Instructions: figure of shares of sales

May 28, 2020

1 May 16th

Input file

1. "JPV_JIG\Trade\0-Raw_Data\Sales\sales_states.csv.xlsx". The file has 6 columns. The year, the identity of the firm that sells the product "seller", the identity of the firm that buys the product "buyer", the province where the seller lives "state_seller", the province where the buyer lives "state_buyer", and finally, the amount of the sale "amount".

Some notation:

Let's refer to firms (seller or buyer) with lower case letters and to provinces to upper case letters. Let's refer to firm i in province P as $i \in P$. Finally, the amount that firm $i \in P$ sells to firm $j \in P'$ would be $X_{ij,PP'}$ (the first subscript represents the seller, the second one the buyer. Then the third one the province of the seller, and the last one the province of the buyer). The unit of observation would be the combination state_buyer-state_seller. So, for example, PP' is one unit of observation. Since there are 20 provinces, then there are 400 units of observation (20 times 20).

For each unit of observation, compute:

1. The number of sellers in the unit of observation PP'. This is

$$n_{PP'} \equiv \sum_{j \in P'} \sum_{i \in P} \mathbb{I}(X_{ij,PP'} > 0),$$

   where $\mathbb{I}(X_{ij,PP'} > 0) = 1$ if $X_{ij,PP'} > 0$ and zero otherwise.

2. The share of the largest seller in PP'. This is

$$\chi_{PP'} \equiv max \left\{ \frac{\sum_{j \in P'} X_{ij,PP'}}{\sum_{k \in P} \sum_{j \in P'} X_{kj,PP'}} \right\}_{i \in P}$$

Main output

Produce vertical bars graph where $n_{PP'}$ is in the Y-axis. The X-axis would have numbers from 1 to 400. Each number would correspond to one interaction betwee the province of the seller and the one of the buyer (the unit of observation). The numbers in the X-axis would be ordered according to $\chi_{PP'}$. Number one would be the combination PP' with the lowest $\chi_{PP'}$ and number 400 would be the combination PP' with the highest $\chi_{PP'}$.

Note: Since there are many units of observations, the bars should be very thin. I would not mind if it looks like a continuous area (actually, this would be preferred!). We can try also with a line graph instead of the vertical bars. Whatever looks better.

2 New: May 19th

2.1 First

Repeat the same graphs as before but defining

$$n_{PP'} \equiv \frac{\sum_{j \in P'} \sum_{i \in P} \mathbb{I}(X_{ij,PP'} > 0)}{\sum_{j \in P'} \sum_{i \in P} 1}$$

This means that $n_{PP'}$ is **the share** of sellers in PP' that have transactions among each other.

2.2 Second

Change in definition: Let's call $\mathcal{P}$ the province/state and $\mathcal{S}$ as the sector/industry. $X_{ij,\mathcal{SS}'}$ is the amount that firm $i \in \mathcal{S}$ sells to firm $j \in \mathcal{S}'$. The unit of observation now would the the combination of sectors (as opposed to states). For instance, one unit of observation now is $\mathcal{SS}'$. Define $d_{ij,PP'}$ as the distance in km between $i \in \mathcal{P}$ and $j \in \mathcal{P}'$. I uploaded a raw data file with the coordinates of each state (so far this is fake data). So far let's use simple distances (not the formula we used last time). Also, **let's focus on the year 2000 only** (but leave this as a parameter yr< -2000 in case we want to focus on a different year later.

For each unit of observation, compute:

1. The share of the largest seller in $\mathcal{SS}'$. This is

$$\chi_{\mathcal{SS}'} \equiv max \left\{ \frac{\sum_{j \in P'} X_{ij,\mathcal{SS}'}}{\sum_{k \in S} \sum_{j \in S'} X_{kj,\mathcal{SS}'}} \right\}_{i \in \mathcal{S}}$$

   This is the same as before, with the only exception that we care about the combinations of sectors instead of the combination of provinces

2. The distance between each firm (based on the coordinates of their provinces) and the natural logarithm of the distances $\ln(d_{ij,\mathcal{PP}'})$. Define the distance between firms in the same province as equal to one. This is: $d_{ij,\mathcal{PP}'} = 1$ for $\mathcal{P} = \mathcal{P}'$.

3. For firms in $\mathcal{SS}$' (this is $i \in \mathcal{S}$ and $j \in \mathcal{S}'$) run the following regression:

$$\mathbb{I}(X_{ij,\mathcal{SS}'} > 0) = \alpha_{\mathcal{SS}'} + \beta_{\mathcal{SS}'} \ln(d_{ij,PP'})$$

   This would give you one $\hat{\alpha}_{\mathcal{SS}'}$ per $\mathcal{SS}'$. Recall that if a firm $i \in \mathcal{S}$ does not sell to firm $j \in \mathcal{S}'$ we have $X_{ij,\mathcal{SS}'} = 0$.

Main output

Produce the graphs where the estimated $\hat{\alpha}$ for each regression is in the Y-axis. The X-axis would be the same as before but ordered according to $\chi_{\mathcal{SS}'}$.

3 New: May 28th

This goes on a new Rmd. Use as an input file "Trade\0-Raw_Data\Sales\sales_new.csv". No need to save figures in the folder. It is enough to leave them in the code. There are 4 sub-tasks:

3.1 Number of sellers

Create four different histograms with the following statistics (use year 2001 only):

1. The number of sellers per buyer (each buyer appears only once)

2. The weighted number of sellers per buyer (each buyer appears only once but each observation is weighted by the sales "buyer_sales")

3. The number of sellers per buyer in each sector (each buyer appears as many times as the sectors it buys from "sector_seller")

4. The same as 3, but weighing each buyer by the total sales

3.2 How persistent are production linkages?

Calculate the fraction of domestic suppliers (buyers) that are retained between two years (this is called the survival rate). I want 3 numbers per year. Two for buyers (unweighted and weighted), and one for sellers (only unweighted because we do not observe the "seller_sales"). To be more specific, let's take for example buyer 3 between 2000 and 2001. In 2000 it has sellers 7, 36, 42, 39, 46. In 2001 it retains only seller 7, 46, 39. So the fraction of suppliers that are retained is 60% (3 out of 5). You calculate this number for each firm, and each year. Then average them with and without weights.

The figures would have on the X-axis the years and on the Y-axis the fractions. In one graphs the weighted and unweighted average for buyers. And in the other one the unweighted average for sellers. Please add a polynomial fit for each variable.

3.3 How does the persistence change with firm size?

Here we will focus on creation and survival rates of firm's production links. Creation rates are the fraction of links at a given moment in time that do not appear the previous year for each firm (new links). Here we want to see whether the rates change with firm size. We would only focus on buyers. Two figures: both would have on the X-axis the natural log of buyer sales and on the Y-axis one would have the creation rate and the other one the survival rate. Also add the polynomial fit.

3.4 Duration of the links

call year 0 the first year two firms start a relationship. The idea here is to have a figure that tells us the share of links that remain after one year (year 1), then the share that remain after two years, and so on until year 4. By construction, the share that remain in year 0 is 100%. Let's use only the relations that start before 2005. So the graph would have on the X-axis the year (starting from year 0 until year 4) and on the Y-axis the share.