

Empirical Exercise 3: Gravity Equation of Trade and Distance

August 18, 2022

Instructor: Marc-Andreas Muendler muendler@ucsd.edu
Teaching Assistant: Carlos Bezerra de Góes cbezerra@ucsd.edu

Due date and time: August 22, 4:50pm

Inputs and products

Please use Stata (any version) for your work. You can base your analysis on the following two original data files

ITPD-E by USITC `itpd.dta`
Gravity data by CEPII `cepii-gravdata.dta`

in the online data folder <https://econweb.ucsd.edu/muendler/teach/22x/101/gen>. To save work time, you may alternatively base your analysis on the ready estimation sample

ITPD-E by USITC and CEPII combined `lec14-sample.dta`

in <https://econweb.ucsd.edu/muendler/teach/22x/101/lec14>.

You may find the code from future Lecture 14 a useful reference: *lec14.do* in the online lecture folder <https://econweb.ucsd.edu/muendler/teach/22x/101/lec14>.

Please submit three products to canvas.ucsd.edu by the due time (**no late submission accepted**): (i) a file with results and your verbal summary titled *ee03.pdf*, (ii) a log file titled *ee03.log*, and (iii) a Stata code file titled *ee03.do*. **Your log file must exhaustively document the steps from the above input files to the output of results.**

Tasks

1. Preliminaries.

- (a) Execute the ready data preparation steps as in future Lecture 14 (*lec14.do*), running only the data preparation code up to line 167.

Hint: The remaining code requires packages `ppmlhdfc` and `est2vec` that you may not yet have installed.

Alternative: To save work time, you may alternatively base your analysis on the ready estimation sample `lec14-sample.dta` and consult the data preparation steps in *lec14.do* as needed.

- (b) From the sample for Lecture 14 (`lec14-sample.dta`), keep only observations with cross-border trade (`src_iso3!=dest_iso3`); keep only observations for the final sample year 2015 when both ITPD-E and CEPII data are available (`year==2015`); and drop the (single) variable `distw` for *Population-weighted Distance (mi)* to avoid any ambiguity with other variable names (`distw_*`).

2. Build up evidence from regressions in five steps.

- (a) Using ordinary least squares, regress *Log Bilateral trade* on *Log Source country production* in a short specification. *Hint:* Consider the command `reg`

- (b) Using ordinary least squares, regress *Log Bilateral trade* on *Log Source country production* in a short specification but include a full set of indicators (dummies) for all source countries. *Hint*: Consider the command `xtreg ... , fe i(src_id)`, where the options `fe i(src_id)` guarantee the inclusion of a full set of indicators.
In one or two sentences, explain the absence of a regression coefficient on *Log Source country production*.
- (c) Using ordinary least squares, regress *Log Bilateral trade* on *Log Source country production* in a short specification but include a set of indicators (dummies) for source countries using the regression command from Lecture 9 (*lec09.do*). *Hint*: Consider the command `reg ... i.src_id`, where the specification `i.src_id` denotes source-country indicators.
In one or two sentences, explain the (awkward) presence of a regression coefficient on *Log Source country production*. *Hint*: Look for the remark “(omitted)” in the Stata output.
- (d) Using ordinary least squares for a traditional gravity equation specification, regress *Log Bilateral trade* on the following predictors: *Log Source country production*, *Log Destination country market size* and a full set of indicators (dummies) for distance categories. *Hint*: The distance categories are in the variables starting with `distw_*` (you can use `desc distw_*` to display the categories to yourself). For the regression, consider the command `reg ... , nocons` where the option `nocons` for no regression constant makes sure that all distance categories are being kept in the regression.
In one or two sentences, explain the findings.
- (e) Using ordinary least squares for a modern gravity equation specification with full sets of source and destination country indicators, regress *Log Bilateral trade* just on the full set of indicators for distance categories and the full sets of source and destination country indicators. *Hint*: From your insights in (2c), you know not to include production or market size variables. For the regression, consider the command `reg ... i.src_id i.dst_id, nocons` where the specification `i.src_id i.dst_id` denotes source-country and destination indicators, and the option `nocons` for no regression constant makes sure that all distance categories are being kept.
In one or two sentences, explain the findings for distance categories as compared to (2d).

3. Make sure to include your one- to two-sentence comments per regression with the pdf file that you submit.