

Econometrics 2 (Fall 2020)

Homework 4: Sample Selection.

Due Wednesday on Sept. 23, 2020.

The data generating process in this code is that of a model of selectivity (as in Davidson and Mackinnon p. 478). The latent variables, y_i^o and z_i^o , are given by

$$\begin{bmatrix} y_i^o \\ z_i^o \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 x_i \\ \gamma_1 w_i \end{bmatrix} + \begin{bmatrix} u_i \\ v_i \end{bmatrix},$$

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim NID \begin{pmatrix} \sigma_u^2 & \rho\sigma_u \\ \rho\sigma_u & 1 \end{pmatrix},$$

and the relationship between observed variables, y_i and z_i , and the latent variables is

$$y_i = \begin{cases} y_i^o & \text{if } z_i^o > 0 \\ \text{unobserved} & \text{otherwise} \end{cases},$$

$$z_i = \begin{cases} 1 & \text{if } z_i^o > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Set the parameters.

There are 50 simulations with 100 observations per simulation. Set $\beta_0 = 0.5, \beta_1 = 3, \gamma_0 = 1, \gamma_1 = 2, \rho = -0.9$ and $\sigma_u = 2$.

Maximum Likelihood Estimation.

In each simulation, draw the error terms, U and V , from the normal distribution, generate the latent variables Y^o and Z^o then generate the observed data, W, X, Y and Z . Estimate the model using maximum likelihood and record the estimates.

Display the results of the last simulation.

The estimate and standard error (in parenthesis) of β_0 in the last simulation is:

The estimate and standard error (in parenthesis) of β_1 in the last simulation is:

The estimate and standard error (in parenthesis) of σ in the last simulation is:

Empirical results.

The average and standard deviation (in parenthesis) of β_0 is:

The average and standard deviation (in parenthesis) of β_1 is:

The average and standard deviation (in parenthesis) of σ is:

Plot the estimated coefficients from all simulations.

Plot a histogram of β_0 .

Plot a histogram of β_1 .

Plot a histogram of σ .