the non-normality bias and its severity in the Tobit model, both truncated and censored, while relaxing the assumption of known disturbance variance. The implicit form of the inconsistency is derived. The numerical values of the bias for different symmetric distributions, under a variety of parameter values, is calculated. Chapter IV also contains a cross-distributional comparison and a comparison of different estimators for each distribution. To generalize the results, the Probit model (binary dependent variable model) has also been considered in this analysis. Appendix A includes the derivation of the first and second truncated moments of some selected symmetric distributions, which are necessary to evaluate the probability limits of the first order conditions. Furthermore, the analysis is extended to a regression with one dummy explanatory variable, and the robustness of estimates of the slope coefficients to violations of the normality assumption is investigated.

There are other works on model specification and related topics in Tobit model. For a general test of

misspecification in the censored Tobit model, see Nelson (1981). His test for a univariate censored normal model is based on the general specification test principle suggested by Hausman (1978). Lee (1981c) introduces a general LM test for selectivity bias, homoskedasticity and normality, based on the Pearson family of distributions. For other specification error tests in limited dependent variable models, see Olsen (1979) and Maddala (1979).

CHAPTER II

ROBUSTNESS TO HETEROSKEDASTICITY

2.1 Introduction

One of the basic assumptions of the regression model is that the disturbances have a common constant variance. This is known as homoskedasticity and the violation of this assumption is known as heteroskedasticity. In many econometric studies, especially those based on cross-section data, the assumption of a constant variance for the disturbance term is unrealistic. In consumer budget studies the variance of the error term very likely increases with household income; see Prais and Houthakker (1955). Likewise in cross-section studies of the firm the disturbance variance probably increases with the size of the firm. Heteroskedasticity also naturally arises (1) when the observations are based on average data, and (2) in a number of "random coefficient" models; see Judge et al. (1980).

In the case of the ordinary regression model, there exists a substantial literature on the nature of heteroskedastic disturbances, their consequences and effects on inference and test procedures, and remedies for the problem.

It is well known that the consequences of heteroskedasticity are two-fold. The ordinary least square estimates of the regression parameters are still unbiased and consistent, but are inefficient, and the estimates of the variances are biased. This will effect tests of hypotheses and inferences on the regression parameters. In most studies, a simple parametric form of heteroskedasticity has been assumed to simplify the problem. In other words, it is assumed that the variance of disturbances is a pre-specified function of one or several independent variables. For an up-to-date review of literature, see Judge et al. (1980).

For limited dependent variable models, there has been relatively little work on the problem of heteroskedastic disturbances. Maddala and Nelson (1975) showed that in limited dependent variable models, contrary to the complete sample regression case, the estimators are not even consistent when the error terms are heteroskedastic. To illustrate the nature of the problem, following Maddala and Nelson, consider a censored Tobit model, as defined in equation (1.1.1). The locus of expected values of y_i is given by

$$E(y_i) = (x_i \beta) \Phi(x_i \beta / \sigma) + \sigma \phi(x_i \beta / \sigma) \quad (2.1.1)$$

where ϕ and Φ are respectively the standard normal density and cdf. Now suppose that the "true" model is

heteroskedastic with parameters σ_{oi} and β_o and designate the "true" locus as $E_o(y_i)$. Then

$$E_o(y_i) = (x_i \beta_o) \phi(x_i \beta_o / \sigma_{oi}) + \sigma_{oi} \phi(x_i \beta_o / \sigma_{oi}) \quad (2.1.2)$$

It is immediately obvious that the presence of the variance term in the expected value locus is the source of the difference in the two expressions and, in turn, of the estimation bias.

The main questions to be answered are (1) what are the consequences of heteroskedasticity; (2) how can we detect it and test for it; and (3) what to do to correct for it.

Hurd (1979) has presented some evidence on these points, for the truncated Tobit model, as defined in equation (1.1.2). For the simplified version of the model, assuming that the constant term is the only regressor, he proved that the maximum likelihood estimator is inconsistent when the error terms are heteroskedastic, and derived an implicit form of the inconsistency of the estimate of the mean of a normal random variable. His assumption on the nature of heteroskedasticity is that the sample comes from a mixture of two normal populations with equal means but unequal variances. Assuming equal proportions for each population, he calculated the

asymptotic bias (or inconsistency) of the estimated mean for different sets of parameter values. He concludes that, as one might expect, the bias depends on the degree of heteroskedasticity, the scale of the problem, and the amount of truncation. The bias increases with increase in the degree of truncation as well as increase in the degree of heteroskedasticity. He argues that the bias is extreme for what seem like modest amounts of heteroskedasticity, and so it may be a serious empirical problem.

In the rest of this chapter, Hurd's analysis is extended to the censored Tobit model, and the asymptotic bias is calculated for a variety of parameter values. The results are surprisingly different than Hurd's. In particular, the results for the censored case are more optimistic than for the truncated case. Heteroskedasticity of given severity causes less inconsistency in the former than latter. In the case considered here, moderate amounts of heteroskedasticity do not cause serious inconsistency, except when the fraction of non-limit observations is very low.

2.2 Derivation of the Inconsistency

In this section we derive (implicit) expressions for the inconsistency of the censored Tobit estimator under a simple form of heteroskedasticity. To keep things

tractable, we consider the case of estimation of only the mean and variance of a censored normal; that is, the only explanatory variable is a constant term.

In this case, the assumption underlying the censored Tobit model is that there is a random sample of size T on an unobservable y which is distributed as $N(\mu, \sigma^2)$, while we observe instead $y = \max(0, y^*)$. Thus negative y^* 's are simply observed as zeroes. This contrasts with the case considered by Hurd, in which only positive y 's were observed and the fraction of limit observations is unknown.

Let the number of positive y 's be n , and index these as y_1, \dots, y_n . (Thus there are $T-n$ values of zero for y , in the sample of size T .) The censored Tobit log likelihood function is

$$L = \text{constant} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 + (T-n) \ln \phi(-\mu/\sigma) \quad (2.2.1)$$

where ϕ is the cdf of the unit normal distribution. If we let $\hat{\mu}$ and $\hat{\sigma}$ denote the MLE's, they satisfy the first-order conditions for maximization of L , which can be written as:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}) - \frac{T-n}{n} \hat{\sigma} m(-\hat{\mu}/\hat{\sigma}) = 0 \quad (2.2.2a)$$

$$-\hat{\sigma}^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 + \frac{T-n}{n} \hat{\mu} \hat{\sigma} m(-\hat{\mu}/\hat{\sigma}) = 0 \quad (2.2.2B)$$

Here $m(t)$ is defined as $\phi(t)/\Phi(t)$, where ϕ is the density of the unit normal distribution and Φ is (as above) the corresponding cdf.

We now wish to take probability limits of (2.2.2) under heteroskedasticity. Following Hurd, we assume that there are two distinct variances, σ_1^2 and σ_2^2 . These are sampled in proportions r_1 and $r_2 = 1-r_1$. That is, with probability r_1 we sample y^* from $N(\mu, \sigma_1^2)$ while with probability r_2 , we sample y^* from $N(\mu, \sigma_2^2)$; then we observe $y = \max(0, y^*)$. Thus of the total of T observations, we suppose that for T_1 of these the underlying variable has variance σ_1^2 while for T_2 it has variance σ_2^2 , where as $T \rightarrow \infty$, $T_1/T \rightarrow r_1$ and $T_2/T \rightarrow r_2$. Correspondingly, of the n positive observations, let n_1 and n_2 represent the numbers of observations with underlying variances σ_1^2 and σ_2^2 , respectively. Then we can note the facts:

$$\frac{n}{T} = \frac{n_1}{T_1} \frac{T_1}{T} + \frac{n_2}{T_2} \frac{T_2}{T} \rightarrow r_1 \phi(t_1) + r_2 \phi(t_2) \quad (2.2.3a)$$

$$\frac{T-n}{n} \rightarrow \frac{r_1 \phi(-t_1) + r_2 \phi(-t_2)}{r_1 \phi(t_1) + r_2 \phi(t_2)} \quad (2.2.3b)$$

$$\frac{n_i}{n} = \frac{n_i}{T_i} \frac{T_i}{T} / \left(\frac{n}{T}\right) + \frac{r_i \phi(t_i)}{r_1 \phi(t_1) + r_2 \phi(t_2)} \equiv s_i, \quad i=1, 2 \quad (2.2.3c)$$

$$\frac{1}{n_j} \sum_i^{n_j} y_i \rightarrow \mu + \sigma_j m(t_j), \quad j = 1, 2 \quad (2.2.3d)$$

$$\frac{1}{n_j} \sum_i^{n_j} (y_i - \mu)^2 \rightarrow \sigma_j^2 - \mu \sigma_j m(t_j), \quad j = 1, 2 \quad (2.2.3e)$$

In all of the above, $t_1 = \mu/\sigma_1$, $t_2 = \mu/\sigma_2$, and all limits are taken as T (hence T_1 , T_2 , n , n_1 , n_2) approaches infinity.

To take probability limits of the first order conditions (2.2.2), we first rewrite them slightly:

$$\begin{aligned} \frac{T_1}{T} \frac{n_1}{T_1} \frac{1}{n_1} \sum_i^{n_1} (y_i - \hat{\mu}) + \frac{T_2}{T} \frac{n_2}{T_2} \frac{1}{n_2} \sum_i^{n_2} (y_i - \hat{\mu}) \\ - \left(\frac{T-n}{T}\right) \hat{\sigma} m(-\hat{t}) = 0 \end{aligned} \quad (2.2.4a)$$

$$\begin{aligned} -\hat{\sigma}^2 + \frac{n_1}{n} \frac{1}{n_1} \sum_i^{n_1} (y_i - \hat{\mu})^2 + \frac{n_2}{n} \frac{1}{n_2} \sum_i^{n_2} (y_i - \hat{\mu})^2 \\ + \left(\frac{T-n}{n}\right) \hat{\mu} \hat{\sigma} m(-\hat{t}) = 0 \end{aligned} \quad (2.2.4b)$$

where $\hat{t} = \hat{\mu}/\hat{\sigma}$. Now we use (2.2.3) to take the probability limits in (2.2.4). Defining $\tilde{\mu} = \text{plim } \hat{\mu}$, $\tilde{\sigma} = \text{plim } \hat{\sigma}$, $\tilde{t} = \tilde{\mu}/\tilde{\sigma}$,

we obtain:

$$\begin{aligned}
& r_1 \phi(t_1) [\mu + \sigma_1 m(t_1) - \tilde{\mu}] \\
& + r_2 \phi(t_2) [\mu + \sigma_2 m(t_2) - \tilde{\mu}] \\
& - [r_1 \phi(-t_1) + r_2 \phi(-t_2)] \tilde{\sigma} m(-\tilde{t}) \\
& = 0 \tag{2.2.5a}
\end{aligned}$$

$$\begin{aligned}
& -\tilde{\sigma}^2 + s_1 \{ \sigma_1^2 - \mu \sigma_1 m(t_1) + 2(\mu - \tilde{\mu}) [\mu + \sigma_1 m(t_1)] \\
& + (\tilde{\mu}^2 - \mu^2) \} + s_2 \{ \sigma_2^2 - \mu \sigma_2 m(t_2) \\
& + 2(\mu - \tilde{\mu}) [\mu + \sigma_2 m(t_2)] + (\tilde{\mu}^2 - \mu^2) \} \\
& + \frac{r_1 \phi(-t_1) + r_2 \phi(-t_2)}{r_1 \phi(t_1) + r_2 \phi(t_2)} \tilde{\mu} \tilde{\sigma} m(-\tilde{t}) = 0 \tag{2.2.5b}
\end{aligned}$$

Equation (2.2.5) implicitly defines the probability limits $\tilde{\mu}$ and $\tilde{\sigma}$ of the MLE's, in terms of the parameters μ , σ_1 , σ_2 , r_1 and r_2 . It is easy to see that when $\sigma_1 = \sigma_2$ (and hence $t_1 = t_2$), the solution $\tilde{\mu} = \mu$ $\tilde{\sigma} = \sigma$ satisfies (2.2.5); the Tobit estimator is consistent under homoskedasticity. It is also not difficult to verify that $\tilde{\mu} = \mu$, $\tilde{\sigma} = \sigma$ does not satisfy (2.2.5) when $\sigma_1 \neq \sigma_2$; the Tobit estimator is inconsistent under heteroskedasticity. Finally, although it is not obvious from (2.2.5), the

inconsistency must depend strongly on the degree of censoring of the sample (which depends on t_1 and t_2). This must be so since in completely uncensored samples, the Tobit estimator becomes least squares, which is robust to heteroskedasticity.

2.3 Calculation of the Inconsistency

For given parameter values (μ , σ_1 , σ_2 , r_1 , r_2) we can calculate the inconsistency of the Tobit MLE's by solving (2.2.5) numerically for $\tilde{\mu}$ and $\tilde{\sigma}$. We have done so, holding $\sigma_2 = 1$ for all cases, for $\mu = 1, 0, -1, -2$, for 15 values of σ_1 between .1 and 10, and for three sets of r_1 and r_2 . These results are given in Tables 2.1 ($r_1 = r_2 = .5$), 2.2 ($r_1 = .2$, $r_2 = 8$), and 2.3 ($r_1 = .8$, $r_2 = .2$). It should be noted that setting $\sigma_2 = 1$ is innocuous, since the results are invariant with respect to scale; for example, the ratio $\tilde{\mu}/\mu$ is the same with $\mu = -1$, $\sigma_1 = 10$, $\sigma_2 = 1$ as with $\mu = -.2$, $\sigma_1 = 2$, $\sigma_2 = .2$.

The solution of (2.2.5) was found using the Newton-Raphson iterative scheme, from arbitrary starting values. Many sets of starting values were tried. In all cases precisely two sets of solutions to (2.2.5) were found, with one set having $\tilde{\sigma}$ negative and the other having $\tilde{\sigma}$ positive. Tables 2.1 to 2.3 give the solution corresponding to positive $\tilde{\sigma}$, for obvious reasons.

We will first discuss the results in Table 2.1, concentrating on the estimate of μ , since under heteroskedasticity it is not entirely clear what consistency of $\hat{\theta}$ would mean. As would be expected, the degree of inconsistency depends heavily on the degree of heteroskedasticity and on the degree of truncation of the sample. For given σ_1 (degree of heteroskedasticity), the inconsistency of $\hat{\mu}$ increases as the degree of truncation increases; that is, as μ decreases. This is entirely as expected, for the reason given at the end of Section 2.2.

For given μ , the inconsistency increases as the degree of heteroskedasticity increases, that is, as σ_1 becomes further from one (the value of σ_2). This is also as expected, though it should be recognized that changing σ_2 for fixed μ also alters the degree of truncation and the scale. This is why, e.g., the results for $\sigma_1 = .1$, $\sigma_2 = 1$ are not identical to those for $\sigma_1 = 10$, $\sigma_2 = 1$.

It should be noted that our results for the Tobit model are much more optimistic than Hurd's results for the related truncated normal model. To pick a few examples: with $\mu = 0$, $\sigma_1 = .5$ we find a bias of $-.02$ whereas in the truncated normal model the bias is -1.47 ; with $\mu = -1$, $\sigma_1 = 2$ we find a bias of $-.31$, whereas Hurd finds a bias of -11.74 . This is a considerable difference. We checked

Hurd's calculations, and were able to reproduce all of his Table 1, except for the entries for $\mu = .1$, $\sigma_1 = .5$ and $\sigma_1 = 2$, where we find probability limits for $\hat{\mu}$ of -6.46 and -12.74 rather than -121.02 and -34.30. Thus, although his most pessimistic cases may be too pessimistic, it still remains that the effects of heteroskedasticity are much more severe in the truncated normal model which he considers than in the censored normal model. One possible explanation is that the inconsistency is caused by the incorrect calculation of the probability of truncation, and this enters every term of the likelihood function for the truncated model, but only some of the terms (those corresponding to the limit observations) of the likelihood function for the censored model. In any case the practical implication is that it is well worth keeping the limit observations, if they are available, to protect against the worst effects of heteroskedasticity. (Presumably there is also an efficiency gain to keeping them.)

Finally, the bias of $\hat{\mu}$ is always of the same sign as μ , for $\mu = -2, -1$, and 1 . The bias of $\hat{\mu}$ is negative for $\mu = 0$, which suggests that it would also be negative for small positive μ , but for larger μ the bias is clearly away from zero.

Tables 2.2 and 2.3 given similar results for $r_1 = .2$, $r_2 = .8$ and for $r_1 = .8$, $r_2 = .2$. The results are quite similar to those found in Table 2.1, and need not be discussed separately.

2.4 Conclusions

We have considered the implications of heteroskedasticity in the censored Tobit model. The specific censored Tobit model considered had only a constant term, and the heteroskedasticity consisted of two distinct variances for the underlying normal variable. Our principal conclusions are: (1) Heteroskedasticity causes inconsistency, the severity of which increases with the severity of the heteroskedasticity and with the extent of truncation of the sample; and (2) Heteroskedasticity causes less inconsistency in the censored Tobit model than in the corresponding truncated model (limit observations missing). While it is dangerous to generalize from the simple model used here to more general cases, our conjecture is that moderate heteroskedasticity (say, variances differing by a factor of two) is not likely to cause substantial inconsistency unless the sample is heavily censored (say, more than half of the observations at the limit). Of course, this optimism must be tempered by the recognition that heteroskedasticity invalidates the usual test statistics, even in complete samples.

TABLE 2.1.--Probability limits of estimates of μ and σ , $r_1 = r_2 = 0.5$

<u>σ</u> <u>1</u>	$\mu = 1.0$			$\mu = 0.0$			$\mu = -1.0$			$\mu = -2.0$		
	<u>μ</u>	<u>σ</u>		<u>μ</u>	<u>σ</u>		<u>μ</u>	<u>σ</u>		<u>μ</u>	<u>σ</u>	
.1	1.02	.66		-.09	.74		-1.62	1.15		-2.47	1.08	
.2	1.02	.68		-.07	.74		-1.62	1.15		-2.47	1.08	
.3	1.02	.70		-.05	.76		-1.62	1.15		-2.47	1.08	
.4	1.02	.73		-.04	.78		-1.55	1.12		-2.47	1.08	
.5	1.01	.76		-.02	.80		-1.40	1.05		-2.47	1.08	
.7	1.00	.85		-.01	.87		-1.12	.94		-2.32	1.04	
.8	1.00	.90		-.00	.91		-1.05	.94		-2.16	.99	
1.0	1.00	1.00		.00	1.00		-1.00	1.00		-2.00	1.00	
1.25	1.00	1.13		-.00	1.13		-1.04	1.16		-2.14	1.21	
1.5	1.01	1.26		-.01	1.28		-1.13	1.37		-2.36	1.49	
2.0	1.03	1.52		-.05	1.60		-1.31	1.83		-2.81	2.10	
2.5	1.05	1.79		-.09	1.94		-1.58	2.31		-3.23	2.71	
3.5	1.10	2.35		-.19	2.64		-2.06	3.32		-4.04	3.95	
5.0	1.18	3.21		-.35	3.72		-2.79	4.84		-5.24	5.81	
10.0	1.42	6.16		-.93	7.39		-5.23	9.96		-9.22	12.04	

TABLE 2.2.--Probability limits of estimates of μ and σ , $r_1 = .2$, $r_2 = .8$

σ_1	$\mu = 1.0$		$\mu = 0.0$		$\mu = -1.0$		$\mu = -2.0$	
	$\underline{\mu}$	$\underline{\sigma}$	$\underline{\mu}$	$\underline{\sigma}$	$\underline{\mu}$	$\underline{\sigma}$	$\underline{\mu}$	$\underline{\sigma}$
.1	1.01	.87	-.04	.91	-1.20	1.05	-2.15	1.03
.2	1.01	.88	-.04	.91	-1.20	1.05	-2.15	1.03
.3	1.01	.88	-.03	.91	-1.20	1.05	-2.15	1.03
.4	1.01	.89	-.02	.92	-1.18	1.04	-2.15	1.03
.5	1.01	.91	-.01	.93	-1.15	1.02	-2.15	1.03
.7	1.00	.94	-.00	.95	-1.06	.99	-2.11	1.02
.8	1.00	.96	-.00	.96	-1.02	.98	-2.07	1.00
1.0	1.00	1.00	.00	1.00	-1.00	1.00	-2.00	1.00
1.25	1.00	1.05	-.00	1.06	-1.03	1.08	-2.13	1.13
1.5	1.01	1.10	-.01	1.12	-1.11	1.19	-2.45	1.37
2.0	1.02	1.22	-.04	1.28	-1.34	1.50	-3.26	1.98
2.5	1.03	1.35	-.08	1.46	-1.62	1.85	-4.11	2.65
3.5	1.06	1.63	-.18	1.86	-2.24	2.61	-5.80	4.02
5.0	1.09	2.09	-.35	2.51	-3.20	3.80	-8.30	6.10
10.0	1.20	3.77	-.99	4.80	-6.50	7.87	-16.64	13.08

TABLE 2.3.--Probability limits of estimates of μ and σ , $r_1 = .8$, $r_2 = .2$

σ_1	$\mu = 1.0$		$\mu = 0.0$		$\mu = -1.0$		$\mu = -2.0$	
	$\bar{\mu}$	$\bar{\sigma}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{\mu}$	$\bar{\sigma}$
.1	1.01	.41	-.10	.48	-2.47	1.33	-3.10	1.19
.2	1.01	.44	-.07	.50	-2.47	1.33	-3.10	1.19
.3	1.01	.49	-.05	.54	-2.44	1.32	-3.10	1.19
.4	1.01	.54	-.03	.58	-2.14	1.19	-3.10	1.18
.5	1.01	.61	-.02	.64	-1.63	.99	-3.08	1.18
.7	1.00	.76	-.01	.77	-1.11	.84	-2.55	1.02
.8	1.00	.84	-.00	.84	-1.04	.87	-2.19	.93
1.0	1.00	1.00	.00	1.00	-1.00	1.00	-2.00	1.00
1.25	1.00	1.20	-.00	1.21	-1.02	1.22	-2.06	1.24
1.5	1.01	1.40	-.01	1.42	-1.06	1.46	-2.15	1.51
2.0	1.02	1.81	-.03	1.85	-1.16	1.96	-2.29	2.05
2.5	1.04	2.21	-.05	2.30	-1.26	2.47	-2.43	2.59
3.5	1.07	3.03	-.10	3.20	-1.45	3.50	-2.70	3.68
5.0	1.13	4.26	-.18	4.56	-1.74	5.04	-3.09	5.31
10.0	1.34	8.37	-.45	9.12	-2.70	10.21	-4.39	10.77

CHAPTER III

TEST FOR HETEROSKEDASTICITY

3.1 Introduction

The robustness of the Tobit estimators to heteroskedasticity was considered in Chapter II. The Tobit estimators are inconsistent when the error terms are heteroskedastic, with the severity of the inconsistency depending on the degree of truncation of the sample and the severity of the heteroskedasticity. The violation of the homoskedasticity assumption in the Tobit model may have serious consequences for inference about the regression coefficients. Therefore, testing for heteroskedasticity in the Tobit model is relevant.

For the complete sample regression case, several methods of testing for heteroskedasticity have been suggested in the literature; for a summary see Judge et al. (1980). One common characteristic of most tests is that they are "constructive" or "nested" in the sense that a specific form of heteroskedasticity is considered as an alternative to the null hypothesis of homoskedasticity. For example, it is generally assumed that the variance of disturbances is (1) a linear function of some exogenous

nonrandom variables, the set of which may or may not have an intersection with the set of explanatory variables of regression; (2) proportional to some power of the mean of the dependent variable; or (3) an exponential function of some exogenous variables (multiplicative heteroskedasticity).

Breusch and Pagan (1979) have suggested a test for heteroskedasticity in the complete sample regression case, using the framework of the Lagrangian multiplier (LM) test. Assuming that the variance of the error terms is a well-behaved function of a linear combination of some exogenous variables, say z_j ($j = 1, \dots, p$), they proved that the LM statistic is one-half of the regression sum of squares from the regression of the standardized squared OLS residuals on the z_j 's. This test statistic is asymptotically distributed as χ^2 with $p-1$ degrees of freedom, when the null hypothesis of homoskedasticity is true. This is equivalent to an F-test of significance of the regression coefficients in the regression of squared residuals on z 's because the denominator of the F statistic will converge in probability to unity.

These usual tests are not applicable to the Tobit model. However, we will use the Lagrangian multiplier test principle to construct a similar test for heteroskedasticity in the Tobit model.

The Lagrangian multiplier test suggested by Aitchinson and Silvey (1958, 1960) and Silvey (1959) is

equivalent to Rao's efficient score test; see Rao (1973). The LM statistic is obtained from the result of maximizing the likelihood function subject to the constraints implied by the null hypothesis, and can be computed from the Lagrangian multipliers corresponding to the constraints as in Aitchison and Silvey (1960), or from the first order conditions as suggested by Rao (1973). The LM test has the same desirable asymptotic properties as the likelihood ratio test, as shown by Silvey (1959). Breusch and Pagan (1980) applied the LM test procedure to several specification problems in Econometrics, and found that in many instances the LM statistic can be computed simply by a regression, using the residuals of the fitted model. The LM test statistic has been derived for a test of normality in the Tobit model by Lee (1981b). Lee (1981c) also uses the LM test procedure to test for specification error in sample selection models, including the joint test of selectivity bias, heteroskedasticity and non-normality based on the Pearson family of distributions.

In this chapter the LM test statistic has been derived to test for heteroskedasticity in the Tobit model, for both the truncated and censored cases. It is assumed that the variance of the disturbances is a linear function of some exogenous variables. For the special case of a constant term as the only regressor, the LM statistic is

expressed as a function of the Tobit residuals, and can be computed using a simple regression.

The organization of the chapter is as follows. The model and a summary of LM tests are presented in Section 3.2. Section 3.3 gives the derivation of the test statistic for the truncated Tobit model. The censored case is covered in Section 3.4. Section 3.5 contains the conclusions.

3.2 The Model and the Test Statistic

Consider the linear model

$$y_i = x_i \beta + \varepsilon_i, \quad (i = 1, \dots, T) \quad (3.2.1)$$

where β is a $(k \times 1)$ vector of coefficients and x_i is a $(1 \times k)$ vector of nonrandom explanatory variables. The ε_i 's are normally and independently distributed with mean zero and variance

$$\sigma_i^2 = z_i \alpha \quad (3.2.2)$$

in which α is a $(p \times 1)$ vector of unrestricted parameters functionally unrelated to the coefficients β , and the first element of z_i , a p -component row vector, is unity. We assume that x_i and z_i are exogenous and obeying the conditions set out in Amemiya (1977). The null hypothesis of homoskedasticity is

$$H_0: \alpha_2 = \alpha_3 = \dots = \alpha_p = 0 \quad (3.2.3)$$

since then

$$\sigma_i^2 = \alpha_1 = \sigma^2$$

is constant.

Let $L(\theta)$ be a log likelihood function depending on a $(k \times 1)$ vector of parameters θ , with $d = \partial L / \partial \theta$ as the first derivative (score) vector and $I = E(\frac{\partial L}{\partial \theta} \times \frac{\partial L}{\partial \theta})$, as the information matrix. The hypothesis to be tested is specified as $q < k$ restrictions on θ

$$H_0: h_j(\theta) = 0, j = 1, \dots, q$$

Setting up the Lagrangian function

$$L(\theta) + \sum_{j=1}^q \lambda_j h_j(\theta)$$

and differentiating with respect to the unknown parameters θ and the Lagrangian multipliers λ_j , we obtain $\tilde{\theta}$ and $\tilde{\lambda}$ as the solutions of the first order conditions

$$\tilde{d} + \tilde{H} \tilde{\lambda} = 0$$

$$h_j(\tilde{\theta}) = 0, j = 1, \dots, q$$

where $\tilde{H} = \{ \frac{\partial h_j}{\partial \theta_i}(\tilde{\theta}) \}$ is a $k \times q$ matrix and $\tilde{\lambda}$ is a $(q \times 1)$

vector of Lagrangian multipliers.

As explained by Breusch and Pagan (1980), the idea underlying the test is that when the null hypothesis is true, the restricted estimates $\tilde{\theta}$ will tend to be near the unrestricted MLE's so that \tilde{d} will be close to the zero vector, as will $\tilde{\lambda}$. This leads to the LM test based on the statistic

$$LM = \tilde{d}' \tilde{I}^{-1} \tilde{d} = \tilde{\lambda}' \tilde{H}' \tilde{I}^{-1} \tilde{H} \tilde{\lambda}$$

where \tilde{I} is information matrix when the null hypothesis is true, evaluated at the restricted estimates $\tilde{\theta}$. The LM statistic is asymptotically distributed as χ^2 with q degrees of freedom. The term $\tilde{d}' \tilde{I}^{-1} \tilde{d}$ is the "score" statistic, Rao (1973), while $\tilde{\lambda}' \tilde{H}' \tilde{I}^{-1} \tilde{H} \tilde{\lambda}$ is called the Lagrangian multiplier statistic, but the two test statistics are identical. We employ the score test form because it is based on the restricted model and the restricted estimates, which are relatively simple to calculate in our case.

3.3 The Statistic for the Truncated Case

Adopting the linear model in (3.2.1) with the heteroskedasticity assumption of (3.2.2), the truncation is introduced into the model by assuming, as in (1.1.2) that we observe y_i if and only if y_i is positive. That is, only the observations with positive values, say $i = 1, \dots, n$, from a possible sample of size T are available for y_i . To

test for the null hypothesis of homoskedasticity, equation (3.2.3), we follow the test procedure explained in Section 3.2. The log likelihood of the sample can be written as:

$$L(\alpha, \beta) = \text{Constant} - \frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^n \sigma_i^{-2}$$

$$(y_i - x_i \beta)^2 - \sum_{i=1}^n \ln \phi(x_i \beta / \sigma_i)$$

where ϕ is cdf of unit normal distribution. The first order conditions are

$$\begin{aligned} d = \frac{\partial L}{\partial (\alpha, \beta)} &= \sum_{i=1}^n (2\sigma_i^4)^{-1} \begin{bmatrix} 0 & z'_i \\ 2\sigma_i^2 x'_i & 0 \end{bmatrix} \begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix} \\ &= \sum_{i=1}^n (2\sigma_i^4)^{-1} K_i \begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix} \end{aligned}$$

where K_i is a $(p+k) \times 2$ nonrandom matrix, and v_{1i} , v_{2i} are random variables, with zero mean conditional on sample inclusion, defined as

$$v_{1i} = \varepsilon_i - \sigma_i m(x_i \beta / \sigma_i)$$

$$v_{2i} = \varepsilon_i^2 - \sigma_i^2 + \sigma_i (x_i \beta) m(x_i \beta / \sigma_i)$$

in which $m(\cdot) = \phi(\cdot)/\Phi(\cdot)$, ϕ is unit normal density and Φ is as above. The information matrix is

$$\begin{aligned}
I &= E \left(\frac{\partial L}{\partial (\alpha, \beta)} \frac{\partial L}{\partial (\alpha, \beta)'} \right) \\
&= \sum_{i=1}^n (4\sigma_i^8)^{-1} K_i \begin{bmatrix} a_i & b_i \\ b_i & c_i \end{bmatrix} K_i'
\end{aligned}$$

where I is a square matrix of dimension $(p+k)$ and

$$\begin{aligned}
a_i &= \sigma_i^2 - \sigma_i m(x_i \beta / \sigma_i) [(x_i \beta) + \sigma_i m(x_i \beta / \sigma_i)] \\
b_i &= \sigma_i m(x_i \beta / \sigma_i) [(x_i \beta)^2 + \sigma_i^2 + (x_i \beta) \sigma_i m(x_i \beta / \sigma_i)] \\
c_i &= 2\sigma_i^4 - (x_i \beta) \sigma_i m(x_i \beta / \sigma_i) [(x_i \beta)^2 + \sigma_i^2 \\
&\quad + (x_i \beta) \sigma_i m(x_i \beta / \sigma_i)]
\end{aligned}$$

The vector of first derivatives (score), evaluated at the restricted MLE's $(\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma})$, and denoted by \tilde{d} , is

$$\tilde{d} = (2\tilde{\sigma}^4)^{-1} \begin{bmatrix} z' \tilde{v}_2 \\ 2\tilde{\sigma}^2 & x' \tilde{v}_1 \end{bmatrix}$$

where Z is a $(n \times p)$ matrix with z_i as its i th row, X is a $(n \times k)$ matrix having x_i as its i th row, and

$$\tilde{v}_j = (\tilde{v}_{j1}, \tilde{v}_{j2}, \dots, \tilde{v}_{jn})'.$$

Note that the first element and the last k elements of \tilde{d} are zeros; only the elements corresponding to

α_i , $i = 2, \dots, p$, are non-zero. The information matrix evaluated at the restricted MLE's is

$$\tilde{I} = (2\tilde{\sigma}^8)^{-1} \begin{bmatrix} Z' CZ & 2\tilde{\sigma}^2 Z' BX \\ 2\tilde{\sigma}^2 X' BZ & 4\tilde{\sigma}^4 X' AX \end{bmatrix}$$

where $C = \text{Diag}(\tilde{c}_i)$, $B = \text{Diag}(\tilde{b}_i)$, and $A = \text{Diag}(\tilde{a}_i)$.

The LM statistic, using the inverse rule for partitioned matrices, is

$$LM = \tilde{V}_2' Z [Z' CZ - Z' BX (X' AX)^{-1} X' BZ]^{-1} Z' \tilde{V}_2 \quad (3.3.1)$$

which is distributed asymptotically as χ^2 with $p-1$ degrees of freedom, given that the null hypothesis of homoskedasticity is true.

Note that the complete sample heteroskedasticity test considered by Breusch and Pagan (1979) is a special case of (3.3.1) in which $B = 0$, $C = 2\tilde{\sigma}^4 I_n$, and $\tilde{v}_{2i} = \tilde{\varepsilon}_i^2 - \tilde{\sigma}^2$. In this case, the test statistic is equal to

$$LM = \frac{1}{2\tilde{\sigma}^4} \tilde{V}_2' Z (Z' Z)^{-1} Z' \tilde{V}_2$$

which is equal to one-half of the regression sum of squares in the regression of $\tilde{\sigma}^{-2} \tilde{\varepsilon}_i^2$ on the z 's.

For the special case of only a constant regressor, for which $x = (1, 1, \dots, 1)'$ is $n \times 1$ and $\tilde{c}_i = c$ for every i , the test statistic can be reduced to

$$LM = \frac{1}{\tilde{c}} \tilde{v}_2' Z (Z'Z)^{-1} Z' \tilde{v}_2$$

which is equal to the regression sum of squares in the regression of $\tilde{c}^{-\frac{1}{2}} \tilde{v}_{2i}$, the standardized Tobit residuals, on the z 's.

3.4 The Statistic for the Censored Case

Using the model in (3.2.1) and the assumption of (3.2.2), we now assume that we observe not y_i but $I_i y_i$, where I_i is an indicator defined as

$$I_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{if } y_i \leq 0 \end{cases}$$

which is equivalent to the model in (1.1.1). The log likelihood function is

$$L(\alpha, \beta) = \text{constant} - \frac{1}{2} \sum_{i=1}^T I_i \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^T I_i \sigma_i^{-2} (y_i - x_i \beta)^2 - \sum_{i=1}^T (1 - I_i) \ln \phi(x_i \beta / \sigma_i)$$

where ϕ is unit normal cdf. Define

$$u_{0i} = I_i - \phi(x_i \beta / \sigma_i) = (I_i - 1) + \phi(-x_i \beta / \sigma_i)$$

$$u_{1i} = I_i \varepsilon_i - \sigma_i \phi(x_i \beta / \sigma_i)$$

$$u_{2i} = I_i \epsilon_i^2 - [\sigma_i^2 \phi(x_i \beta / \sigma_i) - \sigma_i (x_i \beta) \phi(x_i \beta / \sigma_i)]$$

$$w_{1i} = u_{1i} + \sigma_i m(-x_i \beta / \sigma_i) u_{0i}$$

$$w_{2i} = u_{2i} - [\sigma_i^2 + \sigma_i (x_i \beta) m(-x_i \beta / \sigma_i)] u_{0i}$$

where ϕ is the unit normal density, $m(\cdot) = \phi(\cdot)/\sigma(\cdot)$, and w_{1i} , w_{2i} are zero-mean random variables. Then the first order conditions and the information matrix can be written as

$$d = \partial L / \partial (\alpha, \beta) = \sum_{i=1}^T (2\sigma_i^4)^{-1} \begin{bmatrix} 0 & z'_i \\ 2\sigma_i^2 & x'_i & 0 \end{bmatrix} \begin{bmatrix} w_{1i} \\ w_{2i} \end{bmatrix}$$

$$\equiv \sum_{i=1}^T (2\sigma_i^4)^{-1} K_i \begin{bmatrix} w_{1i} \\ w_{2i} \end{bmatrix}$$

$$I = \sum_{i=1}^T (4\sigma_i^8)^{-1} K_i \begin{bmatrix} d_i & g_i \\ g_i & h_i \end{bmatrix} K'_i$$

where K_i is as in Section 3.3, $(w_{1i}, w_{2i})'$ is a random vector, $d_i = E(w_{1i}^2)$, $g_i = E(w_{1i} w_{2i})$ and $h_i = E(w_{2i}^2)$. Let \tilde{d} and \tilde{I} denote the first order conditions and information matrix evaluated with restricted MLE's, $(\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma})$. Then

$$\tilde{d} = (2\tilde{\sigma}^4)^{-1} \begin{bmatrix} z' & \tilde{w}_2 \\ 2\tilde{\sigma}^2 x' & \tilde{w}_1 \end{bmatrix}$$

where Z and X are as in Section 3.3, and $\tilde{w}_j = (w_{j1}, \dots, w_{jn})'$. Note that, as in the truncated case, the first element and the last k elements of \tilde{d} are zeros. Also

$$\tilde{I} = (4\tilde{\sigma}^8)^{-1} \begin{bmatrix} z' HZ & 2\tilde{\sigma}^2 z' GX \\ 2\tilde{\sigma}^2 x' GZ & 4\tilde{\sigma}^4 x' DX \end{bmatrix}$$

where $H = \text{Diag} (\tilde{h}_i)$, $G = \text{Diag} (\tilde{g}_i)$, and $D = \text{Diag} (\tilde{d}_i)$.
The test statistic

$$LM = \tilde{w}_2' Z [Z' HZ - Z' GX (X' DX)^{-1} X' GZ]^{-1} Z' \tilde{w}_2$$

is asymptotically distributed as χ^2 with $p-1$ degrees of freedom.

For the special case of only one constant regressor, $\tilde{h}_i = \tilde{h}$ for every observation. Then the test statistic will reduce to

$$LM = \frac{1}{\tilde{h}} \tilde{w}_2' Z (Z' Z)^{-1} Z' \tilde{w}_2$$

which is equal to the regression sum of squares in the regression of $\tilde{h}^{-1/2} \tilde{w}_{2i}$ on the z 's.

3.5 Conclusions

It is important to be able to test for heteroskedasticity in the Tobit model, because under heteroskedasticity the Tobit estimates are inconsistent and the usual tests are invalid. The Lagrangian multiplier test principle has been adopted, and the test statistic has been derived for the null hypothesis of homoskedasticity (in both the truncated and censored cases), against the alternative that the error variance is a linear function of exogenous variables. The test statistics that result are not too difficult to calculate, and thus should be useful in applied work which uses the Tobit model.

CHAPTER IV

ROBUSTNESS TO NON-NORMALITY

4.1 Introduction

In a standard linear regression model, the least squares estimators are unbiased and consistent even when the assumption of normality of the disturbances is violated. Normality (or some other distributional assumption) is necessary for hypothesis testing in finite samples, but does not affect the mean or probability limit of the least squares estimates.

In the Tobit model, the situation is quite different. The usual MLE which assumes normality (which we will refer to as the "normal MLE") is inconsistent when the disturbances are non-normal. Thus estimates of the Tobit model are not robust to violations of the distributional assumption for the disturbances.

While this chapter is concerned specifically with a special case of the Tobit model, it should be noted that similar considerations apply in a wide variety of models with qualitative and limited dependent variables which might loosely be grouped under the heading of "sample selection models." There has been a proliferation of such

models in recent years, because they allow one to answer questions that could not previously be answered. However, virtually all of these models hinge on normality, or some equally specific distributional assumption, and their robustness has basically not been investigated. This is a potentially important point, since there is typically not any compelling reason to believe that disturbances are normal.

The robustness of estimators of the Tobit model to heteroskedasticity has been considered in Chapter II. The general conclusions were that the MLE's are not consistent under heteroskedasticity, and the bias has a direct relationship with the degree of truncation and the severity of the heteroskedasticity. In Chapter III we developed a computationally simple test for heteroskedasticity in the Tobit model based on the class of Lagrangian multiplier (LM) tests. The LM tests, in general, have the same asymptotic properties and power as likelihood ratio (LR) tests, as proved by Silvey (1959).

As for robustness to non-normality, Goldberger (1980) has considered the truncated version of Tobit model, as defined in (1.1.1), under the assumption that the only regressor is a constant term and that the disturbance variance is known (i.e., estimating the mean of a truncated random variable). He derives the asymptotic bias (inconsistency) of the normal MLE and calculates the numerical

value of the bias for a variety of non-normal errors, by assuming that the true distribution of disturbances is a symmetric distribution (Student, Logistic, Laplace) other than normal. His calculations suggest that bias is small when the truncation is mild and substantial under extreme truncation of the sample. He concludes that the bias for moderate degrees of truncation is unexpectedly large.

This chapter includes an obvious extension of Goldberger's work, namely the non-normality bias in the censored Tobit model. For the sake of tractability, the simple case of only a constant regressor is considered. To provide an insight into whether the assumption of known variance matters, this assumption has been relaxed for both the truncated and censored cases. To generalize the results, the bias in the Probit (binary dependent variable) model has also been investigated.

The organization of this chapter is as follows: Section 4.2 contains the models and their estimators. Section 4.3 includes the derivation of the (implicit form of the) inconsistency for censored, truncated, and binary cases when the variance of the error terms is unknown. The numerical calculations are presented in Section 4.4. The results are tabulated for different distributions in Tables 4.2 to 4.6. Cross-distributional comparisons and

comparisons of different estimators for each distribution can be seen in Figures 4.1 to 4.6.

Section 4.5 contains an extension of the above analysis to the regression case with one dummy explanatory variable. In Section 4.6 the inconsistency of the slope coefficient is derived and numerical calculations of the inconsistency have been illustrated. The results of the regression case are tabulated in Tables 4.7 to 4.9.

Finally, Section 4.7 contains the conclusions. The main conclusions are that (1) the bias is generally less in the censored case than in the truncated case; (2) the assumption of known variance makes a substantial difference in the results; (3) the bias from non-normality can be substantial, and in fact for severely truncated samples can be larger than for the uncorrected least squares estimators. Appendix A includes the derivation of the first and second truncated moments of selected symmetric distributions (Student, Logistic, Laplace) which are necessary to evaluate the bias. A summary of the results in Appendix A is presented in Table 4.1. Appendix B contains the evaluation of some probability limits used in the derivation of the bias in the regression case.

4.2 The Model and Its Estimators

We will concentrate on the special case in which the model contains only a constant term; that is, we are

attempting to estimate the mean of a population. The relevance of this case to more complicated models can be questioned, but it seems to be an obvious starting point, and it allows some results that would not otherwise be possible.

Thus we suppose that we have (in principle) a random sample, y_i^* , $i=1, 2, \dots, T$, from a distribution with mean μ and variance σ^2 . If the y_i^* were all observed, the normal MLE would be the sample average, and it would be robust to non-normality. However, we now consider three alternative assumptions about what is observed.

1. Censored Case. Assume that we observe not y_i^* , but rather $y_i = \max(0, y_i^*)$. Letting the first n y 's be the positive observations, the normal log likelihood is

$$L = \text{constant} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 + (T-n) \ln \Phi(-\mu/\sigma) \quad (4.2.1)$$

where Φ is the $N(0,1)$ cdf. The normal MLE's $\hat{\mu}$, $\hat{\sigma}$ satisfy the first order conditions

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}) - \frac{T-n}{n} \hat{\sigma} m(-\hat{\mu}/\hat{\sigma}) = 0 \quad (4.2.2a)$$

$$-\hat{\sigma}^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 + \frac{T-n}{n} \hat{\mu} \hat{\sigma} m(-\hat{\mu}/\hat{\sigma}) = 0 \quad (4.2.2b)$$

Here $m(\cdot) = \phi(\cdot)/\Phi(\cdot)$, where ϕ is the $N(0, 1)$ density and Φ is as above.

2. Truncated Case. Here we observe y_i ($=y_i^*$) if and only if $y_i^* > 0$. The normal log likelihood is

$$L = \text{constant} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - n \ln \Phi(\mu/\sigma) \quad (4.2.3)$$

so that the first order conditions which yield the normal MLE's are

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}) - \hat{\sigma} m(\hat{\mu}/\hat{\sigma}) = 0 \quad (4.2.4a)$$

$$-\hat{\sigma}^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 + \hat{\mu} \hat{\sigma} m(\hat{\mu}/\hat{\sigma}) = 0 \quad (4.2.4b)$$

3. Binary (Probit) Case. Suppose that we observe

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0. \end{cases} \quad (4.2.5)$$

In this case only μ/σ is identified. Taking $\sigma=1$ as a normalization, the normal log likelihood function is

$$L = n \ln \Phi(\mu) + (T-n) \ln \Phi(-\mu) \quad (4.2.6)$$

so that the first order condition for the normal MLE μ is

$$m(\hat{\mu}) - \frac{T-n}{n} m(-\hat{\mu}) = 0. \quad (4.2.7)$$

For completeness, we also list as possible estimators the sample means

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4.2.8a)$$

$$\bar{y}^* = \frac{1}{T} \sum_{i=1}^T y_i = \frac{n}{T} \bar{y}. \quad (4.2.8b)$$

The first is the mean of the truncated sample (positive observations), and could be used in either the censored or truncated case, while the second is the mean of the censored sample, and could be used in the censored case. Clearly these are estimators which do not attempt to correct for the bias due to censoring or truncation.

4.3 Derivation of the Inconsistency

We now assume that the true distribution is some symmetric distribution other than normal. Let z be a random variable with such a distribution, satisfying $E(z) = 0$, and let f and F denote the density and cdf of z . The variable y^* is assumed to be related to z by $y^* = \mu + bz$ where $b^2 = [\text{var}(z)]^{-1}$. Thus $\varepsilon \equiv bz$ has mean zero and variance one so that comparisons among distributions will not be confused by differences in scale.

To derive the inconsistency of the various normal MLE's, we take the probability limits of the first order conditions, and then solve the resulting equations for the probability limits of the estimators. We note the facts that as $T \rightarrow \infty$,

$$\frac{T-n}{n} \rightarrow \frac{P(y^* \leq 0)}{P(y^* > 0)} = \frac{F(-\mu/b)}{F(\mu/b)} \equiv A \quad (4.3.1a)$$

$$\frac{1}{n} \sum_{i=1}^n y_i \rightarrow E(y^* | y^* > 0) = \mu + bE(z | z > -\mu/b) \quad (4.3.1b)$$

$$\equiv B$$

$$\frac{1}{n} \sum_{i=1}^n y_i^2 \rightarrow E(y^{*2} | y^* > 0) = \mu^2 + b^2 E(z^2 | z > -\mu/b) \quad (4.3.1c)$$

$$+ 2b\mu E(z | z > -\mu/b) \equiv C,$$

where F , A , B , and C are determined by the true distribution. Thus to evaluate the probability limits of the first order conditions, we need the cdf and the first two truncated moments of the distribution of z . These are given in Table 4.1 for the distributions which we use, namely the Laplace, logistic and t distributions.

Let $\tilde{\mu}$ and $\tilde{\sigma}$ represent the probability limits of the MLE's $\hat{\mu}$ and $\hat{\sigma}$. Then taking probability limits of the first order conditions, and using (4.3.1), we obtain the following.

Censored Case

$$\tilde{\mu} + A \tilde{\sigma} m(-\tilde{\mu}/\tilde{\sigma}) - B = 0 \quad (4.3.2a)$$

$$\tilde{\mu}^2 - \tilde{\sigma}^2 + A \tilde{\mu} \tilde{\sigma} m(-\tilde{\mu}/\tilde{\sigma}) - 2B\tilde{\mu} + C = 0 \quad (4.3.2b)$$

Truncated Case

$$\tilde{\mu} + \tilde{\sigma} m(\tilde{\mu}/\tilde{\sigma}) - B = 0 \quad (4.3.3a)$$

$$\tilde{\mu}^2 - \tilde{\sigma}^2 + \tilde{\mu} \tilde{\sigma} m(\tilde{\mu}/\tilde{\sigma}) - 2B\tilde{\mu} + C = 0 \quad (4.3.3b)$$

Binary Case

$$m(\tilde{\mu}) - A m(-\tilde{\mu}) = 0 \quad (4.3.4)$$

(Actually, (4.3.4) can be simplified to $\Phi(\tilde{\mu}) = F(\mu/b)$, but is written as above to maintain uniformity of notation.)

These can then be solved numerically for $\tilde{\mu}$ or $\tilde{\sigma}$. For example, in the censored case we would solve (4.3.2a) and (4.3.2b) for $\tilde{\mu}$ and $\tilde{\sigma}$. This solution will depend on μ and σ and on the form of the distribution chosen. In the case in which σ is assumed to be known, we set $\tilde{\sigma} = \sigma$ and solve (4.3.2a) only for $\tilde{\mu}$ --that is, we ignore (4.3.2b). Similar statements apply to the truncated and binary cases.

The inconsistency of the sample averages given by (4.2.8) can be expressed explicitly. We have

$$\bar{y} \rightarrow E(y^* | y^* > 0) = B \quad (4.3.5a)$$

$$\bar{y}^* \rightarrow P(y^* > 0) E(y^* | y^* > 0) = F(\mu/b)B, \quad (4.3.5b)$$

and the inconsistency is just μ minus this expression.

4.4 Calculations

In this section we report the results of our calculations of the inconsistency of the estimators under non-normality. The distributions considered are t (with 5, 10 and 20 degrees of freedom), Laplace (double exponential), and logistic. In all cases $\sigma = 1$, while μ varies from -3.0 to 3.0.

The results are given in Tables 4.2 to 4.6, with each table representing a different distribution. The first and second columns give μ and $P(y^* > 0)$, the latter being a measure of the degree of truncation or censoring in the population. The next two columns give the asymptotic biases of the censored mean \bar{y}^* and the truncated mean \bar{y} , as defined in (4.2.8) above. The next three columns present the asymptotic bias of the censored, truncated and binary normal MLE's of μ , when σ is known; the next two columns give the same information for the case of unknown σ . Finally, the last two columns present the probability limit $\tilde{\sigma}$ of the estimate of σ , for the truncated and censored cases (with unknown σ , obviously). Our results for the truncated case and known σ correspond

to the results of Goldberger (1980), though with sign changes since he considered upper truncation.

Most of the results are qualitatively similar for all distributions. First, the asymptotic biases of all estimators except the binary estimator disappear as μ gets large (that is, as the sample becomes complete), as they must. Second, it makes a considerable difference whether one knows the disturbance variance. The estimators which assume σ known generally have a much smaller bias than those which also estimate σ . This is not uniformly true (it can't be, since the biases change signs, and thus each equals zero at some point), but for many values of μ the difference is huge. Since σ will in practice never be known, our results are more pessimistic than Goldberger's, which were based on known σ .

Figure 4.1, which plots the biases for the t_{10} case, gives a good visual illustration of the two points just made. It also illustrates a third major conclusion, that the bias of the censored estimator is generally considerably less than that of the estimator from the truncated sample. Again, this is not uniformly so, but for many values of μ the difference is quite considerable. This can be seen perhaps more clearly in Figure 4.2, in which the vertical axis is inflated.

A fourth point concerns the comparison of the normal MLE's with the sample means \bar{y} and \bar{y}^* , which do not

correct for truncation or censoring bias. When σ is known, the bias of the normal MLE is less than the bias of the sample mean, for both truncated and censored samples, for all of the distributions and parameter values we considered. Thus it is better to correct than not to correct, even though the correction is biased, for all cases we considered. (Whether this is true in general is an interesting question.) However, this is not necessarily the case when σ is unknown; then the biased correction can be worse than no correction. This can be seen in tables for both the truncated and censored cases, or in Figure 4.1 for the truncated case.

A comparison across distributions is illustrated in Figure 4.3 for the case of a censored sample with known variance, in Figure 4.4 for a truncated sample with known variance, and in Figure 4.5 for a binary sample. As would be expected, the bias is worse for distributions which are more non-normal (e.g., t_5 or Laplace vs. t_{20}).

Finally, Figure 4.6 illustrates a comparison of the biases in the censored, truncated and binary cases. The binary bias does not go to zero when $\mu \rightarrow \infty$. It is interesting that where any two of the bias curves intersect, all three do; Goldberger has shown that this is so for any distribution (personal communication). For the distributions we consider the censored bias is always

between the binary and truncated biases, though we have not proved this as a general result.

4.5 Extension to the Regression Case

In this section the robustness to non-normality of the Tobit estimator of regression parameters is investigated. For a simple linear regression with one explanatory variable, the asymptotic bias of the slope coefficient and constant term has been derived for both censored and truncated Tobit models under the assumption that the regressor is a dummy variable. Suppose we have (in principle) a random sample of size T on the regression

$$y_i^* = \alpha + \beta x_i + \varepsilon_i$$

where ε_i 's are iid as $N(0, \sigma^2)$ and x_i is a non-random variable with known values for T observations. If the value of y^* were known for all T observations, this would be the ordinary regression case and the normal MLE's (least squares) would be robust to non-normality. Let us now consider two alternative assumptions about what is observed on y_i^* , given that x_i is always observed.

1. Censored Tobit Model. Assume that we observe not y_i^* but rather $y_i = \max(0, y_i^*)$. Assuming that the first n y 's are the positive observations, the normal log likelihood function can be written as

$$\begin{aligned}
L = \text{Constant} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 \\
+ \sum_{i=1}^{T-n} \ln \phi \left(- \frac{\alpha + \beta x_i}{\sigma} \right) \quad (4.5.1)
\end{aligned}$$

where ϕ is the cdf of unit normal. The normal MLE's $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ can be computed from the first order conditions:

$$\frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta} x_i)) - \frac{\hat{\sigma}}{n} \sum_{i=1}^{T-n} m \left(- \frac{\hat{\alpha} + \hat{\beta} x_i}{\hat{\sigma}} \right) = 0 \quad (4.5.2a)$$

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n x_i (y_i - (\hat{\alpha} + \hat{\beta} x_i)) - \frac{\hat{\sigma}}{n} \sum_{i=1}^{T-n} x_i \\
m \left(- \frac{\hat{\alpha} + \hat{\beta} x_i}{\hat{\sigma}} \right) = 0 \quad (4.5.2b)
\end{aligned}$$

$$\begin{aligned}
-\hat{\sigma}^2 + \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta} x_i))^2 + \frac{\hat{\sigma}}{n} \sum_{i=1}^{T-n} (\hat{\alpha} + \hat{\beta} x_i) \\
m \left(- \frac{\hat{\alpha} + \hat{\beta} x_i}{\hat{\sigma}} \right) = 0 \quad (4.5.2c)
\end{aligned}$$

where $m(\cdot) = \phi(\cdot)/\Phi(\cdot)$, ϕ is the unit normal density, and Φ is the unit normal cdf.

2. Truncated Tobit Model. Here we observe

y_i^* ($= y_i$) if and only if $y_i^* > 0$. In other words, only positive values of y_i^* are observed, and therefore used in estimation. The normal log likelihood is

$$\begin{aligned}
L = \text{constant} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 \\
- \sum_{i=1}^n \ln \phi \left(\frac{\alpha + \beta x_i}{\sigma} \right)
\end{aligned} \tag{4.5.3}$$

and the first order conditions which yield the normal MLE's are

$$\frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta} x_i)) - \frac{\hat{\sigma}}{n} \sum_{i=1}^n m \left(\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}} \right) = 0 \tag{4.5.4a}$$

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n x_i (y_i - (\hat{\alpha} + \hat{\beta} x_i)) - \frac{\hat{\sigma}}{n} \sum_{i=1}^n x_i \\
m \left(\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}} \right) = 0
\end{aligned} \tag{4.5.4b}$$

$$\begin{aligned}
-\hat{\sigma}^2 + \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta} x_i))^2 + \frac{\hat{\sigma}}{n} \sum_{i=1}^n (\hat{\alpha} + \hat{\beta} x_i) \\
m \left(\frac{\hat{\alpha} + \hat{\beta} x_i}{\hat{\sigma}} \right) = 0
\end{aligned} \tag{4.5.4c}$$

4.6 Derivation of the Inconsistency and Calculations

Now assume, as before, that the true distribution of disturbances is a symmetric distribution other than normal. Let z be a random variable with such distribution with zero mean, and let f and F denote the density and cdf of z . Furthermore, let bz be the error term, where

$b^2 = [\text{var}(z)]^{-1}$, so that the errors have a variance of one. Thus we have the regression

$$y_i^* = \alpha + \beta x_i + \varepsilon_i$$

where $\varepsilon_i = bz_i$.

In this stage, in order to make the derivations of the bias tractable, we assume that x is a dummy variable which takes the value of unity p percent of the times and therefore, the value of zero $1-p = q$ percent of the times (in the original hypothetical population). Putting it differently, x can be assumed as a Bernoulli random variable such that $P(x = 1) = p$ and $P(x = 0) = q = 1-p$.

Let $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\sigma}$ denote the probability limits of the MLE's $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}$. Then taking the probability limits of the first order conditions in (4.5.2) and (4.5.4), and using the results of Appendix B and rearranging the terms we obtain the following.

Censored Case

$$\tilde{\alpha} + p_1 \tilde{\beta} + \gamma p_2 \tilde{\sigma} m(-\tilde{t}_1) + \gamma q_2 \tilde{\sigma} m(-\tilde{t}_0) - A = 0 \quad (4.6.1a)$$

$$p_1 (\tilde{\alpha} + \tilde{\beta}) + \gamma p_2 \tilde{\sigma} m(-\tilde{t}_1) - p_1 D = 0 \quad (4.6.1b)$$

$$\begin{aligned} & \tilde{\alpha}^2 + p_1 \tilde{\beta}^2 - \tilde{\sigma}^2 + 2 p_1 \tilde{\alpha} \tilde{\beta} + \gamma p_2 (\tilde{\alpha} + \tilde{\beta}) \tilde{\sigma} m(-\tilde{t}_1) \\ & + \gamma q_2 \tilde{\alpha} \tilde{\sigma} m(-\tilde{t}_0) - 2A\tilde{\alpha} - 2 p_1 D \tilde{\beta} + C = 0 \quad (4.6.1c) \end{aligned}$$

where $\tilde{t}_0 = \tilde{\alpha}/\tilde{\sigma}$, $\tilde{t}_1 = (\tilde{\alpha} + \tilde{\beta})/\tilde{\sigma}$ and the values of $A = E(y_1^* | y_1^* > 0)$, $D = E(y_1^* | y_1^* > 0, x_1 = 1)$, $C = E(y^{*2} | y^* > 0)$, $p_1 = (1 - q_1)$, $p_2 = (1 - q_2)$ and γ depend on the distribution of z .

Truncated Case

$$\tilde{\alpha} + p_1 \tilde{\beta} + p_1 \tilde{\sigma} m(\tilde{t}_1) + q_1 \tilde{\sigma} m(\tilde{t}_0) - A = 0 \quad (4.6.2a)$$

$$\tilde{\alpha} + \tilde{\beta} + \tilde{\sigma} m(\tilde{t}_1) - D = 0 \quad (4.6.2b)$$

$$\begin{aligned} & \tilde{\alpha}^2 + p_1 \tilde{\beta}^2 - \tilde{\sigma}^2 + 2p_1 \tilde{\alpha}\tilde{\beta} + p_1 (\tilde{\alpha} + \tilde{\beta}) \tilde{\sigma} m(\tilde{t}_1) \\ & + q_1 \tilde{\alpha}\tilde{\sigma} m(\tilde{t}_0) - 2A\tilde{\alpha} - 2p_1 D\tilde{\beta} + C = 0 \quad (4.6.2c) \end{aligned}$$

Calculations. To illustrate the calculation of the inconsistency of the regression coefficients under non-normality, the Student-t distribution with 5 degrees of freedom is used. Then equations (5.6.1a) and (5.6.1b) are solved for $\tilde{\alpha}$ and $\tilde{\beta}$, assuming the variance is known ($\tilde{\sigma} = 1$). For the unknown variance case (5.6.1) is solved to calculate the values of $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\sigma}$. As mentioned earlier, the simple regressions model with a dummy variable discussed above can be interpreted as a mixture of

two random variables with the expected values $\mu_1 = \alpha$ and $\mu_2 = \alpha + \beta$, where μ_2 is selected p percent of the times. The probability limits of estimates of the two means are: $\tilde{\mu}_1 = \tilde{\alpha}$, $\tilde{\mu}_2 = \tilde{\alpha} + \tilde{\beta}$, and so $\tilde{\beta} = \tilde{\mu}_2 - \tilde{\mu}_1$. The numerical results are presented in Tables 4.7 to 4.9. Table 4.7 provides the asymptotic biases of the estimates of β , $\alpha = \mu_1$ and μ_2 , while moving the means of two random variables in opposite directions from zero. This keeps the degree of truncation more or less constant. The case of moving the means together in the same direction, keeping their distance constant ($=1$), is presented in Table 4.8. Both Tables 4.7 and 4.8 assume $p = .5$. Table 4.9 includes the biases when only the degree of contamination (p) is changed.

Table 4.7 presents the true values and biases of β , μ_2 and $\mu_1 = \alpha$. For the case where σ is known, the results are identical to the non-regression case (constant term as only regressor) considered earlier (e.g., for $\mu_2 = -3$, the bias of estimate of μ_2 is .55 here and in Table 4.2). This is because we are in fact estimating the two means separately. With variance unknown, the results are different from the non-regression case because of the extra restriction (same variance in both samples) which is imposed here. This leads to smaller biases for the estimates of μ_1 , μ_2 and β . In general, the bias of the MLE of β increases as the absolute value of β increases.

Table 4.8 changes the degree of truncation, while keeping the distance between the two means, β , constant. The results suggest that the inconsistency of β , as well as of each mean, increases with an increase in the degree of truncation, but at a smaller rate than the asymptotic bias of separate estimates of each mean (as in Table 4.2).

Table 4.9 presents the biases as p increased (keeping μ_1 , μ_2 and therefore β constant), where p is the proportion of the sample with the larger mean. This may also be viewed as decreasing the degree of truncation, and it will reduce the asymptotic bias of the estimate of the slope coefficient β . It will also reduce the bias of the estimate of $\mu_1 = \alpha$, which is the same type of effect as in Table 4.7, since the means of two samples have been chosen to be symmetric with the equal unknown variance.

4.7 Conclusions

We have considered the asymptotic bias of the normal MLE for censored, truncated and binary samples, when the disturbances are, in fact, non-normal. We state three practical conclusions.

1. The bias for non-normality can be substantial. This is especially true in the realistic case in which the disturbance variance is unknown. In fact, the bias of the normal MLE can be larger than the bias of the uncorrected sample mean.

2. The censored estimator is usually much less biased than the truncated estimator. Therefore, the limit observations should be used if they are available.

3. The bias due to non-normality depends on the degree of censoring (or truncation) of the sample. For example, for samples that are 75 percent complete ($\mu =$ roughly 1.2 above) there is virtually no bias. For samples that are 50 percent complete ($\mu = 0$), the bias is substantial for truncated samples, though not for censored samples. For samples that are largely incomplete, the bias is substantial for all cases.

These results, if general, are of considerable practical importance. For example, they give guidance as to cases in which the approach of Lee (1981a, 1981b), based on a less restrictive distributional assumption than normality, may be worth considering. Therefore, it is important to stress that our results may fail to be general for at least two reasons. First, we have estimated only a simple model, not a general regression. Second, for reasons of tractability we have considered only a limited number of non-normal distributions, all of which are symmetric. Undoubtedly there exist distributions for which one would obtain results that are more pessimistic.

TABLE 4.1.--Truncated-moments of selected zero-mean (Symmetric) distributions

Z	f(z)	F(z)	Var(z)	E(z z > -x)*	E(z ² z > -x)*
Normal	$(2\pi)^{-1/2} e^{-\frac{1}{2}z^2}$	$\int_{-\infty}^z f(t)dt$	1	$\frac{f(x)}{F(x)}$	$1-x \frac{f(x)}{F(x)}$
Student (n)	$h(n) \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)}$	$\int_{-\infty}^z f(t)dt$	$\frac{n}{n-2}$	$\frac{n+x^2}{n-1} \frac{f(x)}{F(x)}$	$\begin{cases} \frac{I_T(\alpha, \beta)}{2F(x)}, & x \leq 0 \\ \frac{2-I_T(\alpha, \beta)}{2F(x)}, & x > 0 \end{cases}$
Logistic	$e^{-z}/(1+e^{-z})^2$	$(1+e^{-z})^{-1}$	$\frac{\pi^2}{3}$	$-x - \frac{\ln(1-F(x))}{F(x)}$	$\begin{cases} x^2 + \frac{2}{F(x)} [x \ln(1-F(x)) - g(-e^x)], & x \leq 0 \\ x^2 + \frac{2}{F(x)} [x \ln(F(x)) + g(-e^{-x})] + \frac{\pi^2}{6} - \frac{x^2}{2}, & x > 0 \end{cases}$
Laplace	$\frac{1}{2} e^{- z }$	$\begin{cases} e^z, & z \leq 0 \\ 1 - \frac{1}{2} e^{-z}, & z > 0 \end{cases}$	2	$\begin{cases} 1-x, & x \leq 0 \\ \frac{1+x}{2e^x-1}, & x > 0 \end{cases}$	$\begin{cases} 1 + (1-x)^2, & x \leq 0 \\ \frac{1 - (1+x)^2}{2e^x-1} + 2, & x > 0 \end{cases}$

*For derivation, see Appendix A.

$$h(n) = \left[\sqrt{n} B\left(\frac{1}{2}, \frac{n}{2}\right) \right]^{-1}, \quad T = \frac{n}{n+x^2}, \quad \alpha = \frac{n-2}{2}, \quad \beta = \frac{3}{2}, \quad B \text{ is beta function, } B_T \text{ is incomplete Beta function.}$$

$$I_T(\alpha, \beta) = B_T(\alpha, \beta) / B(\alpha, \beta), \quad g(t) = \sum_{n=1}^{\infty} (t^n/n^2)$$

TABLE 4.2.--Asymptotic biases-- t_5

μ	Sample Mean				μ (σ known)				μ (σ unknown)			
	$p(y>0)$	Censored	Truncated	Censored	Truncated	Binary	Censored	Truncated	Censored	Truncated	Censored	Truncated
-3.0	.01	3.01	3.92	.55	3.30	.48	-4.64	-18.13	2.03			3.65
-2.8	.01	2.81	3.68	.44	3.01	.38	-4.07	-17.35	1.23			3.44
-2.6	.01	2.61	3.44	.34	2.72	.28	-3.53	-16.58	1.64			3.23
-2.4	.01	2.41	3.21	.25	2.42	.19	-3.03	-15.84	1.45			3.04
-2.2	.02	2.21	2.97	.17	2.12	.11	-2.56	-15.13	1.27			2.85
-2.0	.02	2.02	2.73	.10	1.83	.03	-2.13	-14.45	1.10			2.67
-1.8	.03	1.82	2.51	.03	1.53	-.03	-1.74	-13.82	.93			2.51
-1.6	.05	1.63	2.28	-.02	1.25	-.08	-1.40	-13.23	.78			2.35
-1.4	.07	1.44	2.06	-.06	.97	-.11	-1.09	-12.70	.64			2.22
-1.2	.09	1.25	1.84	-.09	.71	-.14	-.83	-12.22	.51			2.10
-1.0	.13	1.08	1.63	-.10	.47	-.14	-.60	-11.80	.39			2.00
-0.8	.17	.91	1.43	-.10	.26	-.14	-.42	-11.41	.28			1.93
-0.6	.24	.75	1.24	-.10	.09	-.12	-.28	-11.02	.20			1.88
-0.4	.31	.61	1.06	-.08	-.05	-.09	-.18	-10.49	.12			1.83
-0.2	.40	.48	.89	-.06	-.14	-.05	-.10	-9.34	.06			1.74
0.0	.50	.37	.75	-.04	-.18	.00	-.05	-5.47	.02			1.24
0.2	.60	.28	.60	-.02	-.20	.05	-.01	-2.28	-.01			.67
0.4	.69	.21	.48	-.01	-.18	.09	.01	-1.09	-.04			.38
0.6	.76	.15	.38	.01	-.15	.12	.02	-.55	-.05			.21
0.8	.83	.11	.30	.01	-.12	.14	.02	-.27	-.05			.11
1.0	.87	.08	.24	.01	-.08	.14	.02	-.13	-.05			.04
1.2	.91	.06	.19	.01	-.05	.14	.02	-.05	-.05			-.00
1.4	.93	.04	.14	.01	-.03	.11	.02	-.01	-.05			-.03
1.6	.95	.03	.11	.01	-.01	.08	.01	.02	-.04			-.05
1.8	.97	.02	.09	.01	.01	.03	.01	.03	-.04			-.05
2.0	.98	.02	.07	.01	.02	-.03	.01	.03	-.03			-.06
2.2	.98	.01	.06	.01	.02	-.11	.01	.03	-.03			-.06
2.4	.99	.01	.04	.01	.02	-.19	.01	.03	-.03			-.06
2.6	.99	.01	.04	.01	.02	-.28	.01	.03	-.02			-.05
2.8	.99	.01	.03	.00	.02	-.38	.01	.02	-.02			-.05
3.0	.99	.01	.02	.00	.02	-.48	.00	.02	-.02			-.04

TABLE 4.3.--Asymptotic biases-- t_{10}

μ	$p(y>0)$	Sample Mean			μ (σ known)			μ (σ unknown)			σ	
		Censored	Truncated	Binary	Censored	Truncated	Binary	Censored	Truncated	Censored	Truncated	Truncated
-3.0	.00	3.00	3.56		.35	2.15	.32	-1.97	-7.22	.85		1.52
-2.8	.01	2.80	3.35		.27	1.91	.25	-1.70	-7.24	.76		1.48
-2.6	.01	2.60	3.14		.21	1.67	.18	-1.45	-7.19	.68		1.43
-2.4	.01	2.41	2.93		.15	1.45	.13	-1.22	-7.09	.59		1.38
-2.2	.02	2.21	2.73		.10	1.23	.08	-1.02	-7.04	.51		1.34
-2.0	.02	2.01	2.53		.06	1.02	.03	-.83	-6.99	.44		1.31
-1.8	.04	1.82	2.33		.03	.82	.00	-.67	-6.94	.37		1.28
-1.6	.05	1.63	2.13		-.00	.64	-.03	-.53	-6.88	.31		1.26
-1.4	.07	1.44	1.94		-.02	.48	-.05	-.41	-6.76	.25		1.24
-1.2	.10	1.26	1.75		-.03	.34	-.06	-.31	-6.53	.20		1.21
-1.0	.14	1.08	1.57		-.04	.22	-.06	-.23	-6.00	.15		1.15
-0.8	.20	.92	1.40		-.04	.12	-.06	-.16	-4.77	.11		.99
-0.6	.26	.76	1.23		-.04	.04	-.05	-.11	-3.05	.08		.73
-0.4	.33	.62	1.07		-.03	-.01	-.03	-.07	-1.94	.05		.53
-0.2	.41	.50	.91		-.02	-.05	-.02	-.04	-1.21	.02		.37
0.0	.50	.39	.77		-.02	-.07	.00	-.02	-.80	.01		.27
0.2	.59	.30	.65		-.01	-.08	.02	-.01	-.51	-.03		.19
0.4	.67	.22	.53		-.00	-.07	.03	.00	-.31	-.02		.12
0.6	.74	.16	.43		.00	-.06	.05	.01	-.19	-.02		.08
0.8	.80	.12	.34		.01	-.05	.06	.01	-.11	-.02		.04
1.0	.86	.08	.27		.01	-.03	.06	.01	-.05	-.02		.02
1.2	.90	.06	.21		.01	-.02	.06	.01	-.02	-.02		-.00
1.4	.93	.04	.16		.01	-.01	.05	.01	.00	-.02		-.01
1.6	.95	.03	.12		.01	-.00	.03	.01	.01	-.02		-.02
1.8	.96	.02	.09		.01	.01	-.00	.01	.02	-.02		-.03
2.0	.98	.01	.06		.00	.01	-.03	.00	.02	-.01		-.03
2.2	.98	.01	.05		.00	.01	-.08	.00	.02	-.01		-.03
2.4	.99	.01	.03		.00	.01	-.13	.00	.02	-.01		-.03
2.6	.99	.00	.03		.00	.01	-.18	.00	.01	-.01		-.02
2.8	.99	.00	.02		.00	.01	-.25	.00	.01	-.01		-.02
3.0	1.00	.00	.01		.00	.01	-.32	.00	.01	-.00		-.02

TABLE 4.4.--Asymptotic biases-- t_{20}

μ	$P(y^* > 0)$	Sample Mean			μ (σ known)			μ (σ unknown)			σ	
		Censored	Truncated	Censored	Truncated	Binary	Censored	Truncated	Censored	Truncated	Censored	Truncated
-3.0	.00	3.00	3.41	.20	1.30	.19	-.96	-4.20	.41		.41	.82
-2.8	.00	2.80	3.21	.16	1.13	.14	-.81	-4.12	.36		.36	.79
-2.6	.01	2.60	3.02	.12	.97	.11	-.68	-4.24	.32		.32	.79
-2.4	.01	2.40	2.83	.08	.81	.07	-.57	-4.04	.27		.27	.76
-2.2	.02	2.21	2.63	.06	.67	.04	-.47	-4.06	.24		.24	.76
-2.0	.02	2.01	2.45	.03	.54	.02	-.38	-3.99	.20		.20	.75
-1.8	.04	1.82	2.26	.02	.43	.00	-.30	-3.90	.17		.17	.74
-1.6	.05	1.63	2.08	.00	.33	-.01	-.24	-3.62	.14		.14	.71
-1.4	.08	1.44	1.89	-.01	.24	-.02	-.18	-2.83	.11		.11	.60
-1.2	.11	1.26	1.72	-.02	.17	-.03	-.14	-2.10	.09		.09	.48
-1.0	.15	1.08	1.55	-.02	.11	-.03	-.10	-1.49	.07		.07	.38
-0.8	.20	.92	1.38	-.02	.06	-.03	-.07	-1.05	.05		.05	.29
-0.6	.27	.77	1.22	-.02	.02	-.02	-.05	-.79	.03		.03	.24
-0.4	.34	.63	1.07	-.01	-.01	-.02	-.03	-.54	.02		.02	.18
-0.2	.42	.50	.92	-.01	-.02	-.01	-.02	-.36	.01		.01	.13
0.0	.50	.39	.79	-.01	-.03	.00	-.01	-.29	.00		.00	.11
0.2	.58	.30	.66	-.00	-.03	.01	-.00	-.21	.00		.00	.08
0.4	.66	.23	.55	-.00	-.03	.02	.00	-.13	.01		.01	.06
0.6	.73	.17	.45	.00	-.03	.02	.00	-.08	.01		.01	.03
0.8	.80	.12	.36	.00	-.02	.03	.00	-.05	.01		.01	.02
1.0	.85	.08	.28	.00	-.02	.03	.01	-.02	.01		.01	.01
1.2	.89	.06	.21	.00	-.01	.03	.00	-.01	.01		.01	.00
1.4	.92	.04	.16	.00	-.00	.02	.00	.00	.01		.01	.01
1.6	.95	.03	.12	.00	.00	.01	.00	.01	.01		.01	.01
1.8	.96	.02	.09	.00	.00	-.00	.00	.01	.01		.01	.01
2.0	.98	.01	.06	.00	.01	-.02	.00	.01	.01		.01	.01
2.2	.98	.01	.04	.00	.01	-.04	.00	.01	.00		.00	.01
2.4	.99	.00	.03	.00	.01	-.07	.00	.01	.00		.00	.01
2.6	.99	.00	.02	.00	.01	-.11	.00	.01	.00		.00	.01
2.8	1.00	.00	.01	.00	.01	-.14	.00	.01	.00		.00	.01
3.0	1.00	.00	.01	.00	.00	-.19	.00	.00	.00		.00	.01

TABLE 4.5.--Asymptotic biases--Laplace

μ	$P(y^* > 0)$	Sample Mean			μ (σ known)			μ (σ unknown)			σ	
		Censored	Truncated	Binary	Censored	Truncated	Binary	Censored	Truncated	Censored	Truncated	
-3.0	.01	3.01	3.71	.60	.273	.55	-2.36	-7.99	1.19	1.96	1.96	
-2.8	.01	2.81	3.51	.50	.253	.46	-2.19	-8.19	1.13	1.96	1.96	
-2.6	.01	2.61	3.31	.41	.233	.36	-2.02	-8.39	1.07	1.96	1.96	
-2.4	.02	2.41	3.11	.33	.213	.28	-1.86	-8.59	1.00	1.96	1.96	
-2.2	.02	2.22	2.91	.25	.193	.19	-1.70	-8.79	.94	1.96	1.96	
-2.0	.03	2.02	2.71	.17	.173	.11	-1.53	-8.99	.87	1.96	1.96	
-1.8	.04	1.83	2.51	.10	.153	.04	-1.37	-9.19	.80	1.96	1.96	
-1.6	.05	1.64	2.31	.04	.133	-.03	-1.22	-9.39	.73	1.96	1.96	
-1.4	.07	1.45	2.11	-.02	.113	-.08	-1.06	-9.59	.66	1.96	1.96	
-1.2	.09	1.27	1.91	-.07	.093	-.13	-.91	-9.79	.58	1.96	1.96	
-1.0	.12	1.09	1.71	-.10	.073	-.17	-.76	-9.99	.50	1.96	1.96	
-0.8	.16	.91	1.51	-.13	.053	-.19	-.61	-10.19	.41	1.96	1.96	
-0.6	.21	.75	1.31	-.14	.033	-.19	-.47	-10.39	.33	1.96	1.96	
-0.4	.28	.60	1.11	-.13	.013	-.17	-.33	-10.59	.23	1.96	1.96	
-0.2	.38	.47	.91	-.11	-.07	-.11	-.20	-10.79	.13	1.96	1.96	
0.0	.50	.35	.71	-.06	-.27	.00	-.07	-10.99	.02	1.96	1.96	
0.2	.62	.27	.55	-.01	-.34	.11	.01	-8.69	-.05	1.72	1.72	
0.4	.72	.20	.44	.01	-.29	.17	.03	-2.30	-.07	.68	.68	
0.6	.79	.15	.36	.02	-.21	.19	.03	-.78	-.08	.28	.28	
0.8	.84	.11	.29	.02	-.14	.19	.03	-.29	-.08	.10	.10	
1.0	.88	.09	.24	.02	-.08	.17	.03	-.09	-.07	.01	.01	
1.2	.91	.07	.19	.02	-.04	.13	.02	-.01	-.06	-.04	-.04	
1.4	.93	.05	.16	.02	-.01	.08	.02	.03	-.05	-.07	-.07	
1.6	.95	.04	.13	.02	.01	.03	.02	.05	-.05	-.07	-.07	
1.8	.96	.03	.10	.01	.01	-.04	.01	.05	-.04	-.08	-.08	
2.0	.97	.02	.08	.01	.03	-.11	.01	.05	-.03	-.08	-.08	
2.2	.98	.02	.07	.01	.03	-.19	.01	.05	-.03	-.07	-.07	
2.4	.98	.01	.05	.01	.03	-.28	.01	.04	-.02	-.07	-.07	
2.6	.99	.01	.04	.01	.03	-.36	.01	.04	-.02	-.06	-.06	
2.8	.99	.01	.03	.00	.03	-.46	.00	.03	-.02	-.05	-.05	
3.0	.99	.01	.03	.00	.02	-.55	.00	.02	-.01	-.05	-.05	

TABLE 4.6.--Asymptotic biases--Logistic

μ	$p(y^*>0)$	Sample Mean				μ (σ known)				μ (σ unknown)				σ
		Censored	Truncated	Censored	Truncated	Binary	Truncated	Censored	Truncated	Censored	Truncated	Censored	Truncated	
-3.0	.00	3.00	3.55	.40	2.13	.37	2.13	-1.71	-5.55	.79	1.31	.79	1.31	
-2.8	.01	2.80	3.35	.33	1.94	.30	1.94	-1.54	-5.74	.74	1.31	.74	1.31	
-2.6	.01	2.61	3.15	.26	1.74	.23	1.74	-1.38	-5.93	.68	1.31	.68	1.31	
-2.4	.01	2.41	2.96	.19	1.54	.17	1.54	-1.22	-6.12	.62	1.31	.62	1.31	
-2.2	.02	2.21	2.76	.14	1.35	.11	1.35	-1.07	-6.30	.56	1.31	.56	1.31	
-2.0	.03	2.01	2.56	.09	1.16	.06	1.16	-.92	-6.47	.50	1.31	.50	1.31	
-1.8	.04	1.82	2.36	.04	.97	.01	.97	-.78	-6.62	.44	1.31	.44	1.31	
-1.6	.05	1.63	2.17	.01	.80	-.03	.80	-.65	-6.75	.38	1.32	.38	1.32	
-1.4	.07	1.44	1.97	-.02	.62	-.05	.62	-.52	-6.83	.32	1.32	.32	1.32	
-1.2	.10	1.26	1.78	-.04	.46	-.07	.46	-.41	-6.84	.26	1.31	.26	1.31	
-1.0	.14	1.08	1.59	-.05	.32	-.08	.32	-.31	-6.68	.21	1.29	.21	1.29	
-0.8	.19	.92	1.41	-.06	.19	-.08	.19	-.23	-6.18	.16	1.23	.16	1.23	
-0.6	.25	.76	1.24	-.05	.08	-.07	.08	-.15	-4.94	.11	1.07	.11	1.07	
-0.4	.33	.62	1.07	-.04	-.01	-.05	-.01	-.10	-3.27	.07	.81	.07	.81	
-0.2	.41	.49	.91	-.03	-.06	-.03	-.06	-.06	-2.08	.04	.58	.04	.58	
0.0	.50	.38	.76	-.02	-.10	.00	-.10	-.03	-1.30	.01	.41	.01	.41	
0.2	.59	.29	.63	-.01	-.11	.03	-.11	-.01	-.79	-.01	.28	-.01	.28	
0.4	.67	.22	.52	-.00	-.10	.05	-.10	.01	-.47	-.02	.18	-.02	.18	
0.6	.75	.16	.42	.01	-.09	.07	-.09	.01	-.27	-.03	.11	-.03	.11	
0.8	.81	.12	.33	.01	-.07	.08	-.07	.01	-.14	-.03	.06	-.03	.06	
1.0	.86	.08	.26	.01	-.04	.08	-.04	.01	-.07	-.03	.02	-.03	.02	
1.2	.90	.06	.20	.01	-.03	.07	-.03	.01	-.02	-.03	-.01	-.03	-.01	
1.4	.93	.04	.16	.01	-.01	.05	-.01	.01	.01	-.03	-.02	-.03	-.02	
1.6	.95	.03	.12	.01	.00	.03	.00	.01	.02	-.02	-.03	-.02	-.03	
1.8	.96	.02	.09	.01	.01	-.01	.01	.01	.02	-.02	-.04	-.02	-.04	
2.0	.97	.01	.07	.01	.01	-.06	.01	.01	.03	-.02	-.04	-.02	-.04	
2.2	.98	.01	.05	.00	.02	-.11	.02	.00	.02	-.01	-.04	-.01	-.04	
2.4	.99	.01	.04	.00	.02	-.17	.02	.00	.02	-.01	-.03	-.01	-.03	
2.6	.99	.01	.03	.00	.02	-.23	.02	.00	.02	-.01	-.03	-.01	-.03	
2.8	.99	.00	.02	.00	.01	-.30	.01	.00	.02	-.01	-.03	-.01	-.03	
3.0	1.00	.00	.01	.00	.01	-.37	.01	.00	.01	-.01	-.02	-.01	-.02	

TABLE 4.7.--Asymptotic biases of regression coefficients,
Censored - t_5 ($p = .5$)

True Values			Bias (σ -known)			Bias (σ -unknown)			
β	μ_2	μ_1	β	μ_2	μ_1	β	μ_2	μ_1	σ
-6.0	-3.0	3.0	.55	.55	.00	.57	.57	.00	-.01
-2.0	-1.0	1.0	-.11	-.10	.01	-.11	-.10	.01	-.00
-0.8	-0.4	0.4	-.08	-.08	.00	-.08	-.09	-.01	.01
0.0	0.0	0.0	.00	-.04	-.04	.00	-.05	-.05	.02
0.8	0.4	-0.4	.08	.00	-.08	.08	-.01	-.09	.01
2.0	1.0	-1.0	.11	.01	-.10	.11	.01	-.10	-.00
6.0	3.0	-3.0	-.55	.00	.55	-.57	.00	.57	-.01

TABLE 4.8.--Asymptotic biases of regression coefficients,
Censored - t_5 ($p = .5$)

True Values			Bias (σ -unknown)			
β	μ_2	μ_1	β	μ_2	μ_1	σ
1.0	-3.0	-4.0	.25	-5.30	-5.55	2.29
1.0	-2.0	-3.0	.22	-2.51	-2.73	1.28
1.0	-1.0	-2.0	.17	- .76	- .93	.51
1.0	0.0	-1.0	.14	- .08	- .22	.09
1.0	1.0	0.0	.04	.02	- .02	- .03
1.0	2.0	1.0	-.01	.01	.02	- .04
1.0	3.0	2.0	-.01	.00	.01	- .03
1.0	4.0	3.0	.00	.00	.00	- .01

TABLE 4.9.--Asymptotic biases of regression coefficients,
Censored - t_5

True Values			Bias (σ unknown)				
β	μ_2	μ_1	p	β	μ_2	μ_1	σ
2.0	1.0	-1.0	.1	.34	.00	-.34	.18
2.0	1.0	-1.0	.2	.23	.01	-.22	.09
2.0	1.0	-1.0	.3	.17	.01	-.16	.05
2.0	1.0	-1.0	.4	.13	.01	-.12	.02
2.0	1.0	-1.0	.5	.11	.01	-.10	-.00
2.0	1.0	-1.0	.6	.10	.02	-.08	-.02
2.0	1.0	-1.0	.7	.08	.02	-.06	-.03
2.0	1.0	-1.0	.8	.07	.02	-.05	-.04
2.0	1.0	-1.0	.9	.06	.02	-.04	-.05

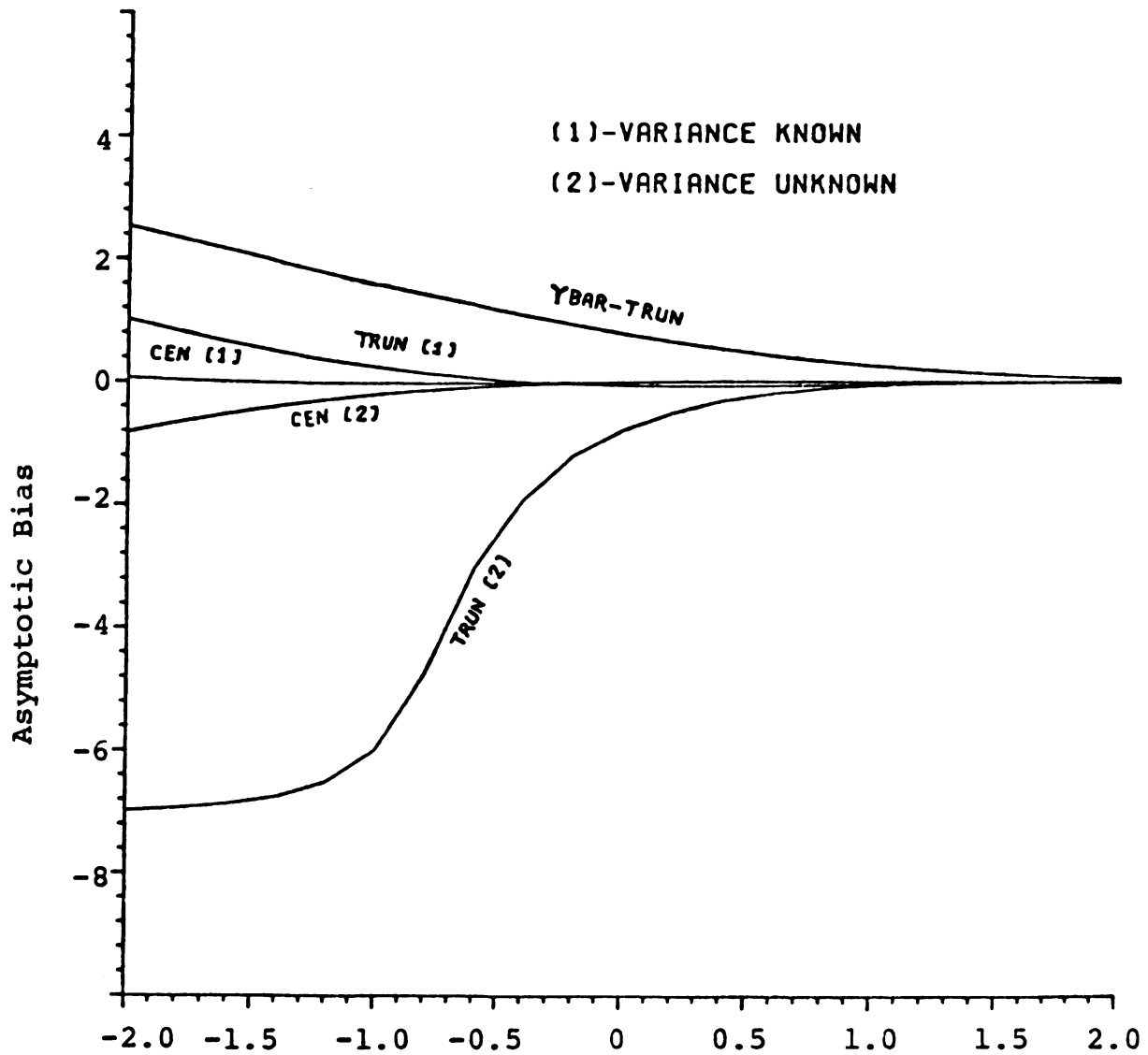


Figure 4.1.--Asymptotic Bias of Estimates of μ , t_{10}
Distribution

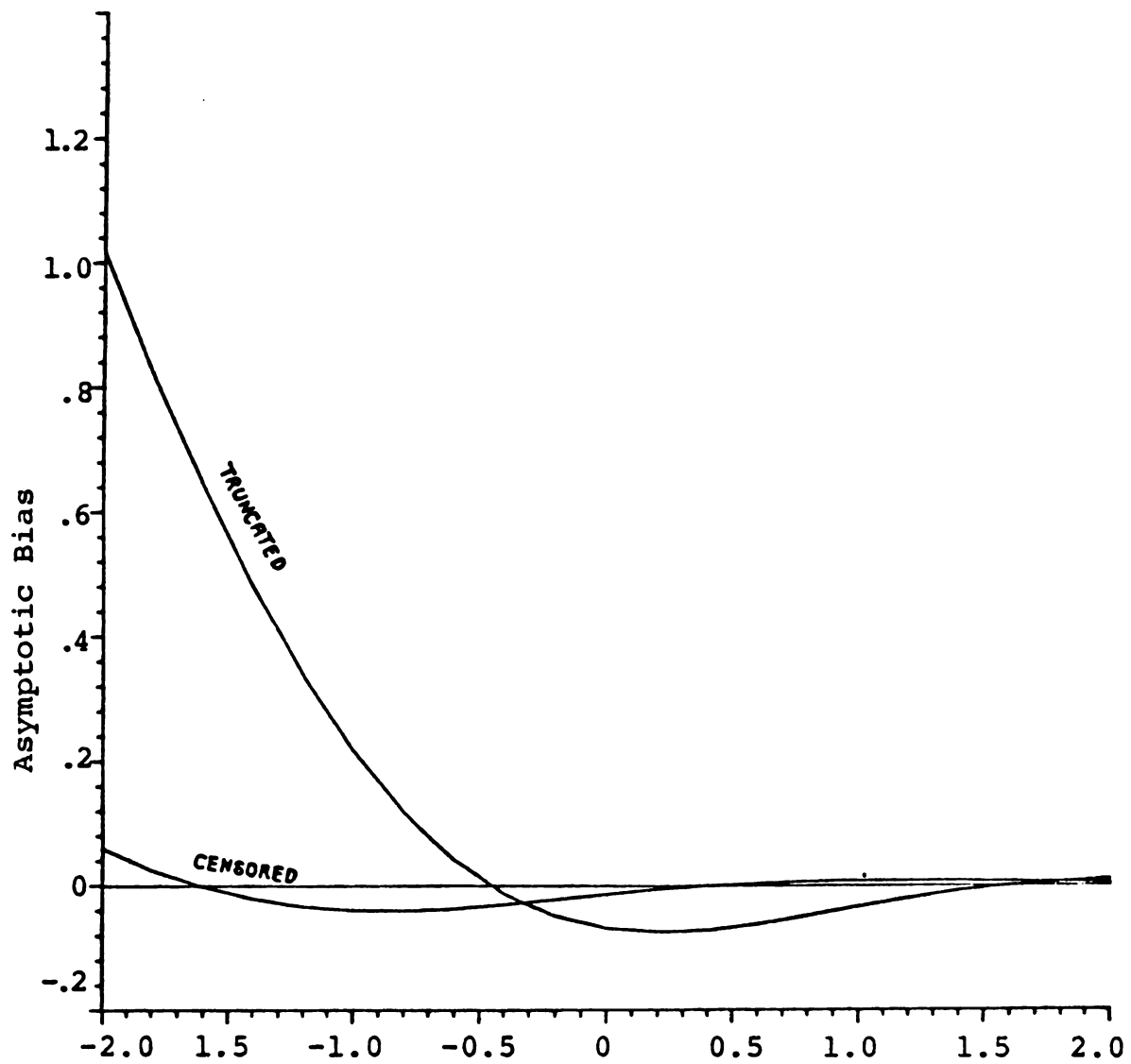


Figure 4.2--Asymptotic Bias of Estimates of μ , t_{10}
Distribution
(Variance known)

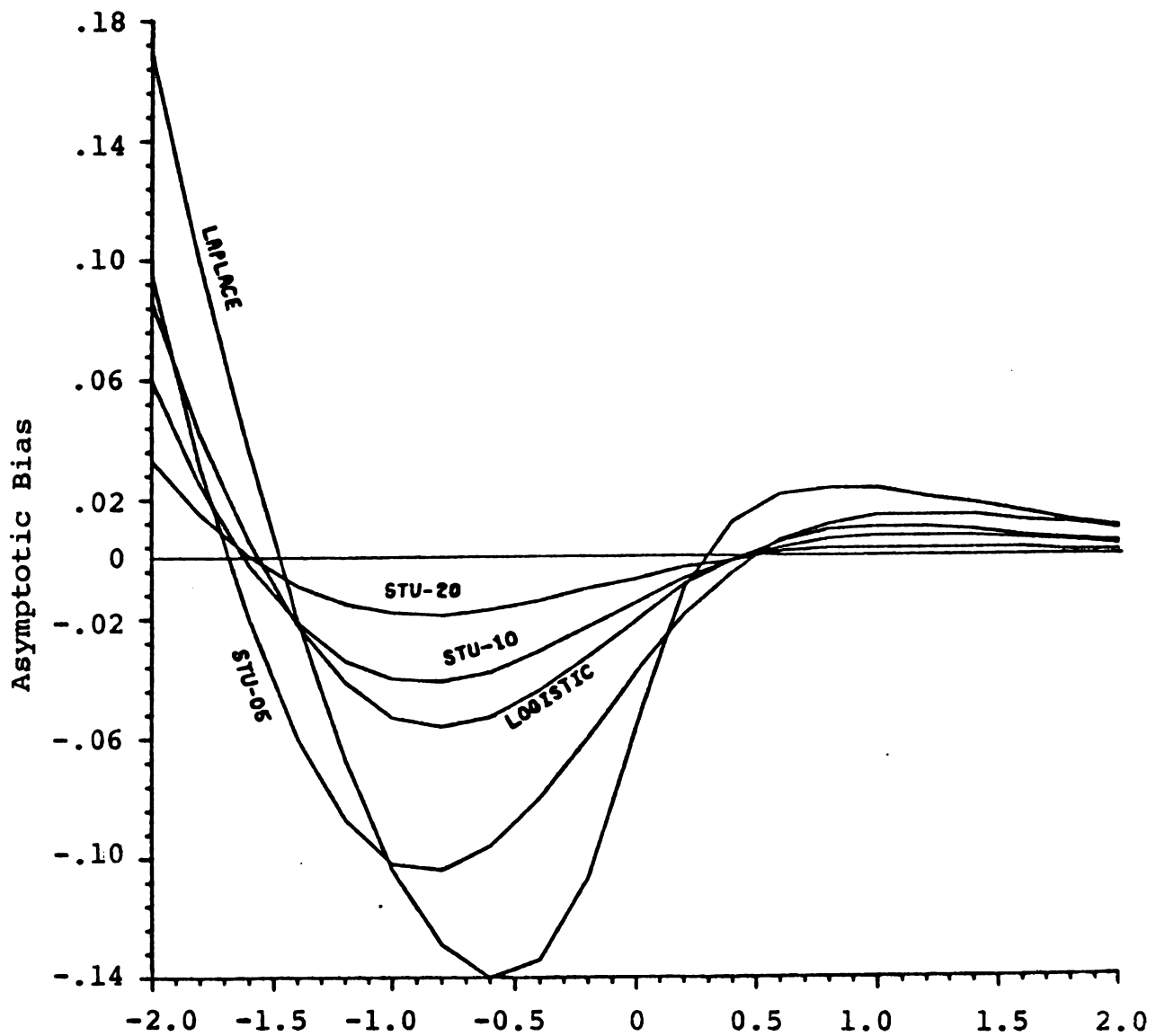


Figure 4.3.--Asymptotic Bias of Estimates of μ , Censored Sample (Variance known).

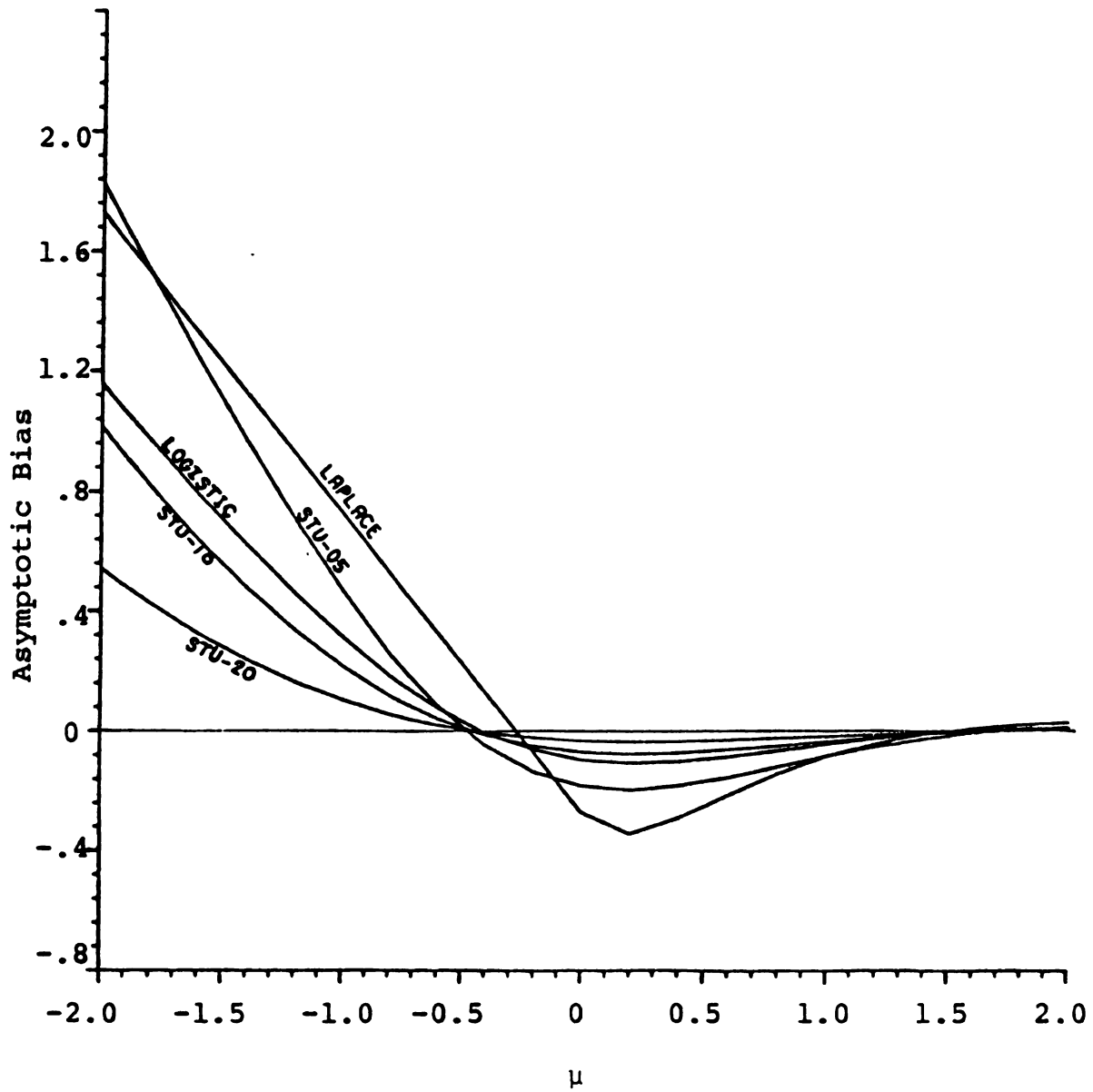


Figure 4.4.--Asymptotic Bias of Estimates of μ , Truncated Sample (variance known)

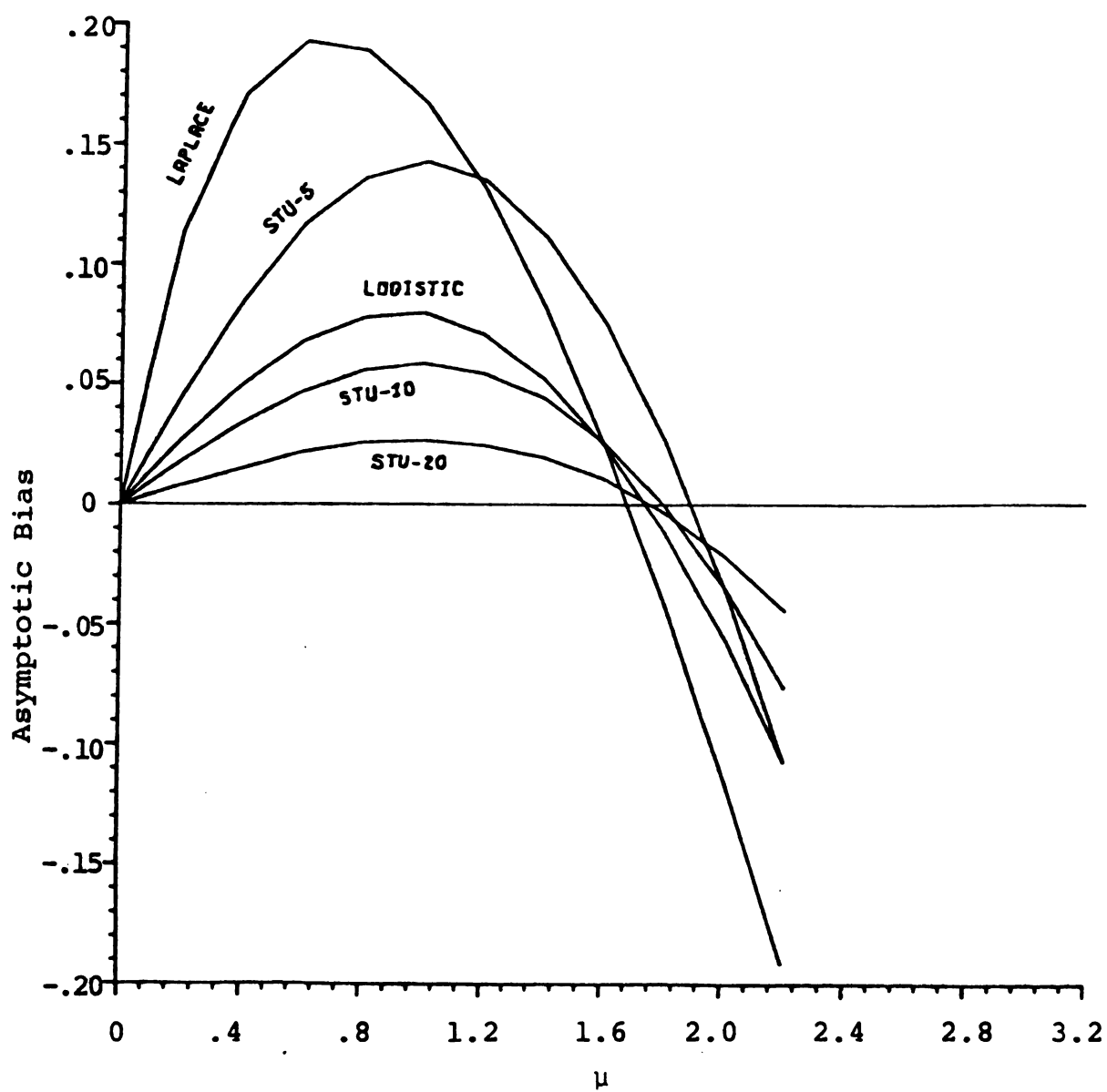


Figure 4.5.--Asymptotic Bias of Estimates of μ , Binary Sample

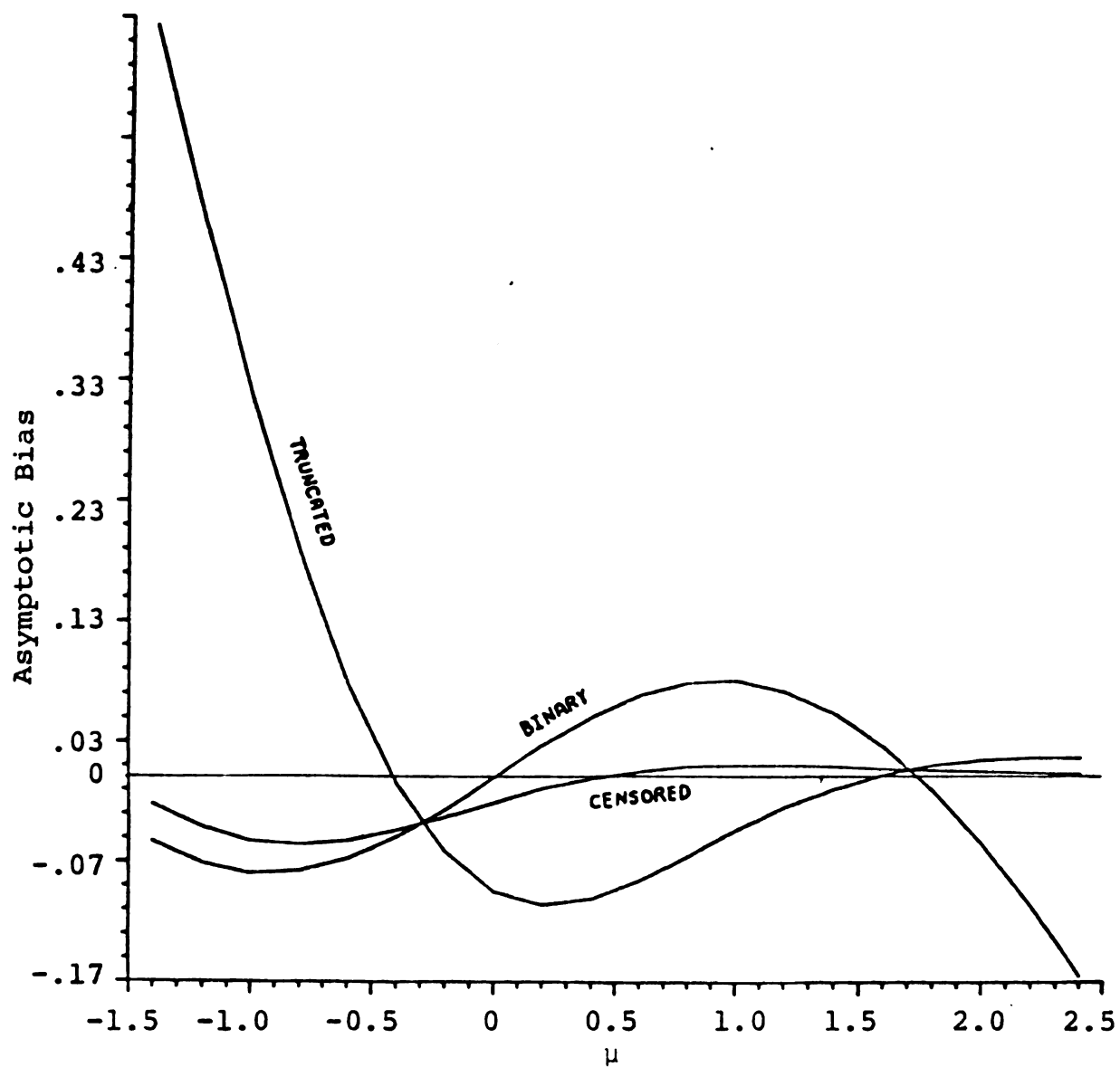


Figure 4.6.--Asymptotic Bias of Estimates of μ , Logistic Distribution (Variance known)

APPENDIX A

TRUNCATED MOMENTS OF SELECTED DISTRIBUTIONS

APPENDIX A

TRUNCATED MOMENTS OF SELECTED DISTRIBUTIONS

1. Student distribution. Let z be distributed as Student with n degrees of freedom and denote the density and cdf of z as f and F . Then

$$f(z) = h(n) \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad z \in (-\infty, +\infty)$$

where $h(n) = [\sqrt{n} B(\frac{1}{2}, \frac{n}{2})]^{-1}$, $E(z) = 0$ and $\text{Var}(z) = \frac{n}{n-2}$

$$F(z) = \int_{-\infty}^z f(t) dt$$

$$E(z | z > -x) = \frac{n+x^2}{n-1} \frac{f(x)}{F(x)}$$

$$E(z^2 | z > -x) = \begin{cases} \left(\frac{n}{n-2}\right) \frac{I_T(\alpha, \beta)}{2F(x)}, & x \leq 0 \\ \left(\frac{n}{n-2}\right) \frac{2-I_T(\alpha, \beta)}{2F(x)}, & x > 0 \end{cases}$$

Where $\alpha = \frac{n-2}{n}$, $\beta = \frac{3}{2}$

Proof:

$$E(z | z > -x) = \frac{1}{F(x)} \int_{-x}^{+\infty} z f(z) dz = \frac{(n + x^2)}{n-1} \frac{f(x)}{F(x)} \quad (\text{Note A1})$$

Using the results of Notes A8 and A9, we get

$$\begin{aligned} E(z^2 | z > -x) &= \frac{1}{F(x)} \int_{-x}^{+\infty} z^2 f(z) dz \\ &= \frac{h(n)}{F(x)} \int_{-x}^{+\infty} z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} dz \\ &= \frac{h(n)}{F(x)} \left[\int_{-x}^0 z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} dz \right. \\ &\quad \left. + \int_0^{+\infty} z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} dz \right] \end{aligned}$$

Using the results in Notes A3 and A4, we get

$$\begin{aligned} &= \begin{cases} \frac{h(n)}{f(x)} \left[\frac{1}{2} n^{\frac{3}{2}} B(\alpha, \beta) I_T(\alpha, \beta) \right], & x \leq 0 \\ \frac{h(n)}{F(x)} \left[n^{\frac{3}{2}} B(\alpha, \beta) \left(1 - \frac{1}{2} I_T(\alpha, \beta)\right) \right], & x > 0 \end{cases} \\ &= \begin{cases} \frac{n}{2F(x)} \frac{B(\alpha, \beta)}{B(\frac{1}{2}, \frac{n}{2})} I_T(\alpha, \beta), & x \leq 0 \\ \frac{n}{2F(x)} \frac{B(\alpha, \beta)}{B(\frac{1}{2}, \frac{n}{2})} (2 - I_T(\alpha, \beta)), & x > 0 \end{cases} \end{aligned}$$

$$= \begin{cases} \left(\frac{n}{n-2}\right) \frac{I_T(\alpha, \beta)}{2F(x)}, & x \leq 0 \\ \left(\frac{n}{n-2}\right) \frac{2 - I_T(\alpha, \beta)}{2F(x)}, & x > 0 \end{cases}$$

2. Logistic distribution. Let z be distributed as logistic, and denote the density and cdf of z as f and F respectively. Then

$$f(z) = e^{-z} (1 + e^{-z})^{-2} = e^z (1 + e^z)^{-2}, \quad z \in (-\infty, +\infty)$$

where $E(z) = 0$ and $\text{Var}(z) = \frac{\pi^2}{3}$

$$F(z) = (1 + e^{-z})^{-1} = e^z (1 + e^z)^{-1}$$

$$\text{and so } [1 - F(z)] = (1 + e^z)^{-1} = e^{-z} (1 + e^{-z})^{-1}$$

$$E(z | z > -x) = -x - \frac{\ln(1-F(x))}{F(x)}$$

$$E(z^2 | z > -x) = \begin{cases} x^2 + \frac{2}{F(x)} [x \ln(1-F(x)) - g(-e^x)], & x \leq 0 \\ x^2 + \frac{2}{F(x)} [x \ln(F(x)) + g(-e^{-x}) \\ + \frac{\pi^2}{6} - \frac{x^2}{2}], & x > 0 \end{cases}$$

$$\text{where } g(t) = \sum_{n=1}^{\infty} (t^n/n^2)$$

Proof:

$$\begin{aligned}
 E(z | z > -x) &= \frac{1}{F(x)} \int_{-x}^{+\infty} z f(z) dz \\
 &= \frac{1}{F(x)} [-xF(x) - \ln(1-F(x))] \quad (\text{Note A6}) \\
 &= -x - \frac{\ln(1-F(x))}{F(x)}
 \end{aligned}$$

$$\begin{aligned}
 E(z^2 | z > -x) &= \frac{1}{F(x)} \int_{-x}^{+\infty} z^2 f(z) dz \\
 &= \begin{cases} \frac{1}{F(x)} \left(-\int_x^0 z^2 f(z) dz + \int_0^{+\infty} z^2 f(z) dz \right), & x \leq 0 \\ \frac{1}{F(x)} \left(\int_{-x}^0 z^2 f(z) dz + \int_0^{+\infty} z^2 f(z) dz \right), & x > 0 \end{cases}
 \end{aligned}$$

Using the results of Notes A5 and A7, we get:

$$\begin{aligned}
 &= \begin{cases} \frac{1}{F(x)} \left[-\frac{\pi^2}{6} + x^2 F(x) + 2x \ln(1-F(x)) - 2g(-e^x) \right. \\ \quad \left. + \frac{\pi^2}{6} \right], & x \leq 0 \\ \frac{1}{F(x)} \left[\frac{\pi^2}{6} - x^2 (1-F(x)) + 2x \ln(F(x)) \right. \\ \quad \left. + 2g(-e^{-x}) + \frac{\pi^2}{6} \right], & x > 0 \end{cases}
 \end{aligned}$$

$$= \begin{cases} x^2 + \frac{2}{F(x)} [x \ln(1-F(x)) - g(-e^x)], & x \leq 0 \\ x^2 + \frac{2}{F(x)} [x \ln(F(x)) + g(-e^{-x}) + \frac{\pi^2}{6} - \frac{x^2}{2}], & x > 0 \end{cases}$$

3. Laplace (double exponential) distribution.

Let z be distributed as Laplace and denote the density and cdf of z as f and F . Then,

$$f(z) = \frac{1}{2} e^{-|z|}, \quad z \in (-\infty, +\infty)$$

where $E(z) = 0$ and $\text{Var}(z) = 2$

$$F(z) = \begin{cases} \frac{1}{2} e^z, & z \leq 0 \\ 1 - \frac{1}{2} e^{-z} = 2e^z (2e^z - 1)^{-1}, & z > 0 \end{cases}$$

$$E(z | z > -x) = \begin{cases} 1-x, & x \leq 0 \\ \frac{1+x}{2e^x - 1}, & x > 0 \end{cases}$$

$$E(z^2 | z > -x) = \begin{cases} 1 + (1-x)^2, & x \leq 0 \\ 2 + \frac{1-(1+x)^2}{2e^x - 1}, & x > 0 \end{cases}$$

Proof:

$$E(z | z > -x) = \frac{1}{F(x)} \int_{-x}^{+\infty} z f(z) dz$$

$$= \begin{cases} \frac{1}{F(x)} \left[\frac{1}{2} e^x (1-x) \right] , & x \leq 0 \\ \frac{1}{F(x)} \left[\frac{1}{2} e^{-x} (1+x) \right] , & x > 0 \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^x , & x \leq 0 \\ \frac{(1+x)}{2e^x - 1} , & x > 0 \end{cases}$$

$$E(z^2 | z > -x) = \frac{1}{F(x)} \int_{-x}^{+\infty} z^2 f(z) dz$$

$$= \begin{cases} \frac{1}{F(x)} \left[\frac{1}{2} e^x (1 + (1-x)^2) \right] , & x \leq 0 \\ \frac{1}{F(x)} \left[\frac{1}{2} e^{-x} (1+x) \right] , & x > 0 \end{cases}$$

$$= \begin{cases} 1 + (1-x)^2 , & x \leq 0 \\ 2 + \frac{1 - (1+x)^2}{2e^x - 1} , & x > 0 \end{cases}$$

Note (A1)

$$\begin{aligned}
 & \int_{-x}^{+\infty} z \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} dz \\
 &= \frac{-n}{n-1} \int_{-x}^{+\infty} -\frac{n-1}{n} z \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} dz \\
 &= \frac{-n}{n-1} \left[\left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n-1)} \right]_{-x}^{+\infty} \\
 &= \frac{n}{n-1} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n-1)} \\
 &= \frac{n+x^2}{n-1} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}
 \end{aligned}$$

Note (A2) on Integration

$$\begin{aligned}
 & \int z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} dz \\
 &= \int z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} \frac{(n+z^2)^2}{-2nz} dw, \quad (w = \frac{n}{n+z^2}) \\
 &= \int z^2 \frac{(n+z^2)^2}{-2nz} \left(\frac{n+z^2}{n}\right)^{-\frac{1}{2}(n+1)} dw \\
 &= \int -\frac{z}{2n} (n+z^2)^{-\frac{1}{2}} (n+z^2)^{\frac{5}{2}} \left(\frac{n+z^2}{n}\right)^{-\frac{1}{2}(n+1)} dw \\
 &= \int -\frac{1}{2} n^{\frac{3}{2}} \left(\frac{z^2}{n+z^2}\right)^{\frac{1}{2}} \left(\frac{n+z^2}{n}\right)^{\frac{5}{2}} \left(\frac{n+z^2}{n}\right)^{-\frac{1}{2}(n+1)} dw
 \end{aligned}$$

$$\begin{aligned}
&= \int -\frac{1}{2} n^{\frac{3}{2}} \left(\frac{n}{n+z^2} \right)^{\frac{1}{2}(n+1)} - \frac{5}{2} \left(\frac{z^2}{n+z^2} \right)^{\frac{1}{2}} dw \\
&= \int -\frac{1}{2} n^{\frac{3}{2}} w^{\alpha-1} (1-w)^{\beta-1} dw
\end{aligned}$$

where $\alpha = \frac{n-2}{2}$ and $\beta = \frac{3}{2}$

Note (A3)

$$\begin{aligned}
&\int_0^{+\infty} z^2 \left(1 + \frac{z^2}{n} \right)^{-\frac{1}{2}(n+1)} dz \\
&= \int_1^0 -\frac{1}{2} n^{\frac{3}{2}} w^{\alpha-1} (1-w)^{\beta-1} dw \quad (\text{Note A2}) \\
&= \frac{1}{2} n^{\frac{3}{2}} \int_0^1 w^{\alpha-1} (1-w)^{\beta-1} dw \\
&= \frac{1}{2} n^{\frac{3}{2}} B(\alpha, \beta)
\end{aligned}$$

where $B(\alpha, \beta)$ is complete Beta function. Also using the properties

$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, it follows that

$$\frac{B(\alpha, \beta)}{B(\frac{1}{2}, \frac{n}{2})} = \frac{B(\frac{n+2}{2}, \frac{3}{2})}{B(\frac{1}{2}, \frac{n}{2})} = \frac{1}{n-2}$$

Note (A4)

$$\begin{aligned}
 & \int_0^x z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}} (n+1) dz \\
 &= \frac{1}{2} n^{\frac{3}{2}} \int_T^1 w^{\alpha-1} (1-w)^{\beta-1} dw, \quad (\text{Note A2}) \quad \left(T = \frac{n}{n+x^2}\right) \\
 &= \frac{1}{2} n^{\frac{3}{2}} \left[B(\alpha, \beta) - \int_0^T w^{\alpha-1} (1-w)^{\beta-1} dw \right] \\
 &= \frac{1}{2} n^{\frac{3}{2}} \left[B(\alpha, \beta) - B(\alpha, \beta) I_T(\alpha, \beta) \right]
 \end{aligned}$$

where $I_T(\alpha, \beta)$ is the incomplete Beta function ratio.

Note (A5) Some facts (summations are from $n = 1$ to $n = +\infty$)

$$\begin{aligned}
 \sum (-1)^{n-1} n^{-2} &= \frac{\pi^2}{12} \\
 \sum (-1)^{n-1} n^{-2} e^{-n|x|} &= - \sum n^{-2} (-e^{-|x|})^n \equiv -g(-e^{-|x|}) \\
 \sum (-1)^{n-1} n^{-1} e^{-n|x|} &= \ln(1 + e^{-|x|}) \\
 \sum (-1)^{n-1} e^{-n|x|} &= e^{-|x|} (1 + e^{-|x|})^{-1} \\
 \sum (-1)^{n-1} n e^{-n|x|} &= e^{-|x|} (1 + e^{-|x|})^{-2} \\
 \int_0^{+\infty} z^2 e^{-z} (1 + e^{-z})^{-2} dz &= \frac{\pi^2}{6}
 \end{aligned}$$

For the last statement, see Johnson and Kotz (1970),
Chapter 22, p. 4.

Note (A6)

$$\begin{aligned}
 & \int_{-x}^{+\infty} z e^{-z} (1 + e^{-z})^{-2} dz \\
 &= \int_{-x}^{+\infty} z \sum_{n=1}^{\infty} (-1)^{n-1} n e^{-nz} dz && \text{(Note A5)} \\
 &= \sum_{n=1}^{\infty} (-1)^{n-1} \int_{-x}^{+\infty} z e^{-nz} dz \\
 &= \sum_{n=1}^{\infty} (-1)^{n-1} (-x e^{nx} + n^{-1} e^{nx}) && \text{(integration by parts)} \\
 &= -x \sum_{n=1}^{\infty} (-1)^{n-1} e^{nx} + \sum_{n=1}^{\infty} (-1)^{n-1} n^{-1} e^{nx} \\
 &= -x e^x (1 + e^x)^{-1} + \ln (1 + e^x) && \text{(Note A5)}
 \end{aligned}$$

Note (A7)

$$\begin{aligned}
 & \int_{-|x|}^0 z^2 e^{-z} (1 + e^{-z})^{-2} dz \\
 &= \int_{-|x|}^0 z^2 \sum_{n=1}^{\infty} (-1)^{n-1} n e^{-nz} dz && \text{(Note A5)} \\
 &= \sum_{n=1}^{\infty} (-1)^{n-1} n^{-2} \int_{-n|x|}^0 s^2 e^s ds && (s = nz)
 \end{aligned}$$

$$\begin{aligned}
&= \sum (-1)^{n-1} n^{-2} [2 - e^{-n|x|} (n^2 x^2 + 2n|x| + 2)] \\
&= 2 \sum (-1)^{n-1} n^{-2} - x^2 \sum (-n)^{n-1} e^{-n|x|} \\
&\quad - 2|x| \sum (-1)^{n-1} n^{-1} e^{-n|x|} - 2 \sum (-1)^{n-1} n^{-2} e^{-n|x|}
\end{aligned}$$

Using the facts in Note A5, we get

$$\begin{aligned}
&= \frac{\pi^2}{6} - x^2 e^{-|x|} (1 + e^{-|x|})^{-1} - 2|x| \ln (1 + e^{-|x|}) \\
&\quad + 2 g(-e^{-|x|})
\end{aligned}$$

where $g(t) = \sum (t^n/n^2)$

Note (A8) Assume $x \leq 0$ and use integration by parts

$$\int_{-x}^{+\infty} \frac{1}{2} z e^{-z} dz = \frac{1}{2} e^x (1 - x)$$

$$\int_{-x}^{+\infty} \frac{1}{2} z^2 e^{-z} dz = \frac{1}{2} e^x [1 + (1 - x)^2]$$

Note (A9) Assume $x > 0$ and do integration by parts

$$\begin{aligned}
&\int_{-x}^{+\infty} \frac{1}{2} z e^z dz \\
&= \int_{-x}^0 \frac{1}{2} z e^z dz + \int_0^{+\infty} \frac{1}{2} z e^z dz
\end{aligned}$$

$$= \frac{1}{2} e^{-x} (1 + x)$$

$$\int_{-x}^{+\infty} \frac{1}{2} z^2 e^z dz$$

$$= 2 - \frac{1}{2} e^{-x} (x^2 + 2x + 2)$$

$$= \frac{1}{2} e^{-x} [4e^x - (x^2 + 2x + 2)]$$

$$= \frac{1}{2} e^{-x} [1 - (1 + x)^2 + 2(2e^x - 1)]$$

APPENDIX B

SOME PROBABILITY LIMITS

APPENDIX B

SOME PROBABILITY LIMITS

Let $y_i = \alpha + \beta x_i + bz_i$

where z_i 's are iid symmetric random variables with $E(z_i) = 0$, and denote the density and cdf of z by f and F . Assume that $b^2 = (\text{Var}(z_i))^{-1}$, which makes the variance of bz unity as a normalization, and x_i is a dummy variable (Bernoulli random variable) with $P(x_i = 1) = p$ and $P(x_i = 0) = q = 1-p$. Then

$$P(y > 0 | x = 1) = P(z > -\frac{\alpha + \beta}{b}) = F(\frac{\alpha + \beta}{b}) \quad (B1)$$

$$P(y > 0 | x = 0) = P(z > -\frac{\alpha}{b}) = F(\frac{\alpha}{b}) \quad (B2)$$

$$P(y > 0) = P(y > 0 | x = 1) P(x = 1) + P(y > 0 | x = 0)$$

$$P(x = 0) = pF(\frac{\alpha + \beta}{b}) + qF(\frac{\alpha}{b}) \quad (B3)$$

See (B1) and (B2)

$$P(y < 0) = pF(-\frac{\alpha + \beta}{b}) + qF(-\frac{\alpha}{b}) \quad (B4)$$

$$\begin{aligned}
P(x = 1 | y > 0) &= \frac{P(y > 0 | x = 1) P(x = 1)}{P(y > 0)} \\
&= \frac{pF\left(\frac{\alpha + \beta}{b}\right)}{pF\left(\frac{\alpha + \beta}{b}\right) + qF\left(\frac{\alpha}{b}\right)} \equiv p_1
\end{aligned} \tag{B5}$$

$$P(x = 0 | y < 0) = \frac{pF\left(-\frac{\alpha + \beta}{b}\right)}{pF\left(-\frac{\alpha + \beta}{b}\right) + qF\left(-\frac{\alpha}{b}\right)} \equiv p_2 \tag{B6}$$

$$\begin{aligned}
E(bz | y > 0, x = 1) &= E(bz | \alpha + \beta + bz > 0) \\
&= bE(z | z > -\frac{\alpha + \beta}{b})
\end{aligned} \tag{B7}$$

$$E(bz | y > 0, x = 0) = bE(z | z > -\frac{\alpha}{b}) \tag{B8}$$

$$\begin{aligned}
E(y | y > 0, x = 1) &= E(\alpha + \beta + bz | \alpha + \beta + bz > 0) \\
&= (\alpha + \beta) + bE(z | z > -\frac{\alpha + \beta}{b})
\end{aligned} \tag{B9}$$

$$E(y | y > 0, x = 0) = \alpha + bE(z | z > -\frac{\alpha}{b}) \tag{B10}$$

$$\begin{aligned}
E(y^2 | y > 0, x = 1) &= E[(\alpha + \beta + bz)^2 | \alpha + \beta + bz > 0] \\
&= (\alpha + \beta)^2 + 2(\alpha + \beta)bE(z | z > -\frac{\alpha + \beta}{b}) \\
&\quad + b^2E(z^2 | z > -\frac{\alpha + \beta}{b})
\end{aligned} \tag{B11}$$

$$\begin{aligned}
E(y^2 | y > 0, x = 0) &= \alpha^2 + 2\alpha bE(z | z > -\frac{\alpha}{b}) \\
&\quad + b^2E(z^2 | z > -\frac{\alpha}{b})
\end{aligned} \tag{B12}$$

$$E(y|y > 0) = p_1 E(y|y > 0, x = 1) + q_1 E(y|y > 0, x = 0), \quad (B13)$$

$(p_1 + q_1 = 1)$. See (B9) and (B10)

$$E(y^2|y > 0) = p_1 E(y^2|y > 0, x = 1) + q_1 E(y^2|y > 0, x = 0) \quad (B14)$$

The following are some probability limits.

Let $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\sigma}$ be probability limits of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}$. As $T \rightarrow +\infty$ ($n \rightarrow +\infty$), we have the following probability limits

$$\frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{1}{n} \sum_{i=1}^n x_i \equiv \frac{k_1}{n} \rightarrow P(x = 1|y > 0) = p_1 \quad (B15)$$

See (B5).

$$\begin{aligned} \frac{1}{T-n} \sum_{i=1}^{T-n} x_i^2 &= \frac{1}{T-n} \sum_{i=1}^{T-n} x_i \equiv \frac{k_2}{T-n} \rightarrow P(x = 1|y > 0) \\ &= p_2 \end{aligned} \quad (B16)$$

See (B6).

$$\frac{T-n}{n} = \frac{1}{\frac{n}{T}} - 1 \rightarrow \frac{1}{p(y > 0)} - 1 = \frac{P(y \leq 0)}{P(y > 0)} \equiv \gamma \quad (B17)$$

See (B3) and (B4).

$$\frac{k_2}{n} = \frac{T-n}{n} \quad \frac{k_2}{T-n} \rightarrow \gamma p_2 \quad (B18)$$

$$\frac{T-n-k_2}{n} = \frac{T-n}{n} - \frac{k_2}{n} \rightarrow \gamma - \gamma p_2 = \gamma(1 - p_2) \equiv \gamma q_2 \quad (B19)$$

$$\frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{k_1}{n} \quad \frac{1}{k_1} \sum_{i=1}^{k_1} y_i \rightarrow p_1 E(y|y > 0, x = 1) \quad (B20)$$

See (B9).

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta} x_i)) &= \frac{1}{n} \sum_{i=1}^n y_i - \hat{\beta} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) - \hat{\alpha} \\ &\rightarrow E(y|y > 0) - p_1 \tilde{\beta} - \tilde{\alpha} \end{aligned} \quad (B21)$$

See (B15) and (B13).

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n x_i (y_i - (\hat{\alpha} + \hat{\beta} x_i)) &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \hat{\alpha} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \\ &\quad - \hat{\beta} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) \rightarrow p_1 \left[\alpha + \beta + bE(z|z > \frac{\alpha + \beta}{b}) \right. \\ &\quad \left. - \tilde{\alpha} - \tilde{\beta} \right] \end{aligned} \quad (B22)$$

See (B15) and (B20).

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta}x_i))^2 \rightarrow E(y^2 | y > 0) + p_1 \tilde{\beta}^2 \\
& + 2p_1 \tilde{\alpha} \tilde{\beta} - 2 p_1 [\alpha + \beta + bE(z | z > -\frac{\alpha + \beta}{b})] \tilde{\beta} \\
& - 2E(y | y > 0) \tilde{\alpha} + \tilde{\alpha}^2
\end{aligned} \tag{B23}$$

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{T-n} m(-\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}) &= \frac{1}{n} \sum_{i=1}^{k_2} m(-\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}}) + \frac{1}{n} \sum_{i=1}^{T-n-k_2} m(-\frac{\hat{\alpha}}{\hat{\sigma}}) \\
&= \frac{k_2}{n} m(-\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}}) + \frac{T-n-k_2}{n} m(-\frac{\hat{\alpha}}{\hat{\sigma}}) \\
&\rightarrow \gamma p_2 m(-\frac{\tilde{\alpha} + \tilde{\beta}}{\tilde{\sigma}}) + \gamma q_2 m(-\frac{\tilde{\alpha}}{\tilde{\sigma}})
\end{aligned} \tag{B24}$$

See (B18) and (B19).

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{T-n} x_i m(-\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}) &= \frac{1}{n} \sum_{i=1}^{k_2} m(-\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}}) \\
&\rightarrow \gamma p_2 m(-\frac{\tilde{\alpha} + \tilde{\beta}}{\tilde{\sigma}})
\end{aligned} \tag{B25}$$

See (18B).

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{T-n} (\hat{\alpha} + \hat{\beta}x_i) m(-\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}) \\
&= \frac{\hat{\alpha}}{n} \sum_{i=1}^{T-n} m(-\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}) + \frac{\hat{\beta}}{n} \sum_{i=1}^{T-n} x_i m(-\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}) \\
&\rightarrow \tilde{\alpha} \text{ (B24)} + \tilde{\beta} \text{ (B25)}
\end{aligned} \tag{B26}$$

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n m\left(\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}\right) &= \frac{1}{n} \sum_{i=1}^{k_1} m\left(\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}}\right) + \frac{1}{n} \sum_{i=1}^{n-k_1} m\left(\frac{\hat{\alpha}}{\hat{\sigma}}\right) \\
&\rightarrow p_1 m\left(\frac{\tilde{\alpha} + \tilde{\beta}}{\tilde{\sigma}}\right) + q_1 m\left(\frac{\tilde{\alpha}}{\tilde{\sigma}}\right)
\end{aligned} \tag{B27}$$

See (B15).

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n x_i m\left(\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}\right) &= \frac{1}{n} \sum_{i=1}^{k_1} m\left(\frac{\hat{\alpha} + \hat{\beta}}{\hat{\sigma}}\right) \\
&\rightarrow p_1 m\left(\frac{\tilde{\alpha} + \tilde{\beta}}{\tilde{\sigma}}\right)
\end{aligned} \tag{B28}$$

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n (\hat{\alpha} + \hat{\beta}x_i) m\left(\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}\right) &= \frac{\hat{\alpha}}{n} \sum_{i=1}^n m\left(\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}\right) \\
&+ \frac{\hat{\beta}}{n} \sum_{i=1}^n m\left(\frac{\hat{\alpha} + \hat{\beta}x_i}{\hat{\sigma}}\right) \rightarrow \tilde{\alpha} \text{ (B27)} + \tilde{\beta} \text{ (B28)}
\end{aligned} \tag{B29}$$

CHAPTER V

CONCLUSIONS

The Tobit model is being employed with increasing frequency in economics and other areas. The assumptions underlying the model are quite strong and more attention must be paid to the effects of violating those assumptions to avoid erroneous inferences.

The implication of heteroskedasticity in the Tobit model is investigated in Chapter II. The model considered has only a constant term, and heteroskedasticity is introduced into the model by assuming two distinct subsamples, each with a different variance of the normal random error. Calculating the asymptotic mean of the estimator for a variety of parameter values, the main conclusions can be summarized as follows: (1) heteroskedasticity will lead to inconsistent estimates of the coefficients, the severity of which increases with severity of heteroskedasticity and the degree of truncation or censoring of the sample; and (2) heteroskedasticity of a given severity causes less inconsistency in the censored Tobit model than in the corresponding truncated model. While it is dangerous

to generalize the results of our simple model, it is reasonable to conclude that moderate heteroskedasticity (say, variances differing by a factor of two) is not likely to cause substantial inconsistency in the censored Tobit model unless the sample is heavily censored (more than half of the observations at the limit). The study of the truncated case by Hurd (1979) is less optimistic. It should also be noted that heteroskedasticity invalidates the usual test statistics, even in the complete sample regression case.

Considering these implications of heteroskedasticity in the Tobit model, some general test for heteroskedasticity would be useful. Such a test is developed in Chapter III, using the Lagrangian multiplier test principle. Assuming (as an alternative to homoskedasticity) that the variance of the error terms is a linear function of some exogenous variables, the test statistics are given for both the truncated and censored cases. Although an interpretation of the form of the test statistics is not simple, the test statistics are not difficult to calculate, using the estimated Tobit residuals. This test has the same asymptotic properties as the likelihood ratio test. It should be added that if a theoretical specification of the cause of heteroskedasticity can be made, the model can be corrected for heteroskedasticity by

taking it into account in the estimation. However, the attraction of the Lagrangian multiplier procedure is that one can test for heteroskedasticity first to see whether or not the more cumbersome estimation procedure is necessary.

The robustness of Tobit estimators to non-normality is considered in Chapter IV. The asymptotic bias of the estimate of the mean of a random variable, for censored, truncated and binary samples, when normality is assumed but the distribution is in fact non-normal, has been calculated using a variety of parameter values. Our analysis suggests that:

1. The bias for non-normality can be substantial. This is especially true in the realistic case in which disturbance variance is unknown. In fact, the bias of the normal MLE can be larger than the bias of the uncorrected sample mean.

2. The censored estimator is usually much less biased than the truncated estimator. Therefore, the limit observations should be included if available.

3. The bias due to non-normality depends on the degree of censoring (or truncation) of the sample. For example, for samples that are 75 percent complete, there is virtually no bias, while for the largely incomplete samples, the bias is substantial. One practical significance of our results is that, for largely incomplete

samples, estimation methods based on less restrictive distributional assumptions may be worth considering.

It is important to stress that our results may fail to be general for two reasons. First, we have estimated only a mean, not a regression. Second, for the sake of tractability, we have considered only a limited number of non-normal distributions, all of which are symmetric.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Aitchison, J., and S. D. Silvey (1958), "Maximum-likelihood Estimation of Parameters Subject to Restraints," Annals of Mathematical Statistics, 29, 813-828.
- _____ (1960), "Maximum-likelihood Estimation Procedures and Associated Tests of Significance," Journal of the Royal Statistical Society, Series B, 22, 145-171.
- Amemiya, T. (1973), "Regression Analysis when the Dependent Variable Is Truncated Normal," Econometrica, 41, 997-1016.
- _____ (1977), "A Note on a Heteroskedastic Model," Journal of Econometrics, 6, 365-370.
- Breusch, T. S., and A. R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," Econometrica, 47, 1287-1294.
- _____ (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," Review of Economic Studies, XLVII, 239-253.
- Cragg, J. G. (1971), "Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods," Econometrica, 5, 829-844.
- Goldberger, A. S. (1980), "Abnormal Selection Bias," SSRI Discussion Paper 8006, University of Wisconsin, Madison.
- Gronau, R. (1973), "The Effect of Children on the Household's Value of Time," Journal of Political Economy, 81, 168-199.
- _____ (1974), "Wage Comparisons--A Selectivity Bias," Journal of Political Economy, 82, 1119-1144.

- Hausman, J. A. (1978), "Specification Tests in Econometrics," Econometrica, 46, 1251-1272.
- Hausman, J. A., and D. A. Wise (1977), "Social Experimentation, Truncated Distributions, and Efficient Estimation," Econometrica, 45, 919-938.
- Heckman, J. (1974), "Shadow Prices, Market Wages, and Labor Supply," Econometrica, 42, 679-694.
- ____ (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimators for Such Models," Annals of Economic and Social Measurement, 5, 475-492.
- ____ (1979), "Sample Selection Bias as a Specification Error," Econometrica, 47, 153-162.
- Hurd, M. (1979), "Estimation in Truncated Samples When There is Heteroskedasticity," Journal of Econometrics, 11, 247-258.
- Johnson, N., and S. Kotz (1969), Discrete Distributions, New York, Wiley.
- ____ (1970), Continuous Univariate Distributions, Vol. I and II, New York, Wiley.
- Judge, G. G., W. E. Griffiths, R. C. Hill, and T. C. Lee (1980), The Theory and Practice of Econometrics, New York, Wiley.
- Lee, L. F. (1981a), "Estimation of Some Non-Normal Limited Dependent Variable Models," Discussion Paper 43, Center for Econometrics and Decision Sciences, University of Florida.
- ____ (1981b), "A Specification Test for Normality Assumption for the Truncated and Censored Tobit Models," Discussion Paper 44, CEDS, University of Florida.
- ____ (1971c), "A Specification Test for Normality in the Generalized Censored Regression Models," Unpublished Paper.
- Lee, L. F., and R. P. Trost (1978), "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand," Journal of Econometrics, 8, 357-382.

- Maddala, G. S. (1979), "Specification Errors in Limited Dependent Variable Models," Discussion Paper, University of Florida.
- Maddala, G. S., and F. D. Nelson (1975), "Specification Errors in Limited Dependent Variable Models," NBER Working Paper No. 96.
- McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in, Frontiers in Econometrics, New York, Academic.
- Nelson, F. D. (1977), "Censored Regression Models with Unobserved, Stochastic Censoring Thresholds," Journal of Econometrics, 6, 309-328.
- _____ (1979), "The Effect of and A Test for Misspecification in the Censored-Normal Model," Caltech Social Science Working Paper No. 291.
- _____ (1981), "A Test for Misspecification in the Censored Normal Model," Econometrica, Forthcoming.
- Olmsted, J. M. (1961), Advanced Calculus, New Jersey: Prentice-Hall.
- Olsen, R. J. (1979), "Tests for the Presence of Selectivity Bias and Their Relation to Specification of Functional Form and Error Distribution," ISPS, Working Paper 812, Yale University.
- Pearson, K. (1968), Tables of the Incomplete Beta-Function, 2nd Ed., Cambridge: The University Press.
- Prais, S. J., and H. S. Houthakker (1955), The Analysis of Family Budget, New York: Cambridge University Press.
- Rao, C. R. (1973), Linear Statistical Inference and Its Applications, 2nd Ed., New York: Wiley.
- Schmidt, P. (1976), Econometrics, New York: Marcel Dekker.
- Schmidt, P., and R. P. Strauss (1975), "The Prediction of Occupation Using Multiple Logit Models," International Economic Review, 16, 471-486.
- Silvey, S. D. (1959), "The Lagrangian Multiplier Test," Annals of Mathematical Statistics, 30, 389-407.

Smirnov, N. V. (1961), Tables for the Distribution and Density Functions of t-Distribution, New York: Pergamon Press.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26, 24-36.