ECO 6375 Prof. T. Fomby
Econometrics III Fall 2008

EXERCISE 6

Purpose: To learn how to estimate and analyze **truncated data** using the **Truncated Regression Model**. **Truncated data** arise when we do not observe a certain segment of the population. This typically happens when a survey targets a particular subset of the population and, perhaps due to cost considerations or oversight, entirely ignores the other part of the population.

Consider again the Fair (1978) data on extramarital affairs analyzed in the Tobit exercise. You should have created an EVIEWS workfile called Affair.wf1. Now consider the following scenario. Suppose instead of having all of the observations in Affair.dat, for some reason, your research assistant has lost the first 451 observations on all those individuals who were not involved in extramarital affairs. (Maybe, in the first place, your research assistant just took the time to interview the 150 persons of the 601 persons that had extramarital affairs thinking (naively) that the persons who had not engaged in extramarital affairs were not important in studying extramarital behavior.) As we know from our textbook ordinary least squares on truncated data provides biased coefficient estimates and can lead us to drawing misleading statistical conclusions. So let's try to recover from our research assistant's "mistake." Go to your Affair.wf1 program in EVIEWS and copy it into a workfile named Affair2.wf1. In this new workfile, change the sample range to be 452 - 601. (This essentially forces you to ignore the observations on those individuals who had no extramarital affairs as if your research assistant had never observed them for you.) Now let's run a **truncated regression analysis** on the last 150 observations on those individuals who had one or more extramarital affairs. Go to **Quick**, **Estimate Equation**, **Method**, then choose **Censored**. Then choose the **Normal** cumulative distribution, **truncated sample,** and input a **left-censoring point of 0.99**. Then enter your initial specification of y_pt as the dependent variable and c, z1, z2, z3, z4, z5, z6, z7, and z8 as explanatory variables.

(a) Estimate the truncated regression model using all of the above explanatory variables. Now using the **backward selection technique** build a final model of extramarital affairs sequentially deleting insignificant variables as determined by the truncated regression model technique. Report the final model with coefficient estimates and accompanying standard errors. How does this final model compare to final model that you obtained in the Tobit exercise using the Tobit censored regressions approach? What did we miss out on by ignoring the observations where y_pt = 0?

(b) Consider the truncated observation data set consisting of observations 452 - 601. Using **Least Squares and the backward selection technique**, build a "final" model for extramarital affairs. (Remember this model provides biased coefficient estimates and incorrect statistical inferences.) How does this final model differ from the final Tobit model obtained in the Tobit exercise?

(c) As one last attempt to recover from the disaster of having to work with the truncated sample our research assistant gave us, try the **Logistic** cumulative distribution with **truncated sample** and the explanatory variables of c, z2, z3, z5, and z8 in a truncated regression model (with censoring value of 0.99). How does this model differ from the final model you obtained in the Tobit exercise analysis? Evidently, it is not

always easy to recover from the mistakes imposed on us from truncated values of the dependent variable. Right?