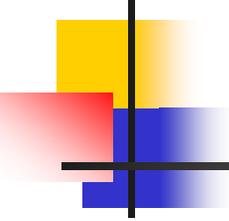# Building a data set for gravity models

Roberta Piermartini
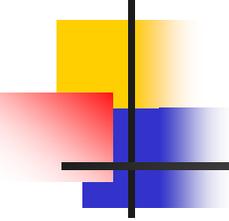
ERSD, WTO

Bangkok, 20 April 2006
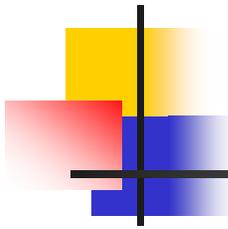
Data sources

- Bilateral trade flows
- Bilateral tariff data
- Bilateral distances
- GDP and population data
- Other sources

Bilateral trade and tariff data

- WITS allows to extract data from:
  - COMTRADE (trade flows)
  - TRAINS (applied bilateral tariffs)
  - IDB (bound and applied tariffs notified to WTO)
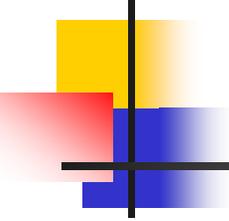- Note the EU appears as a single region

Bilateral distances and bilateral dummies

- CEPII database data are provided in STATA format

http://www.cepii.fr/francgraph/bdd/distances.htm

And explanatory notes are provided

- Including: common border, language (official and ethnical), common coloniser and related variable.
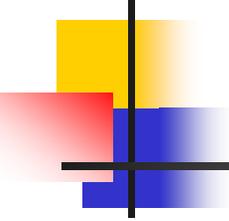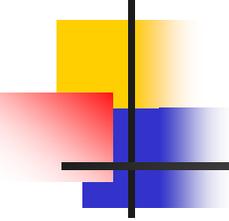
GDP and population

- WDI 2005

Select time series:

  - select countries
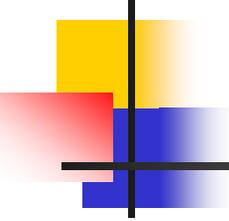 - series
  - periods
  - save as XL

Other data sources

- For institutional variables (World Bank governance indicators)
- For infrastructural variable (WDI)
- Bilateral migration statistics (eg. for the US look at the US Census-USCTIS)

Standard Data transformation

- Check all data are in the same unit of measure
- Log- transformation
- GDP deflation
- Zero trade: 2 ways to treat them

    1. Add 1 and then take logs. Then, use Tobit estimation
    2. Compare exports and imports as in Feenstra. Then, use Helpman et al.
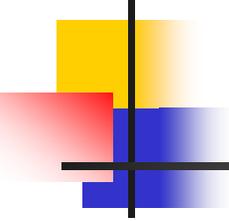
Additional data transformation

- Transformation required by the specific study performed. Some examples include:
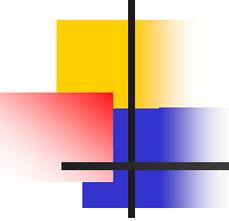 - Calculate remoteness

    $$Rem_j = Sum_k \ distance_{jk}/GDP_k$$

  - Aggregate data for the EU

Introduce data in STATA

- ## Insert XL data in STATA
  - Make sure data cells are numbers
  - Save data in STATA (simply copy and paste XL data in to a new STATA file using Stata editor. Then, preserve)
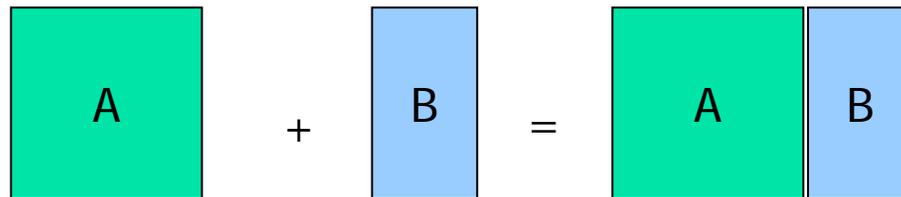 - Save new stata file

Stack data

- To transform a Matrix into a Column use the STATA command

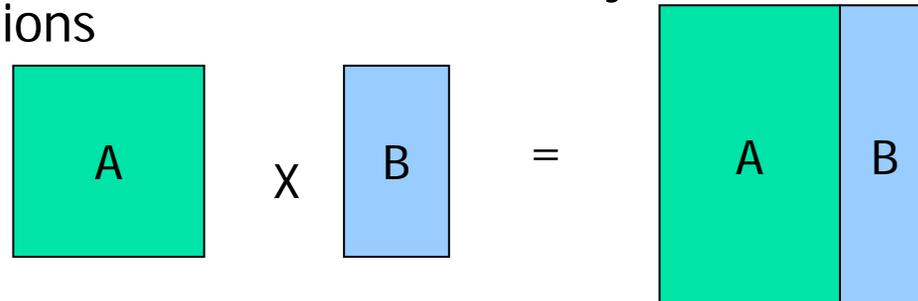 - Stack a b c d e f , into (g h)
  - This will give:

a	b
c	d
e	f

Merge/Join different STATA files

- Merge combines dataset horizontally

- Joint combines datasets horizontally but form all pair combinations

- EX. Join data in CEPII database on distances with WDI database on GDP