Trade Flows and Trade Policy Analysis

October 2013

Dhaka, Bangladesh

Witada Anukoonwattaka
(ESCAP)

Cosimo Beverelli
(WTO)

Selected econometric methodologies and STATA applications

Content

a) Classical regression model
- Linear prediction
- Ordinary least squares estimator (OLS)
- OLS in STATA
- Interpretation of coefficients

b) Introduction to panel data analysis: fixed effects
- Definition and advantages
- Fixed effects versus random effects models
- Fixed effects (FE) estimator
- Alternative: Brute force OLS (dummy variable estimator)
- FE and dummy variable estimator estimation in STATA

c) Binary dependent variable models
   i. Cross-section
   ii. Panel data
   iii. Binary dependent variable models in STATA

a) Classical regression model

- Linear prediction
- Ordinary least squares estimator (OLS)
- OLS in STATA
- Interpretation of coefficients

Linear prediction

1. Starting from an economic model and/or an economic intuition, the purpose of regression is to test a theory and/or to estimate a relationship

2. Regression analysis studies the conditional prediction of a dependent (or endogenous) variable y given a vector of regressors (or predictors or covariates) $\boldsymbol{x}$, $E[y|\boldsymbol{x}]$

3. The classical regression model is:
   - A stochastic model: $y = E[y|\boldsymbol{x}] + \varepsilon$, where ε is an error (or disturbance) term
   - A parametric model: $E[y|x] = g(\boldsymbol{x}, \beta)$, where $g(\cdot)$ is a specified function and β a vector of parameters to be estimated
   - A linear model in parameters: $g(\cdot)$ is a linear function, so: $E[y|\boldsymbol{x}] = \boldsymbol{x}'\beta$

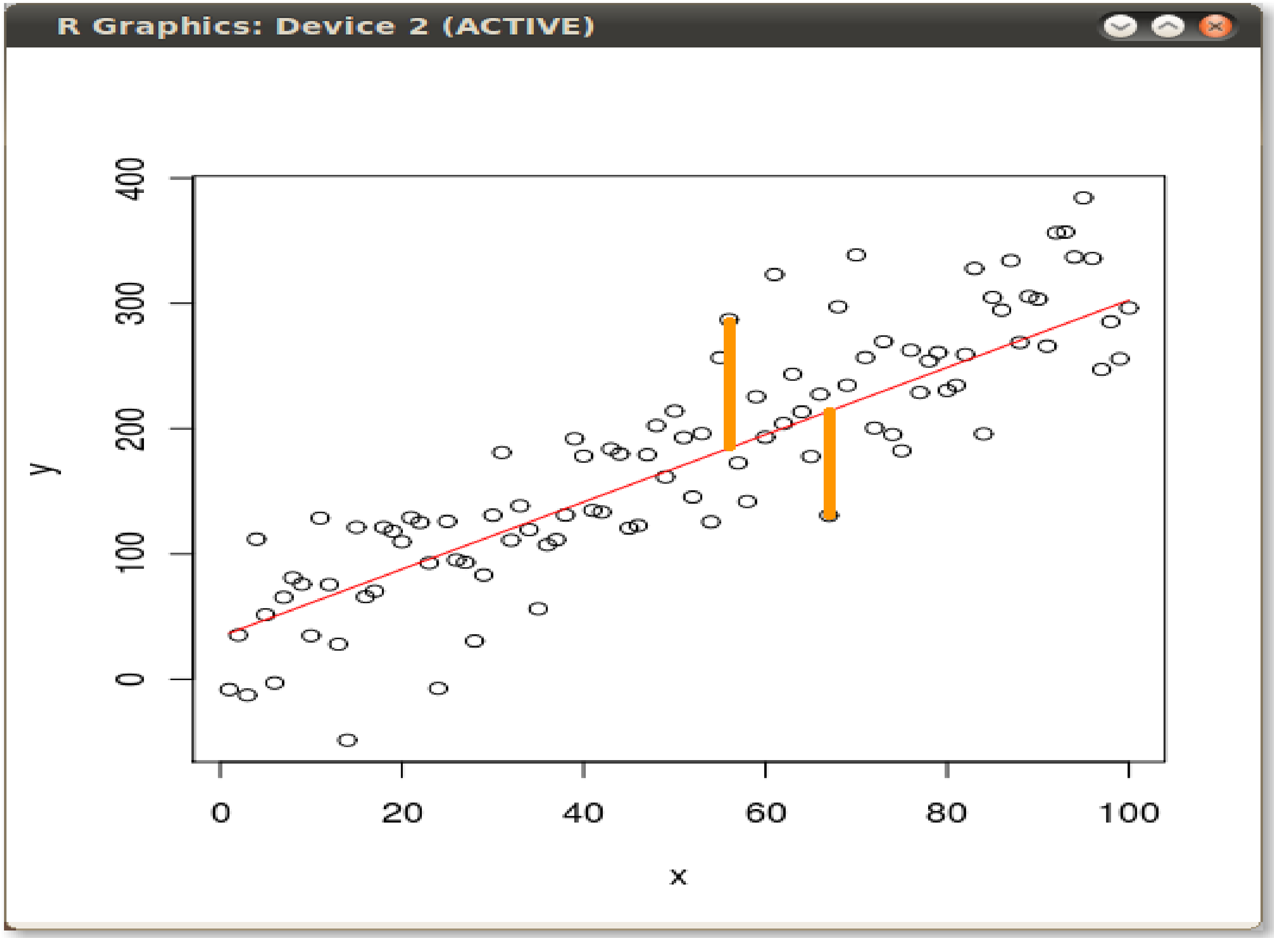
Ordinary least squares (OLS) estimator

- With a sample of N observations ($i = 1, \ldots, N$) on y and $\boldsymbol{x}$, the linear regression model is:
$$y_i = x'_i\beta + \varepsilon_i$$
where x_i is a $K \times 1$ regression vector and β is a $K \times 1$ parameter vector (the first element of x_i is a 1 for all i)
- In matrix notation, this is written as $y = X\beta + \varepsilon$

- OLS estimator of β minimizes the sum of squared errors:
$$\sum_{i=1}^{N} {\varepsilon_i}^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$
which (provided that X is of full column rank K) yields:
$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y = \left(\sum_i x_i x'_i\right)^{-1} \left(\sum_i x_i y_i\right)$$

- This is the best linear predictor of y given $\boldsymbol{x}$ if a squared loss error function $L(e) = e^2$ is used (where $e \equiv y - \hat{y}$ is the prediction error)

Ordinary least squares (OLS) estimator



OLS in STATA

- Stata's regress command runs a simple OLS regression
  - *Regress depvar indepvar1 indepvar2 …., options*

- Always use the option robust to ensure that the covariance estimator can handle heteroskedasticity of unknown form

- Usually apply the cluster option and specify an appropriate level of clustering to account for correlation within groups

- Rule of thumb: apply cluster to the most aggregated level of variables in the model