

GV903: Advanced Research Methods

Class 18 Limited Dependent Variables

Heckman selection model

Read directly in R the dataset from this link <http://www.stata-press.com/data/r12/womenwk.dta>

1. Have a quick look at the dataset and understand the nature of the data we have.
2. Run an OLS regression of *wage* on *age* and *education*.
3. Run an OLS regression of *wage* on *age*, *education*, *married* and *children*.
4. Find the appropriate package in R for sample selection models.
5. Generate a dummy variable that will act as the selection variable. This should have 0 when *wage* is not observed and 1 otherwise.
6. Fit a Heckman selection model, choosing carefully which variables you will use in the outcome equation and which in the selection equation. Compare all models.

Probit model with sample selection (Heckman Probit)

Load in R the Credit Score dataset we used in a previous lab.

1. Make a crosstabulation of *cardholder* status and *default* status.
2. Fit a Probit model of *CARDHLDR* on *AGE*, *INCOME*, *OWNRENT*, *ACADMOS*, *SELFEMPL*, *MAJORDRG* and *MINORDRG*. Do you remember what all these variables measure? The unit of measurement?
3. Fit a Probit model of *DEFAULT* on *AGE*, *INCOME*, *OWNRENT*, *ACADMOS*, *SELFEMPL*, *MAJORDRG* and *MINORDRG*.
4. Fit the same model as in (3) but only for the cardholders in the sample.
5. Fit a Heckman probit model choosing carefully which variables you include in the outcome equation and which in the selection equation. Compare all models.