Entrepot: Dataset Usage Guide
Data Version 1.0

Sharat Ganapati (Georgetown)
Woan Foong Wong (Oregon)
Oren Ziv (Michigan State University)

May 11, 2020

Model Background

Trade does not always travel directly from source i to destination j, it may be routed through a leg from k to l. We construct the cost for containerized trade from k to l, $a_{kl} = t_{kl}^{-\theta}$, up to a dispersion parameter θ to represent idiosyncratic preferences. We then use this to construct the expected trade cost for a shipment from i to j, $b_{ij} \propto \tau_{ij}^{-\theta}$.

We present data on a_{kl} and b_{ij}. Furthermore, this information can be used to construct the probabilities of using various legs to complete a journey from i to j. For example the share of imports (in containerized goods) in destination j from origin i that pass through leg k, l is:

$$\pi_{ij}^{kl} = \frac{\left[(c_i k_{ij})^{-\theta} b_{ik} a_{kl} b_{lj}\right]}{\Phi_j},$$

where c_i is the productivity for source i, Φ_j is a destination fixed effect, and k_{ij} is any origin-destination taste shifter. This can be combined with the unconditional probability for trade

$$\pi_{ij} = \frac{(c_i k_{ij})^{-\theta} b_{ij}}{\Phi_j}.$$

Thus the conditional probability of travel on a route is:

$$\frac{\pi_{ij}^{kl}}{\pi_{ij}} = \frac{b_{ik} a_{kl} b_{lj}}{b_{ij}}.$$

Data Sources

Data on container traffic sourced from Astra Paging (6-months from April to October 2014) using data from Automatic Identification System (AIS) transponders. Aggregate trade data sourced from BACI International Trade Database 2. Land border data sourced from CEPII.

Computation details are in: Ganapati, Wong, Ziv, "Entrepôt: Hubs, Scale, and Trade Costs", CESifo Working Paper No. 8199, 2020.

File Descriptions

- iso_names (csv and xlsx)
  - Description: Crosswalk of country names to ISO codes
  - Columns
    - iso2 - char - 2 digit ISO code
    - iso_name - char - full ISO country name
    - iso3 - char - 3 digit ISO code
- trade_cost_table (csv and xlsx)
  - Description: Full data for both leg trade costs a_{ij} and implied trade costs b_{ij}
  - Columns
    - iso3_i - char - 2-digit ISO code for country i
    - iso3_j - char - 2-digit ISO code for country j
    - a_ij - float - leg trade cost a_{ij}
    - b_ij - float - aggregate route trade cost b_{ij}
    - log_b_ij - float - $\log(b_{ij})$
    - log_x_ij - aggregate observed trade volume $\log(x_{ij})$
- container_trade_table (csv and xlsx)
  - Description: Leg-level trade cost and traffic data
  - Columns
    - iso3_i - char - 2-digit ISO code for country i
    - iso3_j - char - 2-digit ISO code for country j
    - a_ij - float - leg trade cost a_{ij}
    - land_border - integer - indicator for land border from a_{ij}
    - log_predicted_traffic - predicted container traffic from i to j (including implied, but unobserved) traffic on land-crossings)
    - log_actual_traffic - observed seaborne container traffic from i to j
- trade_probabilities
  - Description: Conditional Probabilities of trade using a route:

$$\frac{\pi_{ij}^{kl}}{\pi_{ij}}.$$

    Note that these variables within an i, j pair do not sum to on; as some shipments can use more than one leg on a journey.

  - iso3_i - char - 2-digit ISO code for country i
  - iso3_j - char - 2-digit ISO code for country j
  - iso3_k - char - 2-digit ISO code for country k
  - iso3_l - char - 2-digit ISO code for country l
- a_estimates.png
  - Description: visualization of leg-level trade cost estimates a_{ij}
- b_estimates.png
  - Description: visualization of overall trade cost estimates b_{ij}
- Model_Fit.png
  - Description: visualization of model fit between estimated and actual data on seaborne container shipments