Figure Replication for "Cities, Heterogeneous Firms and Trade" Memo 20230223

Summary

- Previously, I got a negative correlation between city size and export intensity (exports / sales). I then realized that the sales table contains sales for each 2-digit, 3-digit, 4-digit, 5-digit and 6-digit NAICS, so I was overcounting sales.
- After only keeping 2-digit NAICS and adding up the sales in all sectors, I got a slight positive correlation. When I only keep the manufacturing NAICS (31-33), I got a higher positive correlation that resembled the reported plot more, though the actual data points are not identical.

Table 2: Export intensity

- Table 2 said that export intensity for the US is overall exports / manufacturing sales. However, on page 7, it says that "Consequently, city-level export intensity is constructed as overall exports over sales across all sectors." I find that using only manufacturing sales matches up better with the reported data.
- The reported numbers seem strange and inconsistent even with Figure 1. For example, the reported mean is 5.61, or 561%. Taking the natural log of 5 gives 1.6, which is already off the scale of Figure 1.

Table 1: City size (population)

- The numbers for population are not identical, but they are at least somewhat close.
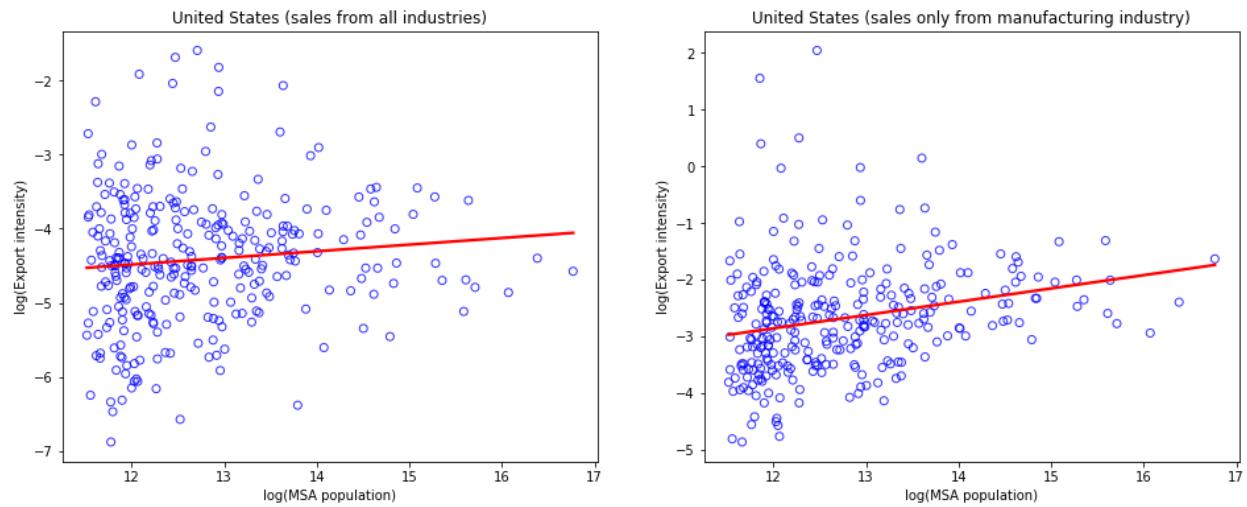
What I got (obtained, using manufacturing sales) vs what was in the paper (reported):

	export_intensity_2012_obtained	export_intensity_2012_reported		pop_2012_obtained	pop_2012_reported
count	298.00000	324	count	298.0	324
mean	0.14763	5.61	mean	734.0	799.3
std	0.54270	NA	std	1686.5	NA
min	0.00767	NA	min	100.3	NA
25%	0.03499	2.51	25%	153.3	158.4
50%	0.06513	4.0	50%	262.3	275.2
75%	0.11355	6.61	75%	595.4	636.6
90%	0.20209	10.2	90%	1446.0	1929.2
95%	0.32141	15.2	95%	2665.9	3176.1
max	7.71054	NA	max	19149.7	NA

Figure 1

- The paper says there are 324 MSAs, but I counted only about 70 circles in the reported plot. There is a visual difference in terms of density of circles between the plot I got (with 298 points) and the plot that was reported (with supposedly 324 points), which is strange.
- Here I show both plots, one using sales = total sales across all sectors, and the other using sales = manufacturing sales only.

Obtained:

Reported: