

## Developing Trade Consultants

Policy | Research | Capacity Building

# Sample Selection Models in Stata

Ben Shepherd

Principal, Developing Trade Consultants

# The Heckman Model

---

- } The Heckman sample selection model combines:
  - } A 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
  - } Which produces maximum likelihood estimates of the two equations jointly.
  
- } If you have trouble obtaining ML estimates, use the twostep option.

# Panel data models

---

- } 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 ... overidvar1  
...) options
  
- } Be very cautious when interpreting the results of such models.