





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

July 2012 Bangkok, Thailand

Cosimo Beverelli (World Trade Organization)

- a) Censoring and truncation
- b) Tobit (censored regression) model
- c) Alternative estimators for censored regression models
- d) Sample selection models
- e) An example with firm-level analysis

## a) Censoring and truncation

- Censoring
- Truncation

#### Censoring

- We want to estimate the effect of x on a continuous variable y\* (latent dependent variable)
- We always observe x but we observe the dependent variable only above a lower threshold L (censoring from below) or below an upper threshold U (censoring from above)
- Censoring from below (or left):

$$y = \begin{cases} y^* & \text{if } y^* > L \\ L & \text{if } y^* \le L \end{cases}$$

- Example: exports by firm *i* are equal to the export value if the export value exceeds *L*, or equal to *L* if the export value is lower than *L*
- Censoring from above (or right):

$$y = \begin{cases} y^* & if \ y^* < U \\ U & if \ y^* \ge U \end{cases}$$

 Example: recorded exports are top-coded at U. Exports by firm *i* are equal to the export value if the export value is below *U*, or equal to *U* if the export value is above *U*

#### Truncation

- We want to estimate the effect of x on a continuous variable y\* (latent dependent variable)
- Truncation from below (or left):

$$y = y^* if y^* > L$$

- All information below *L* is lost
- Example: exports by firm *i* are reported only if the export value is larger than *L*
- Truncation from above (or right):

$$y = y^* if y^* < U$$

- All information above *U* is lost
- Example: in a consumer survey, only low-income individuals are sampled

### b) Tobit (censored regression) model

- Assumptions and estimation
- Why OLS estimation is inconsistent
- Marginal effects (ME) in Tobit
- Problems with Tobit
- Tobit model with panel data
- Example: academic attitude

#### Assumptions and estimation

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

- This implies that the latent variable is also normally  $\sim : y^* \sim \mathcal{N}(\mathbf{x}'\beta, \sigma^2)$
- We observe:

$$y = \begin{cases} y^* & if \ y^* > 0\\ 0 & if \ y^* \le 0 \end{cases}$$

 Tobit estimator is a MLE, where the log-likelihood function is detailed, for instance, <u>here</u>

#### Why OLS estimation is inconsistent

1. OLS estimation on the sample of positive observations:

$$E[y|\mathbf{x}] = E[y^*|\mathbf{x}, y^* > 0] = \mathbf{x}'\beta + E[\varepsilon|\mathbf{x}, \varepsilon > -\mathbf{x}'\beta]$$

- Under the normality assumption:  $\varepsilon | \mathbf{x} \sim \mathcal{N}(0, \sigma^2)$ , the second term becomes  $\sigma \lambda \left(\frac{\mathbf{x}'\beta}{\sigma}\right)$ , where  $\lambda(\cdot) \equiv \frac{\phi(\cdot)}{\Phi(\cdot)}$  is the **inverse Mills ratio**
- If we run an OLS regression on the sample of positive observations, then we should also include in the regression the term  $\lambda\left(\frac{x'\beta}{\sigma}\right)$
- A failure to do so will result in an inconsistent estimate of  $\beta$  due to omitted variable bias ( $\lambda(\cdot)$  and x are correlated in the selected sub-



