Recent Advances in the Field of Trade Theory and Policy Analysis Using Micro-Level Data

July 2012
Bangkok, Thailand

Cosimo Beverelli
(World Trade Organization)
Sample selection models in Stata

• Heckman estimator
• Panel data
Heckman estimator

• Heckman “two-stage” estimator:
  • First stage probit (selection equation)
  • A second stage OLS (outcome equation)

• Do not estimate these two models separately and plug the inverse Mill’s ratio into OLS: you will obtain incorrect standard errors!

• Use Stata’s `heckman` command:
  • `heckman depvar indepvar1 indepvar2 ..., select(indepvar1 indepvar2 ... overidvar1) options`

• It produces maximum likelihood estimates of the two equations jointly
• If you have trouble obtaining ML estimates, use the twostep option (but `vce(robust)` and `vce(cluster clustvar)`) are not available

• See “heckman.do”
Panel data

- Stata does not have any built-in panel data models for Heckman.
- Recall that the first stage probit model is inconsistent in the presence of fixed effects.

- To estimate a fixed effects Heckman model, include the dummy variables manually...
  - `heckman depvar indepvar1 indepvar2 ... dum1 dum2 ..., select(indepvar1 indepvar2 ... dum1 dum2 ...), options`

- ...But you have to be very cautious when interpreting the results of such models.