
Trade Data Cleaning Exercise

LUCA POLL

October 11, 2020

Excercise 1

For each vintage of the Baci dataset, identify relevant exporters and importers list relative to Germany and provide summary statistics of yearly trade volumes in constant euros of 2005 during the period 2005 to 2010 : China, Main export partners, Main import partners.

Trade Flows by years

Table 1-3 display some selected descriptive statistics of the trade flows for the requested groups contrasting imports (to Germany) and exports (from Germany). Note that the number of countries in the two groups are not proportional (21 import partners and 9 (without China) export partners), which leads to not directly comparable numbers of transactions per year for the different groups.

Table 1: Descriptive Statistics for China per year

year	Imports				Exports			
	n	median	sd	range	n	median	sd	range
2005	3858	454.3	81008.8	2909125.3	3839	225.2	30454.5	1209055.0
2006	3943	615.0	107305.5	4145244.7	3910	252.5	41017.5	1580061.3
2007	3844	830.4	97084.1	4572372.5	3741	316.0	43755.5	1578316.5
2008	3835	898.2	100496.4	4654997.7	3794	331.4	50913.4	2043873.1
2009	3875	729.0	90312.8	3503257.0	3760	302.1	58784.6	2450173.0
2010	3886	941.4	148831.0	5340747.2	3781	408.8	110868.5	5528994.7

Table 2: Descriptive Statistics for main export partners (excl. China) per year

year	Imports				Exports			
	n	median	sd	range	n	median	sd	range
2005	35068	464.8	72777.3	7057950.3	36843	1052.7	105830.1	10056478.6
2006	35009	494.7	79606.1	7630115.5	36907	1105.0	109822.0	9656847.8
2007	33560	543.1	70096.8	5746635.9	35005	1286.6	112200.9	10548326.0
2008	33578	533.3	84014.2	8789825.1	34967	1259.5	105464.2	10607698.8
2009	33478	445.4	67999.8	5979074.3	34887	1033.2	83801.2	8447285.5
2010	33475	514.4	89263.1	9382079.1	34880	1144.3	98896.5	7826764.5

From Table 1 we can observe that the number of imports and exports between Germany and China was relatively steady. The median of the CPI adjusted values of the respective

Table 3: Descriptive Statistics for main import partners per year

year	Imports				Exports			
	n	median	sd	range	n	median	sd	range
2005	27703	58.7	56243.6	8557647.3	58898	73.6	12086.7	1509357.2
2006	28212	62.0	64291.3	10017904.1	60770	79.3	14203.8	1681975.8
2007	27301	71.1	61533.1	9209686.8	60112	92.3	15384.9	1550520.9
2008	27494	69.0	69641.8	10746406.3	60619	93.0	15509.0	1331534.1
2009	27643	59.6	47909.1	7111698.0	60091	69.8	10327.1	621308.2
2010	26959	75.6	63263.8	9361663.5	59615	82.0	12449.1	848417.5

transactions in EUR, on the other hand, increased over the years for imports and exports. This might be due to increasing quality of the goods or a shift towards the exchange of more developed goods.

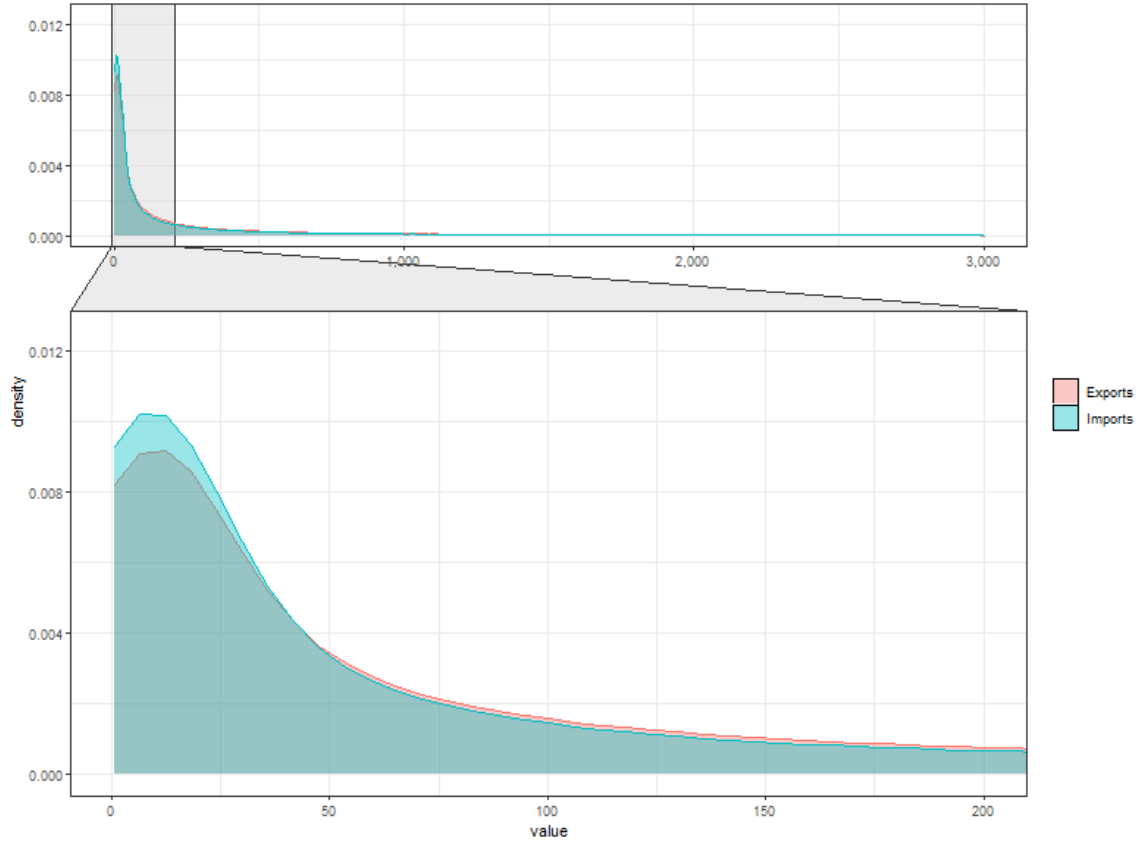
If we compare Table 2 and Table 3 with each others, we can observe at the first glance that the number of imports and exports for the main export partners (without China) are relatively close. The main import partners exhibited less import transactions with lower median values and significantly more exports than the export partners. Figure 1 shows more accurately the distribution of the values for import and export transactions of importing partners. We can see, that given the group, imports are more likely to have lower values than exports which is in line with the histogram in Figure 5 (Appendix) and the overall summations of trade interactions¹, from which we can observe clear differences in the aggregated values across the two groups.

Trade Flows by HS Sections

Besides considering the development of the trade flows over the years, it is also interesting to have a closer look at the composition of the imports and exports. This can be done by considering the HS sections (overview in Appendix Table 9) as reported in Table 7 and 8. One can sport that the most frequent transactions in all of the subgroups occur in the same sections:

¹Import partners Imports: 565,426,515 EUR
Import partners Exports: 678,129,498 EUR
Export partners (no China) Imports: 1,764,461,591 EUR
Export partners (no China) Exports: 2,475,083,478 EUR

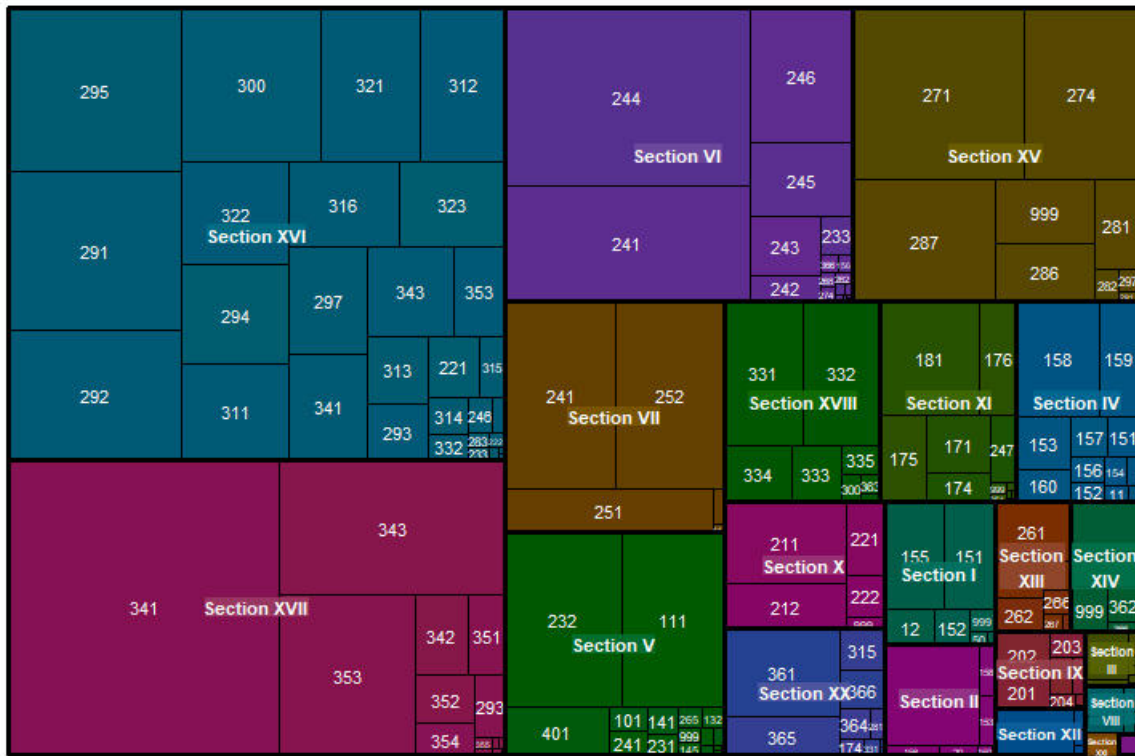
Figure 1: Density of values of trade flows for import partner countries



- ▷ Section XVI: Machinery & similar
- ▷ Section XI: Textiles and textile articles
- ▷ Section VI: Products of chemical or allied industries
- ▷ Section XV: Base metals and articles of base metals

If we do not consider the frequencies, but the aggregated overall values as in Figure 2, we find that the sections with the highest trade frequencies are not necessarily the ones with the largest aggregated trade flow (imports + exports). Section XVII (Vehicles, air crafts,...) and section VII (Plastics, rubbers,...) appear to have lower frequencies but far higher median values which leads to a considerably high aggregated overall value. The most "valuable" industries are the WZ classification industries 341 and 244.

Figure 2: Treemap of aggregated trade volumes by HS section and WZ classification



Exercise 2

For each vintage year of the Baci data, list the 5 largest export industries and the 5 largest import industries for Germany.

Table 4 reports the largest import and export industries per year by net exports. While industry 341 is the largest net export industry throughout the six years, industry 111 is constantly the largest net import industry.

Table 4: Largest five import and export industries

		Export					Import				
year		1st	2nd	3rd	4th	5th	5th	4th	3rd	2nd	1st
2005	Industry	341	295	343	244	292	181	353	300	11	111
	Net exports	45,023,184	12,543,764	12,120,623	10,458,753	8,914,788	-4,037,703	-4,582,893	-4,597,189	-5,110,853	-11,133,609
2006	Industry	341	343	244	295	292	181	11	300	353	111
	Net exports	45,451,011	14,104,814	13,752,792	13,260,810	9,432,246	-3,782,089	-5,138,525	-5,819,178	-6,691,685	-10,674,352
2007	Industry	341	343	244	295	292	321	181	300	11	111
	Net exports	51,149,098	14,787,590	13,928,938	13,356,262	10,266,109	-2,006,810	-3,498,909	-4,254,446	-4,439,993	-14,462,158
2008	Industry	341	244	295	343	292	232	181	300	11	111
	Net exports	48,776,675	15,781,234	13,388,093	11,920,936	10,704,491	-2,813,854	-3,462,169	-4,901,910	-5,294,548	-16,630,128
2009	Industry	341	244	295	291	292	323	181	11	300	111
	Net exports	29,907,095	11,486,605	9,145,245	7,988,552	7,922,368	-2,976,073	-3,487,433	-4,749,210	-4,871,635	-9,748,666
2010	Industry	341	343	295	244	291	321	11	300	232	111
	Net exports	44,458,200	10,982,827	10,336,960	9,913,309	9,908,141	-4,330,983	-4,983,193	-7,159,531	-7,304,669	-14,556,201

5

Table 5: Panel Dataset

country	t	wz	hs6	import	impcount	v	hs92_labels	temp_weight	hs2section
31	2005	11	60220	0	1	3.2	Trees,shrubs,bushes,grafted or not,of kinds whi	1.0	Section II
31	2005	11	60299	0	1	4.5	Other live plants, nes	1.0	Section II
31	2005	11	70110	0	1	18.5	Seed potatoes	1.0	Section II
31	2005	11	80222	1	1	19655.4	Hazlenuts without shells, fresh or dried	1.0	Section II
31	2005	11	80232	0	1	0.8	Walnuts without shells, fresh or dried	1.0	Section II
31	2005	11	80410	0	1	1.6	Dates, fresh or dried	1.0	Section II
31	2005	11	81310	0	1	0.8	Dried apricots	1.0	Section II
31	2005	11	81320	0	1	2.4	Dried prunes	1.0	Section II
31	2005	11	90111	0	1	24.0	Coffee, not roasted or decaffeinated	1.0	Section II
31	2005	11	90240	0	1	15.0	Black tea (fermented) and partly fermented tea,	1.0	Section II

Exercise 3

Generate a panel of Germany's imports and exports towards its main partner between 2005 and 2010.

The structure of the panel looks like presented in Table 5. In order to create a data frame with the right amount of observations, duplicates from the "cw_HS_WZ93_3_final2.dta" that were not detected by the common duplicate detection functions² had to be localized and deleted. If not, 204015 additional observations were created.

Exercise 4

Plot aggregate time series exports and imports volume of Germany with its main partners between 2005 and 2010.

Figure 3 shows the aggregated import and export values per year for Germany's main import and main export partners. From the graph, we can clearly see that the yearly aggregated trade volume with the export partners is larger than with the import partners. Furthermore, it is evident that the exports from Germany to its trading partner countries are of larger value than the imports from Germany's partner countries to Germany. If we disaggregate the partner groups, we are left with Figure 4. Focusing on trading partner countries with large aggregated volumes, one can see that across all years Germany exported the most to France, followed by the US, followed by the UK and Italy. The largest volume of imports came from 2009 onward not from France but from China, which increased the volume of exports to Germany (so imports for Germany) by roughly 25 mio. EUR in just five years.

²like `distinct(.)` or `unique(.)`

Figure 3: Aggregated trade volumes by type of partner and import/export

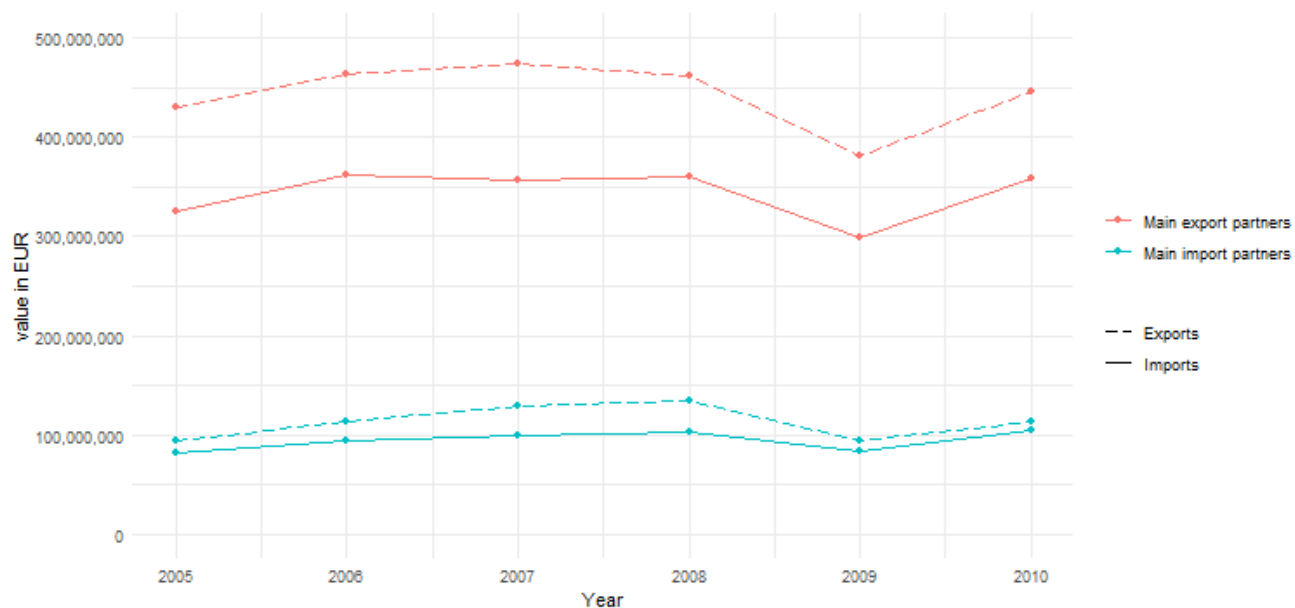
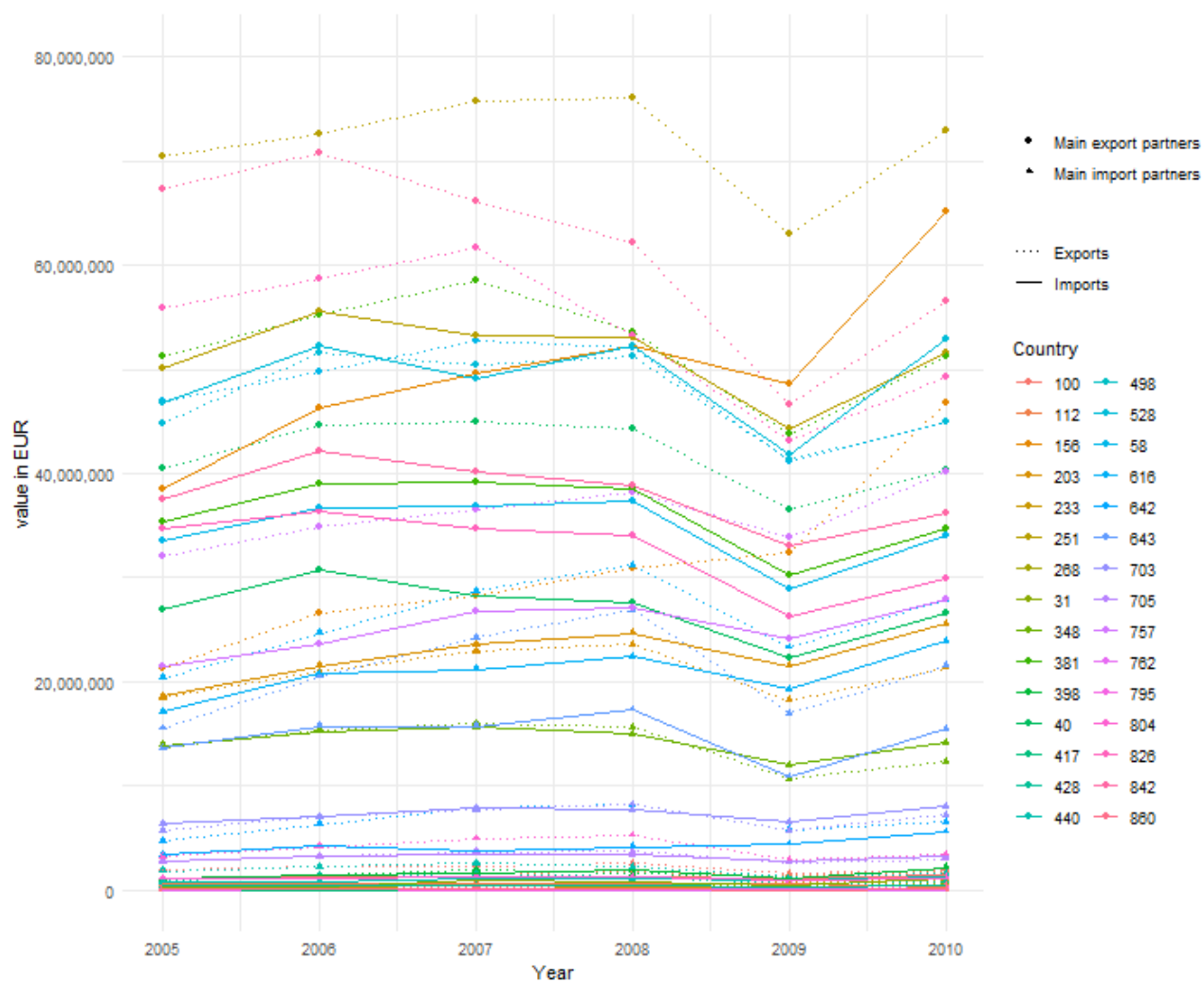


Figure 4: Aggregated trade volumes by country and import/export



Exercise 5

Estimate the correlation between trade volumes (exports and imports separately) on Germany's GDP when controlling for industry and origin (or destination) country effects.

Since the question asks for correlation and not causation, I choose to specify a pooled OLS model:

$$\mathbf{y} = \widetilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \varepsilon$$

which corresponds to the stacked version of

$$\log(gdp)_{it} = \alpha + \beta_1 \log(value)_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, \dots, 30, t = 2005, \dots, 2010$$

where $\log(gdp)_{it}$ is the dependent variable³, the log of Germany's GDP, $\log(value)$ the value of the transaction as a measure for the trade volume. \mathbf{x}_{it} a vector containing dummy variables for the respective country and the industry. α is the intercept and ε_{it} the error term. If the error terms are strictly exogenous and homoskedastic, if there is no serial correlation, ε_i and ε_j independent for $i \neq j$ and there is no multicollinearity, the pooled OLS (POLS) estimator is BLUE:

$$\hat{\tilde{\boldsymbol{\beta}}}_{POLS} = \left(\widetilde{\mathbf{X}}' \widetilde{\mathbf{X}} \right)^{-1} \widetilde{\mathbf{X}}' \mathbf{y} = \left[\sum_{i=1}^N \sum_{t=1}^T \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}'_{it} \right]^{-1} \sum_{i=1}^N \sum_{t=1}^T \tilde{\mathbf{x}}_{it} y_{it}$$

where $value$ is included in the matrix $\widetilde{\mathbf{X}}$. However, since the assumptions are most likely violated in the proposed model, the standard errors are wrong and spurious correlation hinders us from identifying underlying causality. In order to find the fitting model, one would consider specifying a random effects or a fixed effects model and conduct a Hausman test in order to decide for the correct specification. Efficient estimators depend on the context, but most likely the Within Estimator, the (Conditional) Maximum Likelihood or a FGLS Estimator would be the the best choice. Furthermore, if we would have wanted

³the logs were taken to normalize the observations

to establish a causal relationship, the limited variation in the dependent variable (only six different values) would have been concerning as the model would exhibit relatively low external validity and problems of multicollinearity.

Table 6: Correlation trade volumes and Germany's log GDP

	Imports	Exports
Constant	28.468*** (0.001)	28.468*** (0.001)
log(value)	0.0002*** (0.00001)	0.0004*** (0.00001)
Observations	392,721	596,419
R ²	0.001	0.002
Adjusted R ²	0.0003	0.001
F Statistic	1.937***	6.857***

Note: *p<0.1; **p<0.05; ***p<0.01

The individual coefficients for the categorical variables "country" and "industry" are not reported

From the correlation results in Table 6 we see that a percent change in the trade volume has a larger correlation for exports than for imports with an accompanying percentage change in Germany's GDP. This might be due to the difference in size and importance of imports and exports for the German GDP.

Appendix

Other interesting graphical descriptive statistics

Figure 5: Histogram trade volume by importers/exports for import partner countries

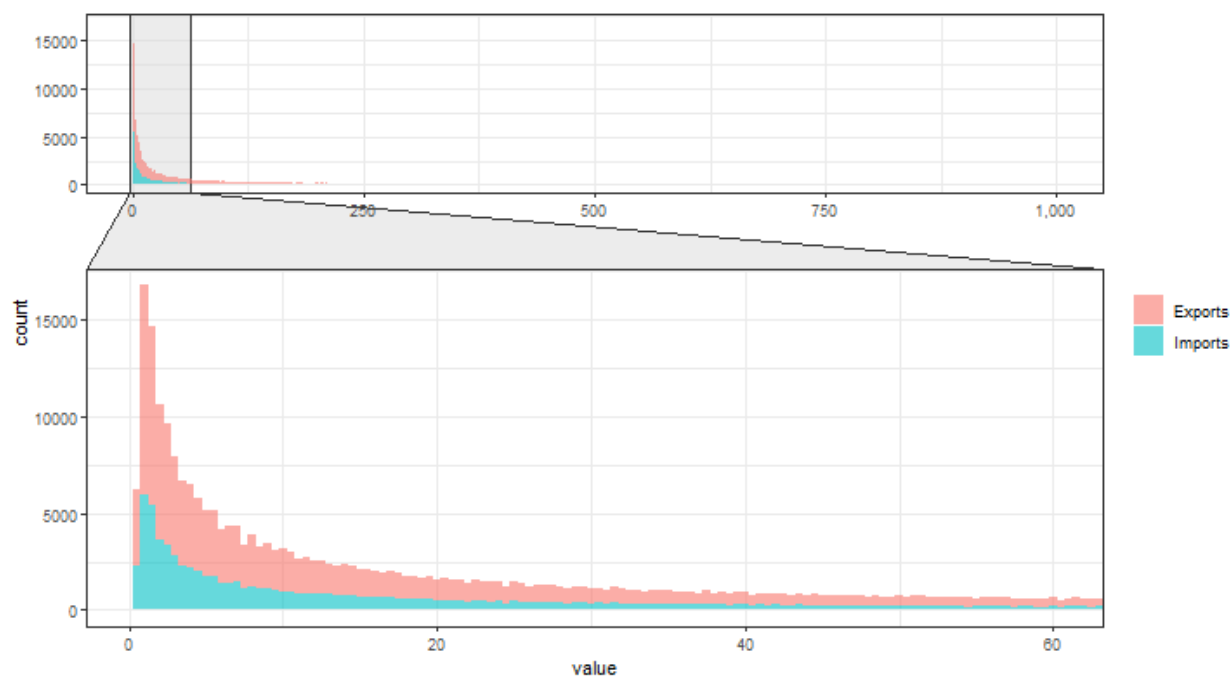


Figure 6: Boxplots of aggregated trade volume for main export partner countries

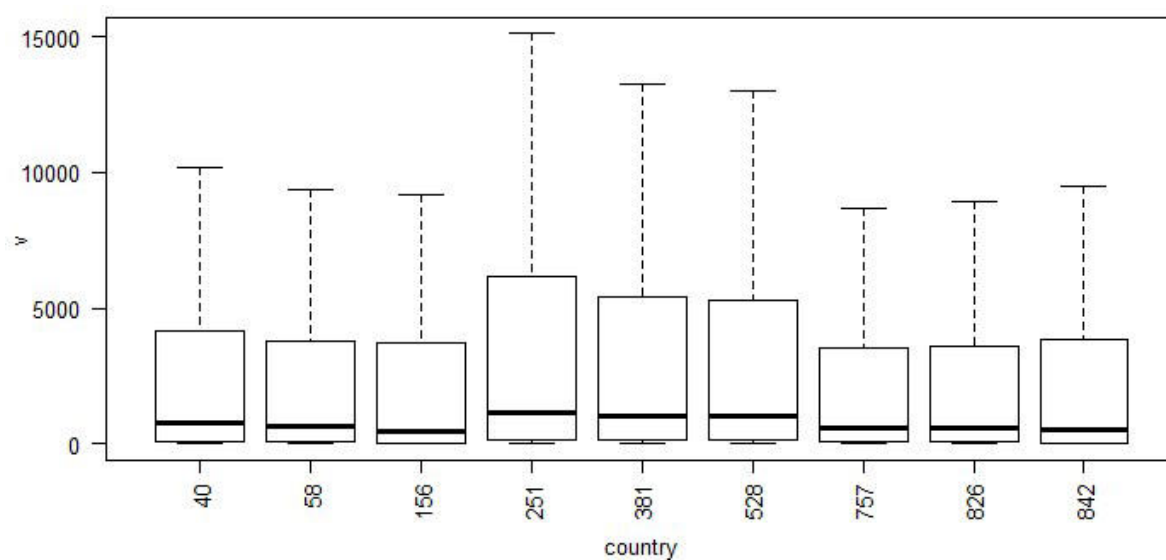


Figure 7: Boxplots of aggregated trade volume for main import partner countries

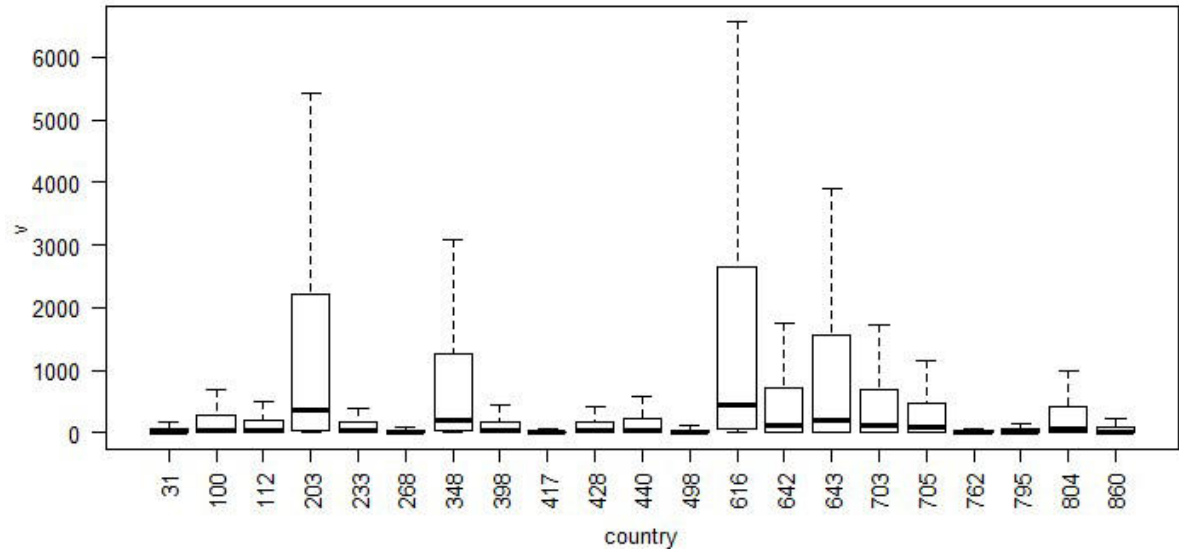
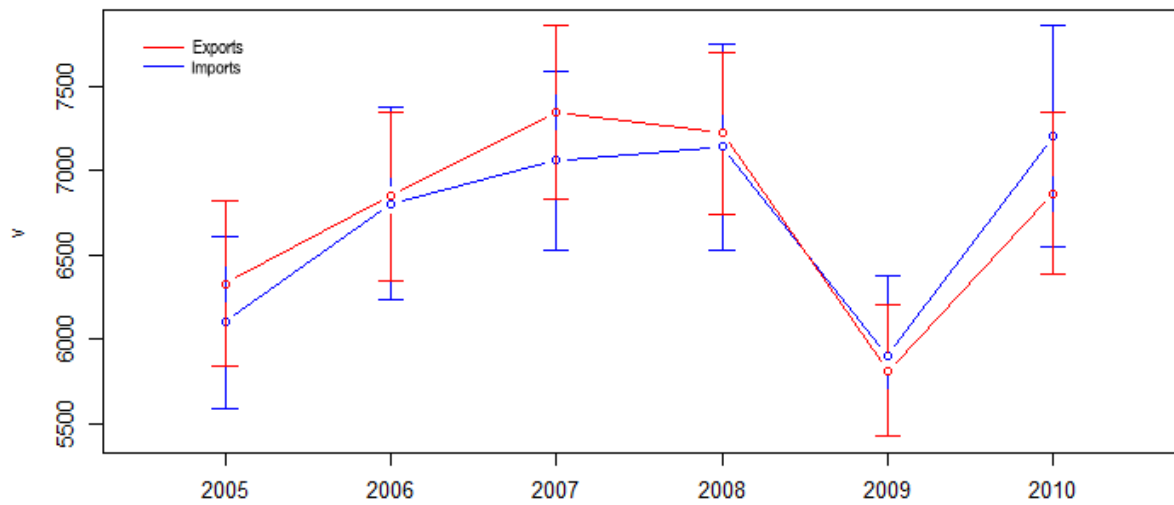


Figure 8: Heterogeneity across time for means of values by exports and imports



Tables

Table 7: Descriptive Statistics for main export partners by HS section

Section	Imports				Exports			
	n	median	sd	range	n	median	sd	range
Section I	7120	180.7	33007.6	793878.6	7444	240.5	24683.3	394861.6
Section II	11123	178.3	28847.0	773105.4	11290	319.1	8415.0	230474.4
Section III	2151	254.4	15176.2	197095.8	2369	417.4	8687.2	139053.6
Section IV	8344	573.2	28606.8	520457.4	8528	1266.2	24575.5	546241.2
Section IX	3277	321.9	11537.3	158104.1	3505	873.7	19638.2	359209.2
Section V	5360	279.4	316851.4	9382079.1	5973	318.9	147986.7	3275513.7
Section VI	33864	483.2	66361.8	4279017.4	36302	918.1	132683.3	10607698.8
Section VII	9792	2243.2	30322.6	484300.6	9979	4607.3	29352.5	315931.5
Section VIII	2451	301.7	15014.4	196189.8	2539	361.4	4355.9	47496.4
Section X	6672	888.0	15073.6	181214.1	6911	1667.7	25878.3	291383.0
Section XI	36940	142.3	12245.0	527323.2	37982	184.8	6032.8	146228.7
Section XII	2653	283.4	25626.2	373334.5	2662	325.7	7916.1	101832.5
Section XIII	6727	447.5	11866.2	252518.9	6972	1302.8	8431.8	190403.9
Section XIV	2365	449.3	55400.9	1474204.1	2516	637.1	68198.7	2433481.3
Section XIX	732	162.1	2886.2	33559.7	802	525.9	5613.8	60205.8
Section XV	27071	843.3	25135.0	700089.4	27816	1964.2	23248.3	589731.8
Section XVI	37535	1336.9	90725.7	5340747.3	38673	3759.5	47666.2	1810893.7
Section XVII	5694	2000.4	263632.4	8693844.7	6127	3851.8	474915.6	10056478.7
Section XVIII	10783	675.0	25014.1	547349.4	11167	1681.6	34236.4	723789.3
Section XX	6377	673.0	46955.0	1433699.0	6379	1186.9	24628.5	451570.3
Section XXI	378	571.1	12519.6	91118.5	378	1019.6	26933.1	245595.4

Table 8: Descriptive Statistics for main import partners by HS section

Section	Imports				Exports			
	n	median	sd	range	n	median	sd	range
Section I	3726	71.0	10639.7	213640.8	8811	44.1	7341.1	208848.6
Section II	6184	42.6	6892.2	158430.8	14376	27.7	2161.0	79429.9
Section III	825	59.2	6385.3	79903.3	2913	36.2	2577.7	42252.2
Section IV	5982	55.9	6204.2	146640.8	13754	69.1	5329.9	133659.1
Section IX	3431	199.8	9717.8	144843.4	5315	60.2	4278.0	83212.3
Section V	2532	93.7	456784.4	10746406.3	5910	40.3	29165.1	1010148.8
Section VI	14423	40.7	6495.2	237642.5	48574	51.0	10574.6	848417.6
Section VII	8239	89.4	14330.7	311800.5	17462	251.2	11785.2	350736.0
Section VIII	2056	41.2	2764.5	60925.7	3679	54.8	4264.3	85821.0
Section X	4781	61.4	12232.6	330497.7	10728	94.5	8954.3	292933.9
Section XI	29804	38.2	3393.3	123391.1	56525	34.9	2162.6	96254.4
Section XII	2014	44.3	7456.7	98712.2	4376	28.0	2653.7	68067.1
Section XIII	4843	59.0	5272.7	97695.3	11347	69.7	2490.9	84660.8
Section XIV	1271	71.5	14288.5	203976.0	2611	37.8	2661.8	72836.8
Section XIX	339	18.5	644.7	3605.6	1244	42.9	695.3	10818.8
Section XV	21817	118.6	19591.5	768763.6	41977	114.5	8632.3	288972.7
Section XVI	33705	86.8	33944.1	1652369.5	69361	255.2	18947.9	1681975.9
Section XVII	5325	206.9	64220.4	956642.6	10051	337.1	44447.8	1077557.5
Section XVIII	8209	36.3	10584.2	491973.1	18856	88.2	5978.8	268074.0
Section XX	5488	85.8	30955.2	569212.6	11805	67.4	5200.1	145515.8
Section XXI	318	21.5	164.1	1182.5	430	13.1	1226.0	14238.8

Table 9: HS2 classification system

Code	Section	Description
01 - 05	I:	Live animals, animal products
06 - 14	II:	Vegetable products
15	III:	Animal or vegetable fats and oils and there cleavage products
16 - 24	IV:	Prepared foodstuffs, beverages, spirits, vinegar, tobacco, manufactured tobacco substitutes
25 - 27	V:	Mineral products
28 - 38	VI:	Products of chemical or allied industries
39 - 40	VII:	Plastics, rubber and articles thereof
41 - 43	VIII:	Raw hides and skins, leather, furskins and articles thereof, saddlery and harness, travel goods, handbags and similar containers, articles of animal gut
44 - 46	IX:	Wood and articles of wood, wood charcoal, cork and articles of cork, manufactures of straw, esparto or of other planting materials, basket ware and wickerwork
47 - 49	X:	Pulp of wood or of other fibrous cellulosic material, recovered (waste and scarp) paper or paperboard, paper and paperboard and articles thereof
50 - 63	XI:	Textiles and Textile articles
64 - 67	XII:	Footwear, headgear, umbrellas, sun umbrellas, walking-sticks, seat-sticks, whips, riding-crops and parts thereof, prepared feathers and articles made therewith, artificial flowers, articles of human hair
68 - 70	XIII:	Articles of stone, plaster, cement, asbestos, mica or similar materials, ceramic products, glass and glassware
71	XIV:	Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal and articles thereof, imitation jewelry, coin
72 - 83	XV:	Base metals and articles of base metals
84 - 85	XVI:	Machinery and mechanical appliances, electrical equipment, parts thereof, sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles
86 - 89	XVII:	Vehicles, aircraft, vessels and associated transport equipment
90 - 92	XVIII:	Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments and apparatus, clocks and watches, musical instruments, parts and accessories thereof
93	XIX:	Arms and ammunition, parts and accessories thereof
94 - 96	XX:	Miscellaneous manufactured articles
97 - 98	XXI:	Works of art, collectors' pieces and antiques