Empirical Exercise 6: Gravity with and without self-trade

October 20, 2020

Instructor: Marc-Andreas Muendler   muendler@ucsd.edu
Teaching Assistant: Jacob Orchard   jdorchard@ucsd.edu

Due date and time: **October 27, 5pm**

Inputs and products

Please use Stata (any version) for your work. You may call any other software from within Stata (including Python, R, Perl, and system-level commands). Please base your analysis on the following files

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

in the online data folder at `https://econweb.ucsd.edu/muendler/teach/20f/435/gen`.

   You may find the code from the previous lecture 4 a useful reference: `lec04upd.do` in the online lecture folder `https://econweb.ucsd.edu/muendler/teach/20f/435/lec04`. **Note the update** from `lec04.do` to `lec04upd.do`.

   Please submit three products to `canvas.ucsd.edu` by the due time: (i) a file with results titled *ee05.pdf*, (ii) a log file titled *ee05.log*, and (iii) a Stata code file titled *ee05.do* (which may call other software). **Your log file must exhaustively document the steps from the above input files to the output of results.**

Tasks

1. Preliminaries.

   (a) Construct two data files, one for trade in goods and one for trade in services. Use the ITPD-E data by USITC, do not remove self trade, keep only the years 2000 and 2015, and aggregate the trade flows to the source country (exporter), destination country (importer), and year level (over all industries).

   (b) Use the gravity data by CEPII at the source country (origin), destination country, and year level, keep only the years 2000 and 2015, and extract the following variables: population-weighted distance, contiguity (shared border), and common official language.

   (c) Similar to Empirical Exercise 5, follow Eaton and Kortum (2002, p. 1761-1762) and construct categories of (population weighted) distance in miles using the CEPII data (transforming the distance measure from kilometers into miles). Assign the distances into the following six categories: [0,375), [375,750), [750,1500), [1500,3000), [3000, 6000), and [6000,maximum) miles.

   (d) Combine (merge) the ITPD-E trade data for goods and services, each type of trade in a separate file, with the CEPII gravity data (including the distance categories) at the source country, destination country, and year level. (Make sure your log file reports the merge results.)

2. Regressions.

   (a) Similar to Empirical Exercises 4 and 5, run two regressions for trade in goods, on a sample that includes zero trade. Decide which of the six distance categories you want to make your reference category (e.g. [0,375) miles) and omit it from all specifications.

    i. For a subsample that excludes self trade, specify a Poisson pseudo maximum likelihood (PPML) regression for trade (not log trade) on five of the six constructed distance category indicators, on an indicator for contiguity, and an indicator for common official language. Condition on source-year and destination-year fixed effects. Make sure your sample includes zero trade flows. Note that the set of regressors for log distance needs to be mutually exclusive and set the distance categories to zero whenever a flow is self trade. Cluster the standard errors at the level of source-destination pairs.

    Hint: In your regression, make sure you specify that there is no constant, otherwise one of the five remaining distance categories could be omitted.

    ii. For the full sample including self trade, specify a PPML for trade (not log trade) on five of the six constructed distance category indicators, on an indicator for contiguity, and an indicator for common official language. Condition on source-year, destination-year, and self-trade country fixed effects. Make sure your sample includes zero trade flows. Note that the set of regressors for log distance needs to be mutually exclusive and set the distance categories to zero whenever a flow is self trade (given self-trade country fixed effects to be included). Cluster the standard errors at the level of source-destination pairs.

    Hint: In your regression, make sure you specify that there is no constant, otherwise one of the five remaining distance categories could be omitted.

(b) Run the same two PPML regression specifications for trade in services, on a sample that includes zero trade (whenever the estimator allows).

3. Interpretation.

- Compare the four specifications. Once the PPML estimator is used and zero trade flow are accounted for here, do you find the inclusion of self-trade to make a strong difference?

References

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741–79.