Empirical IO

PS2: Production function estimation

Hans Martinez

Apr 12, 2021

Introduction

For this assignment, I hard-coded everything from scratch in `R`. It didn't cross my mind that someone might have already coded this estimation procedure. In any case, this allowed me to use in the first stage a nice tool from Machine Learning (ML), called Multivariate Adaptive Regression Splines (MARS). Intuitively, this algorithm creates automatically a piecewise linear model. MARS performs well in comparison to other non-parametric non-linear interpolation routines. The advantage is that the algorithm splits the variable space optimally. That is, it searches over the best possible partition of the variables fed into it.

However, one of the challenges to code everything from scratch was the estimation of the standard errors. These are usually bootstrapped as in Levinsohn and Petrin, and discussed by Ackerberg et al. The challenge is to bootstrap from panel data respecting the structure. In our case, it is also important to consider the assumption on the productivity's markovian process. One suggestion is to bootstrap by block. Alternatively, we can also do *jackknife* resampling, which is what I ended up doing. In the *jackknife* procedure, I removed randomly one firm with all its time observations per resampling iteration. I resampled 100 times, then I estimated the standard errors from the distribution of coefficients. Ideally, I would have liked to jackknife over the entire sample, not just over 100 randomly selected firms. However, the procedure was taking too long.

This report and the code can be found in my GitHub repo.

ACF production function estimation

1. I used the ACF estimation procedure using intermediate input as the proxy variable estimator, assumed a Cobb-Douglas functional form in logs, and took the log of the real gross domestic output as the production function to estimate. Table 1 displays the coefficient estimates with their standard errors per industry; table 2 the summary statistics about the distribution of productivity. Figure 1 shows the distribution of the four industries for graphical comparison.

2. In this second exercise, I drop all firms with zero investment. Table 3 displays the estimates of the coefficients, table 4 the summary statistics of the productivity, and figure 2 contrast the original estimates for the full sample and without zero-investment firms.

   Discussion: In general, we can appreciate that the firms that do invest have higher productivity mean than the full sample. However, the productivity dispersion seems to be greater for these firms. In particular, for the Textile industry (321), the change is highly significant. The dispersion and the mean are notoriously higher. In comparison, the change is not so significant for the Food

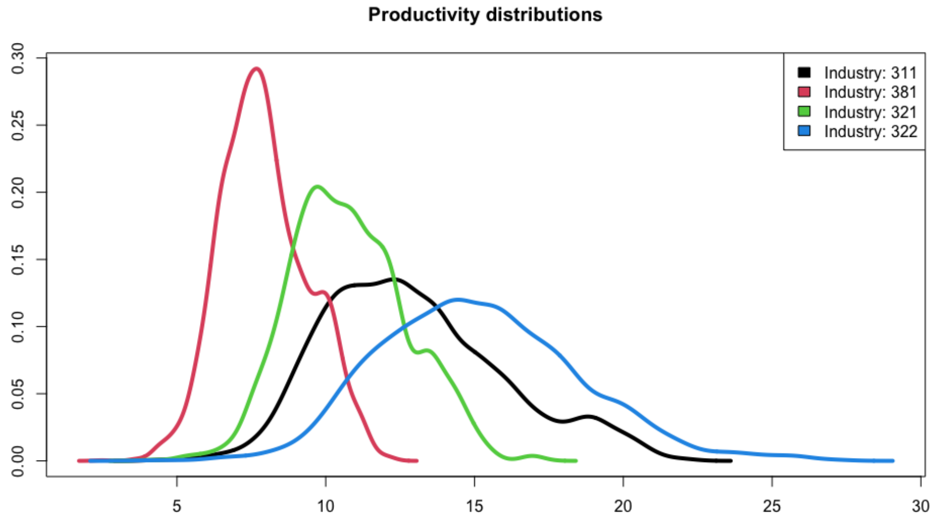

Figure 1: Productivity distributions per industry.

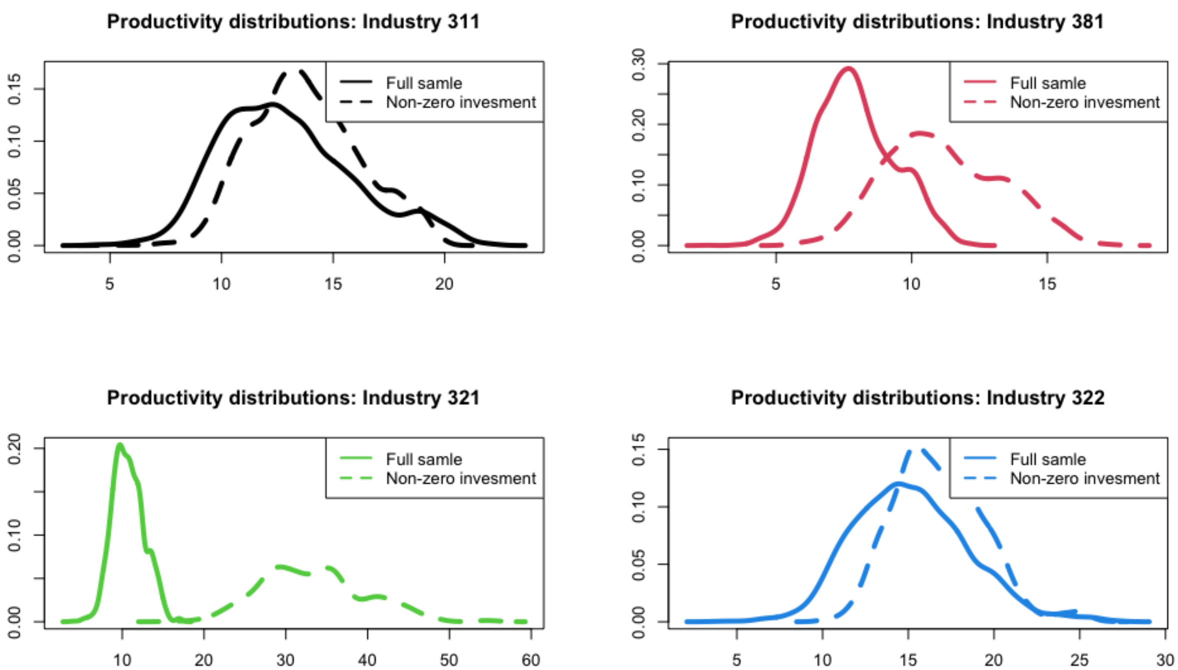

Figure 2: Productivity distributions per industry comparing full sample and without zero-investment firms.

Table 1: ACF production function estimation [a]

	Industry 311	Industry 381	Industry 321	Industry 322
Log K	0.3590	0.4409	0.1789	0.5578
S.E.	0.0019	0.0006	0.0019	0.0006
Log L	-1.4768	-0.4720	-0.4278	-2.6441
S.E.	0.0005	0.0008	0.0005	0.0008

[a] Coefficient estimates and standard errors calculated using jackknife resampling with 100 iterations.

Table 2: Summary statistics of the productivity distribution

	Industry 311	Industry 381	Industry 321	Industry 322
Min	4.3157	2.4795	3.8215	3.7769
Q1	10.6732	6.8976	9.3796	12.7645
Mean	12.5520	7.7932	10.6148	14.9202
Q3	14.8167	8.9107	12.0396	17.2562
Max	22.1834	12.2850	17.3285	27.3638
SE	0.0386	0.0247	0.0370	0.0510
Obs	6,129.0000	3,470.0000	2,745.0000	4,251.0000

Table 3: ACF without zero-investment firms[a]

	Industry 311	Industry 381	Industry 321	Industry 322
Log K	-0.0762	0.3480	-1.4939	-0.2300
S.E.	0.0007	0.0018	0.0007	0.0018
Log L	-0.4357	-1.0040	-1.6894	-1.2066
S.E.	0.0006	0.0009	0.0006	0.0009

[a] Coefficient estimates and standard errors calculated using jackknife resampling with 100 iterations.

Products industry (311). This result is in line with the economic intuition. We would expect firms that invest to have higher productivity. Likewise, we can imagine that there are some industries for which it is more relevant to keep investing than for others. For example, constant capital investment is likely to be relevant to keep productivity high for manufacturing plants in the Textile industry. New toolings for new fashion designs might be needed every season. In contrast, this is not likely to be the case for the firms in the Food products industry, where perhaps adapting the process for a new flavor might not need capital spending.

3. Lastly, I split the data into exporting firms and non-exporting firms, and importing and non-importing firms. In the international trade literature, it is often found that exporting and importing firms have higher productivity. Intuitively, for exporting firms this must be the case because they have to be competitive enough to be attractive for foreign consumers after netting off transportation costs and trade barriers, such as exporting/importing tariffs, for example.

   The coefficient estimates for the different samples are displayed in tables 5 and 6. The summary statistics of the productivity distributions are shown in tables 7 and 8. Finally, figures 3 and 4 show the contrast between the two samples within industries.

Table 4: Summary statistics productivity distribution without zero-investment firms

	Industry 311	Industry 381	Industry 321	Industry 322
Min	4.7699	5.6562	15.6180	9.9556
Q1	12.0594	9.5896	28.7046	14.9265
Mean	13.5843	10.9532	32.7905	16.5591
Q3	15.3504	12.7642	36.9017	18.5391
Max	20.0129	17.5568	55.6406	26.3343
SE	0.0367	0.0430	0.1426	0.0514
Obs	4,313.0000	2,436.0000	1,949.0000	2,623.0000

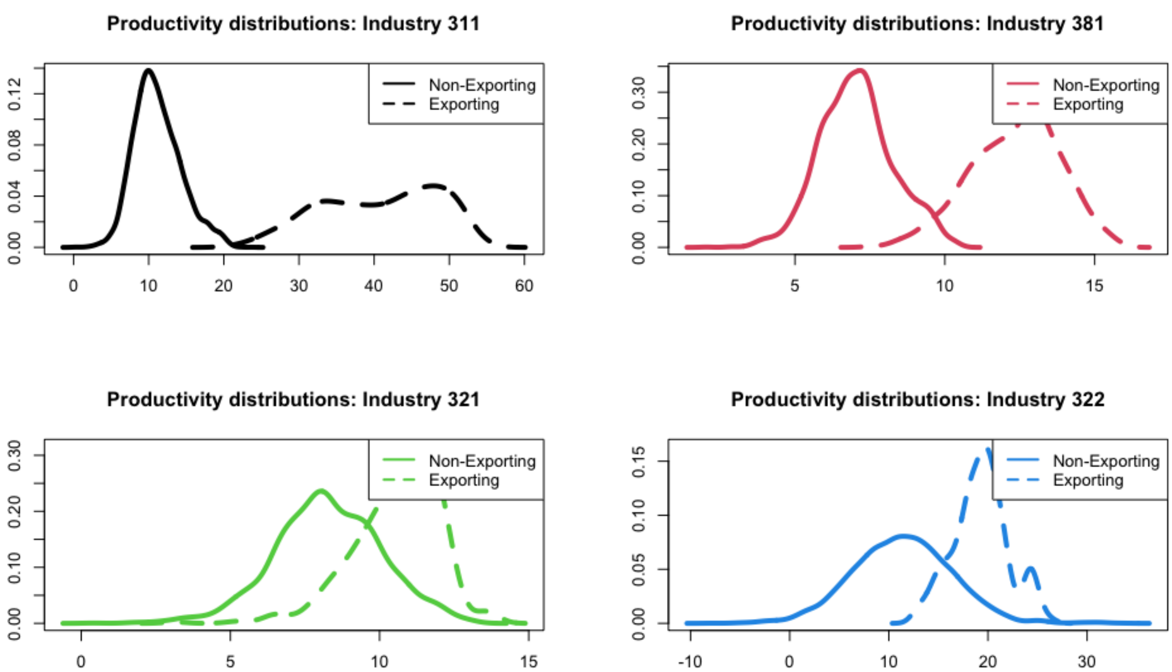

Figure 3: Productivity distributions for non-exporting and exporting firms

Table 5: ACF with and without exporting firms[a]

	Industry 311	Industry 381	Industry 321	Industry 322
Non-Exporting				
Log K	0.8724	0.5053	0.6700	2.5094
S.E.	0.0025	0.0015	0.0025	0.0015
Log L	-2.3451	-0.4613	-1.0396	-5.8614
S.E.	0.0004	0.0010	0.0004	0.0010
Exporting				
Log K	-1.1121	0.0798	0.1442	-0.9338
S.E.	0.0005	0.0033	0.0005	0.0033
Log L	-2.8645	-0.2877	0.0265	0.0734
S.E.	0.0004	0.0010	0.0004	0.0010

[a] Coefficient estimates and standard errors calculated using jackknife resampling with 100 iterations.

Table 6: Summary statistics productivity distribution without zero-investment firms

	Industry 311	Industry 381	Industry 321	Industry 322
Non-Exporting				
Min	0.0628	2.0423	0.4142	-7.8321
Q1	8.9733	6.2008	7.1034	8.0863
Mean	10.7971	7.0147	8.2360	11.3541
Q3	13.0990	7.7871	9.4941	14.6292
Max	23.7292	10.5252	13.2261	33.7358
SE	0.0419	0.0230	0.0377	0.0809
Obs	5,644.0000	2,895.0000	2,268.0000	3,809.0000
Exporting				
Min	21.9430	7.6502	3.1436	12.3177
Q1	34.0786	11.3153	9.7293	17.6037
Mean	41.9820	12.5532	10.8762	19.4020
Q3	47.5792	13.4106	11.6322	20.9199
Max	54.0033	15.7004	13.7670	26.3939
SE	0.3536	0.0617	0.0670	0.1332
Obs	485.0000	575.0000	477.0000	442.0000

Discussion: Overall, exporting and importing firms do have higher productivity means than non-exporting and non-importing firms, respectively. However, now the dispersion of the productivity differs across industries and between the importing/exporting condition.

For the exporting firms, we observe higher mean productivities consistently across industries. Notice, however, that the dispersion looks higher for all except for the Apparel industry (322), which seems to be tight around the mean than the non-exporting firms.

For the importing firms, we observe that these firms have higher productivity means and higher dispersion consistently for Fabricated Metals (381) and Textiles (321) industries. The exceptions are Food Products (311) and Apparel (322) for which we have weird results. This could point out maybe some code error. Something I could have done to improve if I had more time would be to use the jackknife estimators, instead of the whole sample estimators.

Table 7: ACF with and without exporting firms[a]

	Industry 311	Industry 381	Industry 321	Industry 322
Non-Importing				
Log K	0.2951	0.9665	0.3988	14.3288
S.E.	0.0018	0.0016	0.0018	0.0016
Log L	-2.4163	-1.4109	-0.7044	-63.1266
S.E.	0.0009	0.0014	0.0009	0.0014
Importing				
Log K	208.2941	0.1388	-0.6270	0.1416
S.E.	0.0028	0.0023	0.0028	0.0023
Log L	-344.4136	-0.1520	-1.2774	0.1471
S.E.	0.0003	0.0006	0.0003	0.0006

[a] Coefficient estimates and standard errors calculated using jackknife resampling with 100 iterations.

Table 8: Summary statistics productivity distribution without zero-investment firms

	Industry 311	Industry 381	Industry 321	Industry 322
Non-Importing				
Min	4.5968	-0.3772	2.4110	-68.5774
Q1	13.4028	5.0612	8.1371	92.2393
Mean	15.7995	6.1767	9.1463	126.3039
Q3	18.5669	7.2316	10.2730	160.0491
Max	28.8470	12.6623	13.8702	329.9741
SE	0.0519	0.0341	0.0359	0.8140
Obs	5,338.0000	2,494.0000	2,046.0000	4,033.0000
Importing				
Min	-1,249.7336	7.5838	14.9518	5.7531
Q1	-802.0182	9.8969	24.1100	8.7140
Mean	-600.7986	10.9918	26.3159	9.3764
Q3	-398.0652	11.7838	29.0798	9.9396
Max	378.3056	13.8763	37.7022	11.8683
SE	9.5581	0.0411	0.1428	0.0818
Obs	791.0000	976.0000	699.0000	218.0000

Final comments

It is interesting to note that the ratio of firms that export/import versus non-export/import is very low, around 10%. Consider the number of firms that invest, where the ratio is almost 50%. If so many firms invest, why so few get to trade with foreign markets? Why the production is not reallocated to the most competitive firms and we observe a high dispersion of the productivity distribution?

Also, it would be interesting to have geospatial data and data about the demand side. Does the spatial distribution help explain the wide dispersion in productivity? Maybe there are more dense demand markets for some industries than for others? Furthermore, information about the history of each industry in the country and on regulation will help understand these differences. Are the export/import tariffs the same for every industry? Are some industries relatively new or with a long tradition in the country? Likewise, on the firms that trade with foreign markets,

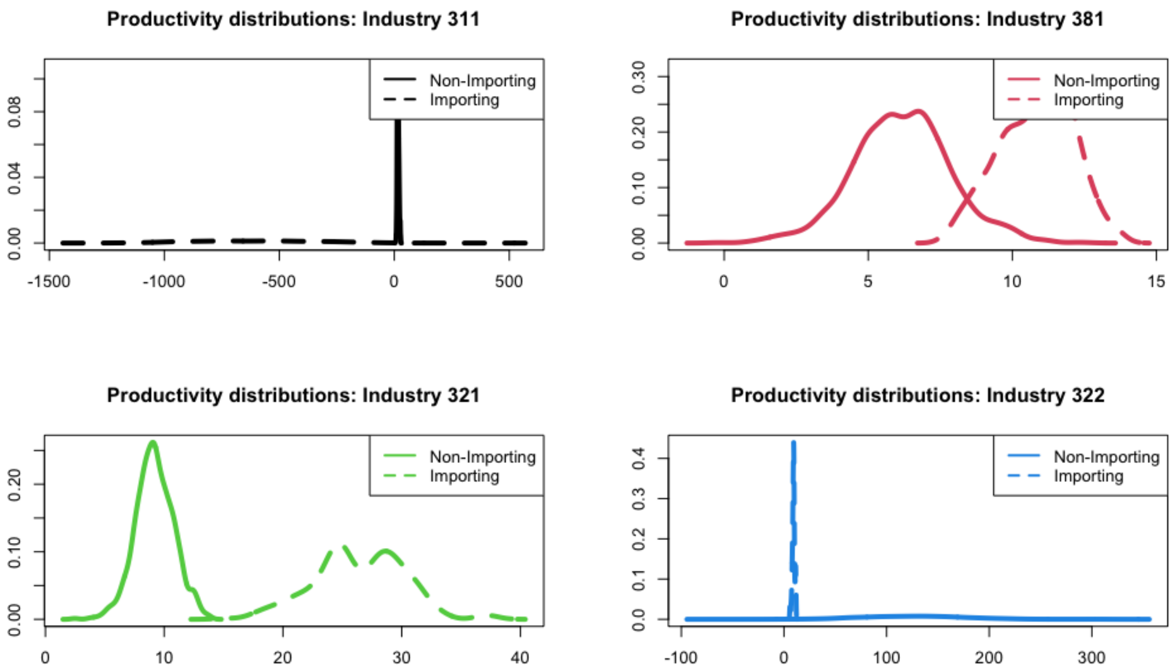

Figure 4: Productivity distributions for non-importing and importing firms

information about foreign partners, like which countries do firms buy from or sell to.

Among other things I would have loved to try is to compare my estimates with the pre-coded packages available in `R` or `Stata`. I'm curious to see if there are any gains in using MARS over other non-parametric techniques. I would also like to compare the bootstrap versus the jackknife results. Finally, I would have liked to explore other functional forms of the production function.