Tobit (Censored) and Ordered Models

In the microcenter, change directories to c:\temp and copy programs and data to this directory by typing (once you are in Gauss):

```plaintext
Chdir c:\temp;
Dos copy w:\classdata\kleine401\Tobit
```

1. There is an incomplete program, Ltobit.g, which calculates the likelihood (observation by observation) for the Tobit model. Edit the program Ltobit.g by typing:

```plaintext
Edit Ltobit.g
```

Complete all lines with a ?, save the program, and run it. When there are no errors, proceed.

2. The file Dtobit generate data on a censored or Tobit model. Look at this program, a copy of which is shown below, and make sure that you understand the data generating process

```plaintext
N = 2000;
X1 = (rndn(N,1)^2)/sqrt(2);
X2 = rndn(N,1);
Ystar = -2*X1 + 2*X2 + 4 + 2*rndn(N,1);
s = (Ystar .gt 0);
Y = Ystar.*s;
X = X1~X2~ones(N,1);
Xp = selif(X,s);
Yp = selif(Y,s);
```

When you type run dtobit, every line above will be executed. Notice that Y and X hold all of the original data while Xp and Yp only hold those observations for which Y is positive.

a) OLS. Employing all if the data, estimate the model by OLS, report the estimates, and compare them with the true parameter values. In your answer, you should informally discuss why the OLS estimator is not consistent.
b) **OLS.** Employing only the data for which Y is positive (i.e. Yp and Xp), estimate the model by OLS, report the estimates, and compare them with the true parameter values. In your answer, you should informally discuss why the OLS estimator is not consistent.

c) **Maximum Likelihood.** Letting bols be the OLS estimator from either (a) or (b), calculate the column of 2000 squared residuals, r2. Hint: r2 = ( ? )^2. With mean giving the average of a column of numbers, estimate and report the error variance, sig2. With sig = sqrt(sig2) as the square-root of the variance, employ bols and sig to construct starting values. Be careful, the Tobit program, as discussed in class, estimates parameters relative to sig and also sig (last). Being careful to make the maximum likelihood library available by typing:

```
Library Maxlik,
```

report the maximum likelihood estimates and compare them with the true parameter values.

d) Employing either a J-test or a likelihood ratio test, at the 10% significance level test the null hypothesis that X1 and X2 do not belong in the model (the alternative is simply Not H0).

Remark: The command cdfchic(r,T) calculates the probability above T for a chi-square variable with r degrees of freedom.

3) **Reported demand data.** Load the demand data by typing:

```
Load L,Q,X;
```

Here, L is a label for the variables in X (each column represents a different variable). To see the list of variable names (a string), Type: $L; These names are in the same order as are the variables in X. For example, X[.,1] will be the first variable in the list of names.

Calculate the sample size by typing rows(Q);
4) A Tobit Demand Model.

a) Data Preparation.

i) To prepare that data on the exogenous variables, create a high income dummy variable that is one if the individual is above median income and is zero otherwise. Replacing the ?, the high income dummy variable can be calculated as:

\[
\text{Income} = X[, ?]; \\
H = \text{Income} \geq \text{median(Income)};
\]

Here, the first command specifies which X-variable is income (i.e. specifies which column of X is income) and the second command creates the dummy variable.

ii) Finally, to prepare the data, we need a variable that is income when income is high and is zero otherwise. We will review in class the reason for this. Calculate this variable as:

\[
\text{Hincome} = H.*\text{Income};
\]

Then, append this variable to the X-data by typing: \( \text{XX} = X\sim\text{Hincome} \). Modify the label to name the new X-variable by typing: \( \text{LL} = L["Hinc"] \); The quotes tell gauss that Hinc is a string (list of characters and not numbers). Type $\text{LL}$ and you will see the new list of names.

b) Using the program Ltobit, estimate a Tobit model by maximum likelihood. Report the column (vector) of estimates (b).

c) Store the diagonal elements from the covariance matrix under the name S2 \( (S2= \text{diag(C)} \), where C is the name for the covariance. Store the square roots of estimated coefficients on the X-variables under the name S by typing \( S = \text{sqrt(S2[1:?:])} \). Here, you will need to specify the value for ?.

d) Find "t-ratios" for the variables by typing:

\[
t = b[1 :?]/s;
\]

Here, b and s are columns of numbers. The ./ command results in every element in b being divided by the corresponding element in s. Calculate
and report P-values for all of the coefficients on the X-variables. Note that \(1 - \text{cdfn}(\text{abs}(t))\) calculates the tail probability above the positive number(s) \(\text{abs}(t)\), where \(\text{abs}(t)\) takes the positive value of \(t\). By typing one line, you should be able to get all P-values at once.

e) For every X-coefficient that is significantly different from 0 at the 10% level and to the extent that it is possible, discuss whether or not the signs are reasonable (something you should do in all empirical work). A very brief answer here will suffice. Remark: Any applied paper should discuss whether or not the signs are as expected for "important" variables (i.e. those that belong in the model).

f) At the 10% income level, test for whether or not income has the same impact on demand for both high and low income groups (you already have the computer output to do this).

g) At the 10% significance level, test the null hypothesis that every coefficient, except possibly the constant term, is zero.

Remark: The type of calculations in (c) -(g) should be in all term papers.

5. Optional

a) Assuming that true demand follows a Tobit model, discuss how you might test for whether or not there is accurate reporting.

b) It is always good practice to actually look at your data. For this problem, type:

\[
\begin{align*}
\text{uq} & = \text{unique}(q, 1); \\
\text{f} & = \text{counts}(q, \text{uq})/\text{row}(s(q)); \\
\text{uq} \sim \text{f};
\end{align*}
\]

Here, \(\text{uq}\) holds the distinct or unique values of \(q\) and \(f\) holds the corresponding sample frequency. Given the above results, comment on whether or not a Tobit model is appropriate for these data (assuming accurate reporting in this part).