CESifo Working Papers
www.cesifo.org/wp

Sectoral Trade Freeness and Agglomeration in the EU: An Empirical Test Strategy

Gianluca Cafiso

CESIFO WORKING PAPER NO. 4000
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
NOVEMBER 2012

CESifo *Working Paper No. 4000*

Sectoral Trade Freeness and Agglomeration in the EU: An Empirical Test Strategy

Abstract

This paper studies the evolution of trade freeness and of the agglomeration of production, as well as their relationship, at the sectoral level in a group of EU countries. Our main objective is to test at the sectoral level the conclusions of previous aggregate analyses which find that an inverse relationship between trade costs and agglomeration holds in case of the Home Market Effect. Our sectoral focus requires an original testing approach based on the combination of different bootstrap distributions.

JEL-Code: F100, F120, F150.

Keywords: trade freeness, agglomeration of production, home market effect, bootstrap.

Gianluca Cafiso
University of Catania
Department of Economics and Business
Corso Italia 55
Italy – 95129 Catania
gcafiso@unict.it

November 6, 2012
I thank the participants to the seminar series at the Economics Department of the University of Catania for very helpful comments; a special thank to Antonio Punzo for statistical support. Graditude goes to Luca De Benedicts for comments and suggestions at the SIE conference in Rome as well.

Contents

1 Introduction

Over the last two decades trade theorists have devoted much effort to the understanding of how trade costs condition trade patterns. As Anderson & van Wincoop (2004) explain, the group of costs which actually affect trade is large and difficult to disentangle. Shipment costs play an important role, but other factors raise the cost of consuming a product in a different location. Border-related formalities, technical and non-technical barriers, tariffs and standards, trade insurance and financing are just some other members of the trade costs family. In many regards, globalization and trade costs are the two sides of the same medal. Indeed, many believe that the large increase in trade observed during the last decades has been possible thanks to the evolution of communication and transportation technologies, evolution which has turned trade costs down. A big step forward which made proclaim *the death of distance* (Cairncross 1997). Even though trade costs have truly decreased over time (Jacks 2009), distance still matters (Disdier & Head 2008) and trade costs continue to condition trade flows significantly.

Production activities are not equally spread throughout the geographic space. As a matter of fact, agglomeration patterns always emerge (Brulhart & Traeger 2005). Patterns are a constant for some real activities such as manufacture, but they are common for immaterial activities too. There are many factors which cause activities to agglomerate in specific locations or to settle as far apart as possible from others. Resource endowments, economies of scale, infrastructures availability or congestion costs are just some possible explanations. In this paper we focus on agglomeration from a trade theory perspective (Combes et al. 2008). Under the new focus on trade costs, New Trade Theory models have generated predictions on how trade costs affect the agglomeration of production. Proximity to consumers is thought to influence firms' location decision according to a mechanism known as Home Market Effect (hereafter, HME) (Helpman & Krugman 1985). Accordingly, firms prefer to settle close to consumers because trade costs are proportional to distance and they influence the market price and firms' mark-up (Melitz & Ottaviano 2008). A recent theoretical paper by Behrens et al. (2009) shows that if one controls for the effect of "geographic differences" (Relative Centrality) and of "differences in productivity" (Comparative Advantage), agglomeration might truly follow the HME also in a multi-country world. In Behrens et al.'s (2009) framework the HME entails that when trade freeness (an indirect trade-costs indicator, see below) increases, agglomeration increases consequently because the HME itself strengthens. A general empirical test of this relationship is provided by Head & Ries (2001) for North America and by Niepmann & Felbermayr (2010) for the OECD countries.[1]

In this paper we study the evolution of trade freeness and agglomeration at the sectoral level

[1] Given the different possible causes of any observed agglomeration pattern, a relationship between trade costs and agglomeration is not to take for given. To wit, if agglomeration occurs because firms settle in a region to exploit that region's factor endowment, trade costs are not likely to explain that sector's pattern of agglomeration. Trade costs might also influence agglomeration through other channels. For example, lower trade costs might decrease a firm's incentive to settle in the large market with respect to another location's comparative advantage in fiscal terms. Given that trade-costs are less binding, the firm could serve the large market from a close-by fiscal-advantageous country.

(2-digit ISIC rev.3) in some EU countries in the period 1995-2006, we use trade freeness to account for the trade-costs effect. The edge of our analysis lies in its sectoral focus. Our objective is to test the conclusions of previous aggregate analyses [Head & Ries (2001), Niepmann & Felbermayr (2010)] which find that an inverse relationship between trade costs and agglomeration emerges. In the process to achieve our main objective, we develop an analysis which provides information about the net change of trade freeness and agglomeration in the period considered, which informs on the above-mentioned relationship regardless of the HME and which tests the HME hypothesis in the sectors considered.[2]

The sectoral focus conditions the analysis in two ways. First, we need indicators of trade costs and agglomeration which match by sector, time and country; this turns out to be feasible only using the Trade Freeness Indicator and the Theil Index for absolute concentration. Second, the sectoral testing makes regression analysis not feasible because too few observations are available for each sector.[3] Then, we have devised an original sign-test of the contemporaneous variation between trade freeness and agglomeration to test their relationship a the sectoral level.

As above-mentioned, our analysis has as theoretical reference the HME theory, to back our analysis with a clear theoretical reference, we sketch Behrens et al. (2009) conclusions in the next section. Afterwards, we study the net evolution of trade freeness and agglomeration separately and their relationship regardless of the HME hypothesis (section 4). Then, we focus on how they evolve when the HME hypothesis holds in a specific sector (section 5). In section 6 we draw the main conclusions. In appendix 1 we provide a detailed discussion of the data used, while appendix 2 introduces an alternative approach to test the sign of the contemporaneous variation discussed in section 4.3.

2 Trade Freeness and Agglomeration in a multi-country world with the HME

As mentioned in section 1, New Trade Theory (NTT) models which embed the Home Market Effect (HME) predict that decreasing trade costs foster agglomeration because the HME itself is supposed to strengthen.[4] In this section we provide a theoretical reference for this as recently modelled by Behrens et al. (2009) who use trade freeness as an indicator of trade costs.

In a multi-country framework, Behrens et al. (2009) affirm that the HME holds when:

$$\frac{\lambda_1}{\theta_1} \geq \cdots \geq \frac{\lambda_i}{\theta_i} \geq \cdots \geq \frac{\lambda_M}{\theta_M} \tag{1}$$

[2] The larger part of the research presented in this paper is informative regardless of the HME hypothesis which, in the end, turns out to be a sub-case of our analysis (section 5).

[3] Previous research on this issue [Head & Ries (2001), Niepmann & Felbermayr (2010),] was developed through regression analysis but only at an aggregate level.

[4] For a discussion of the HME in different settings, see Head et al. (2002).

where λ_i is country i's production share, θ_i is country i's demand share, M is the number of countries and $\theta_1 \geq ... \geq \theta_i \geq ... \geq \theta_M$. Condition 1 means that countries with larger demand shares host a relatively larger production share.

All other things being equal, NTT models of intra-industry trade indicate that condition 1 should hold because, given trade costs, firms have a motivation to set in large markets to be close to larger portions of their customers. However, condition 1 is not expected to hold in real-world data because, Behrens et al. (2009) explain, third-country effects offset the large-country motivation to settle in a specific location. In particular, they refer to differences in relative centrality and Ricardian comparative advantage.[5]

Behrens et al. (2009) argue that when it is possible "to separate the effect of relative centrality and comparative advantage on the one side, from the impact of relative demand driven by relative size and relative wages on the other side" (page 263) condition 1 is likely to hold. They demonstrate this by deriving the production shares $\left(\lambda_i^{SA}\right)$ which would prevail in the case of no comparative-advantage and no centrality-advantage in the network of countries. Consequently, the authors demonstrate that condition 1 is verified when using the λ_i^{SA} shares and they derive the following relationship [eq. 19 in Behrens et al. (2009)]:

$$\lambda_i^{SA} = \frac{1 + (M-1)\overline{\phi}}{1-\overline{\phi}}\theta_i - \frac{\overline{\phi}}{1-\overline{\phi}} \qquad (2)$$

where ϕ_{ij} is bilateral trade freeness and in the case of no geographic advantage in the network of countries $\phi_{ij} = \overline{\phi}$ (equal trade freeness for any country pair), and θ_i is country i's share of world demand.

Equation 2 in Behrens et al. (2009)'s model involves the direct relationship between trade freeness and agglomeration which we aim to test in our analysis: increasing trade freeness $\left(\overline{\phi}\uparrow\right)$ strengthens the HME towards the largest country.[6] This is made clearer through the following numerical example based on equation 2 where in case B trade is freer than in A (e.g. trade freeness parameter from 0.40 to 0.45):

- Case A: $M = 3; \overline{\phi} = 0.40$;

  $\theta_1 = 0.45 \rightarrow \lambda_1^{SA} = 0.683 \rightarrow \lambda_1^{SA}/\theta_1 = 1.517$

  $\theta_2 = 0.30 \rightarrow \lambda_2^{SA} = 0.233 \rightarrow \lambda_2^{SA}/\theta_2 = 0.773$

  $\theta_3 = 0.25 \rightarrow \lambda_3^{SA} = 0.083 \rightarrow \lambda_3^{SA}/\theta_3 = 0.332$

- Case B: $M = 3; \overline{\phi} = 0.45$;

  $\theta_1 = 0.45 \rightarrow \lambda_1^{SA} = 0.763 \rightarrow \lambda_1^{SA}/\theta_1 = 1.695$

  $\theta_2 = 0.30 \rightarrow \lambda_2^{SA} = 0.218 \rightarrow \lambda_2^{SA}/\theta_2 = 0.726$

  $\theta_3 = 0.25 \rightarrow \lambda_3^{SA} = 0.045 \rightarrow \lambda_3^{SA}/\theta_3 = 0.180$

[5] In Behrens et al.'s (2009) reasoning, when demand is equally distributed across regions, locations with a relative advantage in terms of better centrality and higher productivity always attract a larger share of production.

[6] Remember that the direct relationship between trade freeness and agglomeration traslates into an inverse relationship between trade costs and agglomeration because decreasing trade costs mean increasing trade freeness.

as expected, λ_1^{SA}/θ_1 is larger in case B.

In Behrens et al. (2009) the λ_i^{SA} shares guarantee that the HME occurs in a multi-country framework, but it is the HME which implies the inverse relationship between trade costs and agglomeration. Indeed, the HME strengthens when trade freeness increases because the less trade costs impact profit maximization, the more firms settle in the large market (by so increasing agglomeration) and serve periphery through exports. On the contrary, if trade costs are prohibitive, the motivation to stay close to each portion of consumers (not only to the largest one) is stronger. Then, when condition 1 holds in real-world sectors of activity, a direct relationship between trade freeness and agglomeration should emerge if the theoretical model applies to the real world. At the aggregate level and using regression analysis, Head & Ries (2001) and Niepmann & Felbermayr (2010) conclude that it holds for North America and Europe. Our objective is to test this relationship in different sectors to see how our results relate to aggregate ones, but first we need to develop a general analysis functional to such objective in section 4.

3 The data set

The empirical analysis in this paper uses a data set which considers twenty-one sectors of activity. These sectors are: "Mining and Quarrying" (ISIC Rev.3 10-14), "Electricity, Gas and Water Supply" (ISIC Rev.3 40-41), 18 subsectors of the manufacture aggregate (ISIC Rev.3 15-37), and "Agriculture, Hunting and Forestry" (ISIC rev 3 01-02); the full list of sectors is in Table 1. Our group of countries consists of the EU-15 countries before the May 2004 enlargement. All the data are for the 1995-2006 period, yearly figures. For analytic convenience, we will consider four subperiods defined as: period 1 (P1), 1995-1997; period 2 (P2), 1998-2000; period 3 (P3), 2001-2003; period 4 (P4), 2004-2006.[7] In section 4.1 we quantify trade freeness using bilateral export and national trade. Agglomeration in section 4.2 is computed using employment figures at the NUTS-2 regional level; for the agglomeration analysis "Agriculture, Hunting and Forestry" is not available (sector 21 in Table 1). To test the HME hypothesis in section 5.1, we use value added and *domestic absorption* figures computed as production plus import less export. All the data have the sector k, country i, year t or period p dimension; bilateral export has also the partner j dimension. More information on the data are in appendix 1.

4 Analysis of Trade Freeness and Agglomeration: All sectors

The objective of this section is twofold. First, we study the evolution of trade freeness and agglomeration separately and calculate their overall net change. Indeed, given deepening

[7] We aggregate years in 4 periods mainly to consider more significant variations of the indicators used in the analysis and to deal with missing-values by using intra-period averages.

Table 1: Sectors of Activity

Sector Number	Nace 1.1 ISIC rev 3	Sector Name
1	C, 10-14	Mining and quarrying
2	D, 15-16	manuf. of food products, beverages and tobacco
3	D, 17-19	manuf. of Textiles, textile products, leather and footwear
4	D, 20	manuf. of wood and of products of wood and cork, except furniture; manuf. of articles of straw and plaiting materials
5	D, 21-22	manuf. of pulp, paper and paper products; publishing and printing
6	D, 23	manuf. of coke, refined petroleum products and nuclear fuel
7	D, 24	manuf. of chemicals and chemical products
8	D, 25	manuf. of rubber and plastic products
9	D, 26	manuf. of other non-metallic mineral products
10	D, 27	manuf. of basic metals
11	D, 28	manuf. of fabricated metal products, except machinery and equipment
12	D, 29	manuf. of machinery and equipment n.e.c.
13	D, 30	manuf. of office machinery and computers
14	D, 31	manuf. of electrical machinery and apparatus n.e.c.
15	D, 32	manuf. of radio, television and communication equipment and apparatus
16	D, 33	manuf. of medical, precision and optical instruments, watches and clocks
17	D, 34	manuf. of motor vehicles, trailers and semi-trailers
18	D, 35	manuf. of other transport equipment
19	D, 36	manuf. of furniture; manufacturing n.e.c
20	E, 40-41	Electricity, gas and water supply
21	A, 01-02	Agriculture, hunting and forestry

Notes: · sector 21 is not available for the analysis of agglomeration.

integration in the EU during the period under consideration (to wit, the Euro was introduced in 1999), we want to check whether and how much trade freeness and agglomeration have changed. Secondly, we consider their contemporaneous variation to asses how one evolves given the other's specific change in one direction. Accordingly, in subsection 4.1 and 4.2 we consider the evolution of trade freeness and agglomeration across the first (P1:1995-97) and last period (P4:2005-2006) to draw conclusions about their net evolution, while in section 4.3 we consider their sequential evolution across all the periods available (P1, P2, P3, P4) in order to study their relationship.

4.1 Trade Freeness

In our analysis we use an indicator which reflects trade costs known as Trade Freeness Indicator (TFI) (Head & Mayer 2004). Measures of trade costs are difficult to obtain, researchers therefore resort often to this indirect measurement [Behrens et al. (2009) and Niepmann & Felbermayr (2010)]. Actually, the TFI is to interpret as the endogenous trade-flows outcome to trade costs as well as consumption preferences. Ideally, one would like to offset the effect of preferences; unfortunately this is not possible in the context of our analysis and we use it in its typical formulation.[8]

The TFI is computed using bilateral export and national trade, it quantifies trade freeness through the difference between foreign and domestic trade, assuming that this difference depends only upon restrictions to foreign trade.[9] In our analysis, we use a sectoral and time specification of this indicator:

$$TFI_{ij,t}^{k} = \sqrt[2]{\frac{x_{ij,t}^{k} \cdot x_{ji,t}^{k}}{x_{ii,t}^{k} \cdot x_{jj,t}^{k}}} \tag{3}$$

where $x_{ij,t}^{k}$ is country i's export of k to j in year t, and $x_{ii,t}^{k}$ is country i's national trade, computed as country i's total production less total export of k in year t.

Implicit to its construction is the hypothesis of symmetric trade barriers ($TFI_{ij} = TFI_{ji}$). The TFI ranges in the 0-1 interval, where $TFI = 0$ suggests prohibitive trade costs (no bilateral trade) and $TFI = 1$ suggests no trade costs (free trade).[10] In Table 2 we report summary statistics for $TFI_{ij,t}^{k}$ and in Figure 1 we plot the across-sectors and pairs average of $TFI_{ij,t}^{k}$ ($TFI_{ij,t}^{k} \Rightarrow TFI_t$).

From Table 2 it emerges that the bulk of TFI values is below 0.15 ($95th$ percentile); the mean value is only 0.035 and the median is 0.009. Then, trade freeness is low among the EU-15 countries.[11] Figure 1 shows that the mean trade-freeness (TFI_t) has decreased from 1995 to 2001 and risen again afterwords.

For the purpose of our analysis, we average the yearly bilateral trade-freeness values along two dimensions: first, to obtain a unique sectoral estimate of the index ($TFI_{ij,t}^{k} \Rightarrow TFI_t^{k}$), second, to obtain average values for 4 subperiods of 3 years each ($TFI_t^{k} \Rightarrow TFI_p^{k}$). Then, we consider the trade-freeness evolution from the first to the last period: $\Delta TFI^k = TFI_{p4}^{k} - TFI_{p1}^{k}$; $\Delta TFI^k > 0$ indicates a trade freeness increase, $\Delta TFI^k < 0$ viceversa.

Average trade-freeness values for the whole period (1995-2006), the 4 subperiods and the time difference are in Table 3. The subperiod values are plotted in Figure 2 to display the temporal evolution. The lowest TFI value (highest trade costs) is for sector 20 "Electricity, gas and water supply", the highest TFI value (lowest trade costs) is for sector 13 "manufacture of office machinery and computers".[12]

[8] Chen & Novy (2011) follow this direction by engineering a modified version of the TFI which requires elasticity-of-substituon estimates. However, we cannot apply their enhanced measurement. Indeed, our level of aggregation makes any elasticity estimate unreliable and meaningless, and at the same time we need to maintain it to own data for the agglomeration analysis in section 4.2 which match with the trade freeness indicator. Then we choose to use the typical TFI in order not to introduce any bias.

[9] The role of preferences is disregarded by the TFI. For a discussion of how the difference between imported and domestic consumption is due to the combined effect of preferences and restrictions to trade, see Anderson & van Wincoop (2003).

[10] To wit, for $TFI = 0.5$ (very high level, compare with Table 3) the product of bilateral trade (numerator) is one-forth the product of national trade (denominator), for $TFI = 1$ bilateral trade is as much as national trade. TFI>1 is computationally possible, however, it is not at all realisitic. This is why the TFI range is restricted to 0-1 in practice, where 1 represents the free trade case.

[11] Our values are in line with those reported by other authors. For a comparison, see Niepmann & Felbermayr (2010) Figure 2b and Table 9, or Head & Mayer (2004) Table 5.

[12] It comes with no surprise that trade costs impact relatively less sector 13. Indeed, this is a high value-added sector where products are realized in large plants and shipped at low cost to several markets for sale.

Table 2: Trade Freeness, summary statistics.

sec	mean	min	max	p50	iqr	p5	p95
1	0.00792	1.98E-07	0.29210	0.00208	0.00642	0.00002	0.03101
2	0.01736	2.49E-04	0.50727	0.00637	0.01372	0.00054	0.06754
3	0.08167	1.58E-03	0.98252	0.03611	0.07159	0.00339	0.33165
4	0.01332	4.19E-06	0.31361	0.00456	0.00844	0.00018	0.06098
5	0.01515	3.63E-05	0.23039	0.00744	0.01339	0.00075	0.05747
6	0.02339	5.73E-07	0.91860	0.00506	0.01617	0.00002	0.07947
7	0.07671	1.30E-03	0.98201	0.03378	0.07365	0.00279	0.32332
8	0.03395	2.31E-04	0.80879	0.01522	0.03414	0.00142	0.12783
9	0.01217	5.27E-05	0.22000	0.00476	0.01177	0.00036	0.04975
10	0.07092	8.71E-05	0.97016	0.03348	0.06808	0.00185	0.29623
11	0.01105	2.50E-05	0.13189	0.00492	0.00898	0.00052	0.05111
12	0.04805	8.57E-04	0.75270	0.02416	0.04826	0.00298	0.16426
13	0.19093	2.84E-04	0.97799	0.07578	0.24522	0.00598	0.72742
14	0.05391	3.99E-04	0.92431	0.02715	0.05412	0.00232	0.19227
15	0.08233	7.30E-05	0.91629	0.02926	0.08594	0.00122	0.37948
16	0.06513	7.55E-04	0.81816	0.03994	0.06645	0.00345	0.21045
17	0.09874	5.02E-06	0.89268	0.04285	0.14493	0.00146	0.32461
18	0.04979	4.68E-05	0.68285	0.01772	0.04719	0.00126	0.22424
19	0.01955	1.53E-04	0.28260	0.00720	0.01612	0.00056	0.09172
20	0.00306	1.07E-08	0.04653	0.00042	0.00290	0.00000	0.01542
21	0.01169	3.34E-06	0.71168	0.00238	0.00791	0.00006	0.04477
Total	0.03567	1.07E-08	0.98252	0.00983	0.03014	0.00032	0.15215

Notes: • "p50" is the Median, "iqr" id the 75-25 interquantiles range, "p5" is the 5th , "p95" is the 95th .

Figure 1: Trade Freeness.

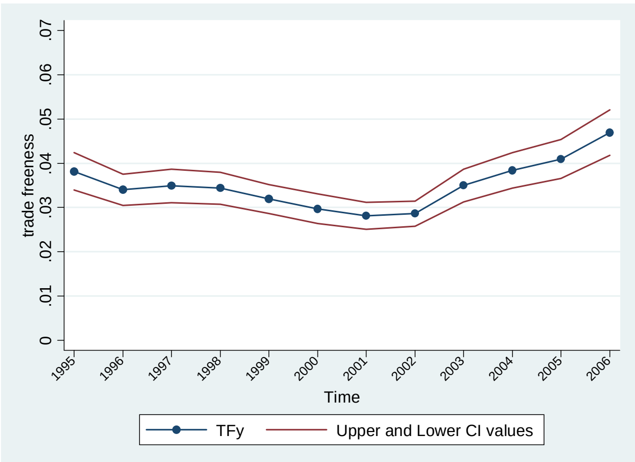


Table 3: Trade Freeness by sector.

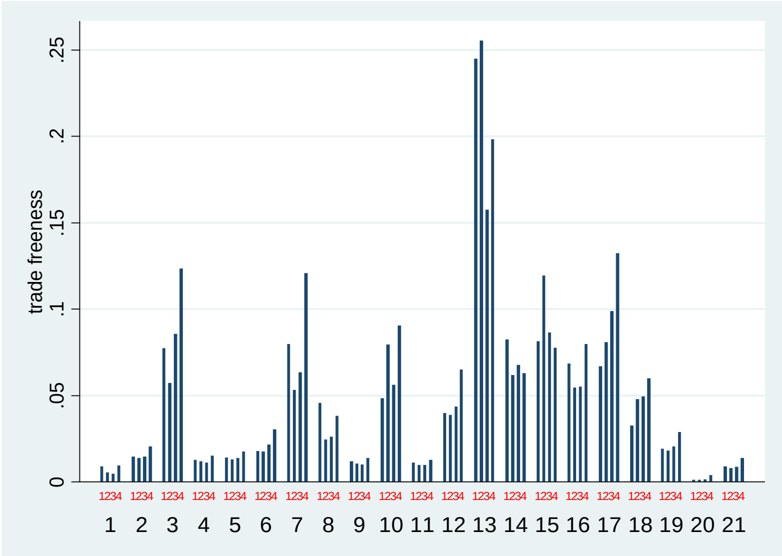
	a	b	c	d	e	f
sec	TFI_{all}^{k}	TFI_{p4}^{k}	TFI_{p3}^{k}	TFI_{p2}^{k}	TFI_{p1}^{k}	$\Delta TFI_{p4.p1}^{k}$
1	0.0078	0.0094	0.0045	0.0055	0.0088	0.0006†
2	0.0169	0.0205	0.0144	0.0138	0.0146	0.0059*
3	0.0965	0.1234	0.0855	0.0572	0.0773	0.0461*
4	0.0131	0.0151	0.0111	0.0120	0.0126	0.0025*
5	0.0150	0.0175	0.0137	0.0129	0.0140	0.0035*
6	0.0228	0.0304	0.0216	0.0175	0.0179	0.0125*
7	0.0923	0.1206	0.0634	0.0532	0.0798	0.0408*
8	0.0483	0.0383	0.0262	0.0245	0.0458	-0.0075†
9	0.0120	0.0137	0.0101	0.0105	0.0120	0.0017*
10	0.0884	0.0904	0.0561	0.0795	0.0483	0.0421*
11	0.0110	0.0126	0.0096	0.0098	0.0111	0.0015*
12	0.0483	0.0651	0.0435	0.0387	0.0396	0.0254*
13	0.2301	0.1981	0.1574	0.2553	0.2448	-0.0467 na
14	0.0911	0.0627	0.0675	0.0618	0.0824	-0.0198†
15	0.1050	0.0775	0.0863	0.1195	0.0814	-0.0038†
16	0.0662	0.0797	0.0550	0.0546	0.0685	0.0113†
17	0.1022	0.1322	0.0989	0.0808	0.0669	0.0653*
18	0.0491	0.0598	0.0495	0.0477	0.0325	0.0273*
19	0.0213	0.0288	0.0204	0.0179	0.0192	0.0096*
20	0.0016	0.0038	0.0014	0.0011	0.0013	0.0025 na
21	0.0109	0.0136	0.0088	0.0080	0.0090	0.0046*

Notes: · Bootstrap-based test:"†" signals no rejection of "Ho: ΔTFI=0" against "H1:ΔThIk $\neq$0" at 5%, "*" signals rejection, "na" stands for test-output not available. · columns *a* reports the whole-period average of TFI, column *b*-*e* the subperiod values, column *f* the difference value between P4 and P1.

To study the evolution of trade freeness, we consider the ΔTFI^k values in Table 3. However, such differences might not be statistically significant (to wit, they emerge because of measurement errors). We therefore use bootstrap-simulations for hypothesis testing to check the significance of the TFI variation. The column ΔTFI^k in Table 3 reports the outcome of the bootstrap-based test.[13] Given non-normality of the bootstrap distribution in the bulk of sectors, we resort to *Bias-corrected and Accelerated Confidence Intervals* (BCA-CI) to define rejection areas. The system of hypotheses is "$H_0 : \Delta TFI^k = 0$" against "$H_1 : \Delta TFI^k \neq 0$". A significant trade freeness variation emerges in fourteen out of nineteen sectors available: sectors 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 17, 18, 19, 21 exhibit a significantly positive variation. Then, whenever trade-freeness changes significantly, this reflects a trade costs reduction. We expected a generalized trade-freeness increase because the period under analysis comprises the implementation of many EU integration policies (first of all, the Euro's introduction in 1999). Nonetheless, the analysis shows that some sectors were unaffected.

[13] For more information about the boostrap procedure used, check appendix 1.

Figure 2: Trade Freeness by sector and period.



4.2 Agglomeration of Production

The distribution of production is uneven throughout the geographic space. Agglomeration emerges because of different causes and, usually, it can be explained through direct and evident arguments such as infrastructure availability, manufacture tradition, resource endowments, etc; for a discussion about this in the case of the EU, see Cafiso (2011). By abstracting from specific cases, New Economic Geography models (Baldwin et al. 2005) provide a formalization of these processes and stress the role of trade costs. For the purpose of our analysis, we need to measure agglomeration through a synthetic index which quantifies how much each sector is far away from an even distribution over the geographic space.

Agglomeration can be measured in absolute or relative terms through concentration indices. An industry is concentrated in absolute terms if the bulk of its production takes place only in few locations (no other sector considered), while it is in relative terms if its geographic distribution is concentrated with respect to the average distribution of all the other sectors considered (Haaland et al. 1999). Since we are interested in scale economies and trade, we consider absolute concentration as measured through the Theil Index (ThI). The Theil Index formula for absolute concentration is:

$$ThI^k = \frac{1}{R} \cdot \sum_{r=1}^{R} \left[\left(\frac{x_r^k}{\overline{x}^k} \right) \cdot \log \left(\frac{x_r^k}{\overline{x}^k} \right) \right] \quad (4)$$

where $0 \leq ThI^k \leq \log R$, x_r^k is activity x (employment, production, value-added, etc.) in sector k ($k = 1, ..., K$) and region r ($r = 1, ..., R$), $\overline{x}^k = \frac{1}{R} \sum_{1=r}^{R} x_r^k$ is the across-regions average and each region belongs exclusively to one country i ($i = 1, ..., M$).[14]

[14]The Theil Index is obtained from the formula of the Generalized Class of Entropy Indices when the sensitivity parameter α is set equal to 1; if $\alpha = 2$ one obtains the Half Square Coefficient of Variation. The more positive α is, the more sensitive the index to differences at the top of the distribution; the less α is, the more sensitive to differences at the bottom of the distribution. For a discussion of inequality indices, check

Among the inequality indices available, we choose the Theil Index because it is additively decomposable by groups:

$$ThI^k = ThI^k_{between} + ThI^k_{within}$$

This allows us to quantify how much inequality derives from within-groups and between-groups differences. Indeed, we are particularly interested in between-groups inequality because it will be matched with the TFI discussed in section 4.1.

We use regional employment available for the first twenty sectors listed in Table 1, and set countries as groups of region.[15] The index is calculated both for the whole time-span (12 years, from 1995 to 2006) and for the 4 subperiods of 3 years each. As for trade freeness, we check the evolution of agglomeration $\Delta ThI^k = ThI^k_{p4} - ThI^k_{p1}$; $\Delta ThI^k > 0$ indicates an increase in agglomeration. To check the statistical relevance of the variation ($H_0 : \Delta ThI^k = 0$), we employ a bootstrap-based test as the one discussed in Brulhart & Traeger (2005); we resort again to BCA-CI to define rejection areas.[16] The index values (total and between component) are reported in Table 4 and plotted in Figure 3; the time-differences and the test output are also in Table 4.

Cowell (2000).

[15] In case of relevant productivity differences across regions and sectors, the use of employment figures might bias the indicators of agglomeration. For this reason some authors prefer using value added figures (Behrens et al. 2005). We use employment because this is available at the Nuts-2 regional level for the sectors, countries and years to match with the TFI. Besides, we reckon that such differences are not particularly relevant in our analysis because we consider a homogenous set of countries. In the Eurostat Regio data set, employment figures were not available for sector 21 "Agriculture, Hunting and Forestry".

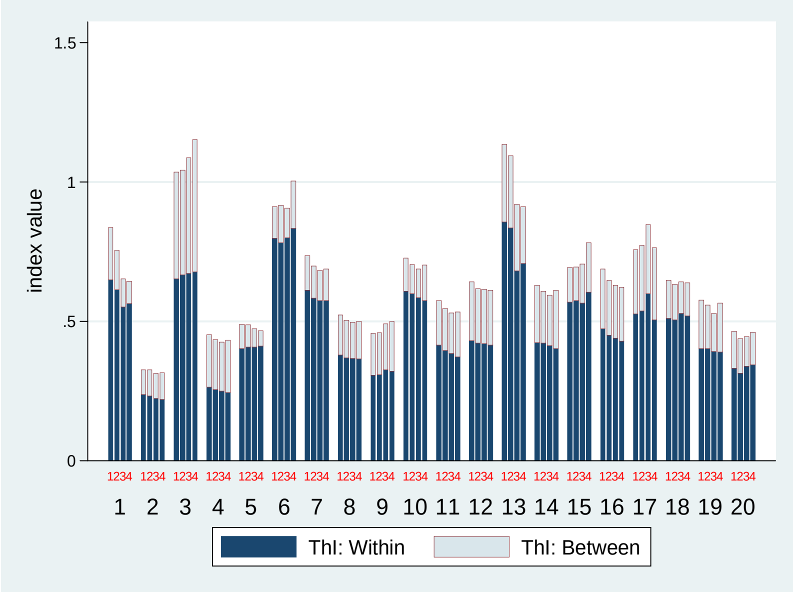
[16] The Shapiro-Francia test (Shapiro & Francia 1972) signals a normal distribution for sectors 3, 4, 5, 6, 8, 10, 11, 14, 16, 17, 18, 19. We use again Bias-corrected and accelerated Confidence Intervals (BCA-CI) for coherence with the ΔTFI test. As a matter of fact, non-normality does not bias the Z-test since we get the same sectoral outcome when we refer to rejection areas based on the normal distribution. This holds both for trade freeness and agglomeration.

Table 4: Theil Index by sector: Total and Between component.

	a	b	c	d	e	f	g	h	i	l	m	n
	Theil Index, Total						Theil Index, Between Component					
sec	ThI^T_{all}	ThI^T_{p4}	ThI^T_{p3}	ThI^T_{p2}	ThI^T_{p1}	$\Delta ThI^T_{p4.p1}$	ThI^B_{all}	ThI^B_{p4}	ThI^B_{p3}	ThI^B_{p2}	ThI^B_{p1}	$\Delta ThI^B_{p4.p1}$
1	0.735	0.644	0.652	0.756	0.837	-0.192*	0.124	0.081	0.101	0.143	0.188	-0.107*
2	0.321	0.315	0.313	0.326	0.327	-0.011	0.090	0.096	0.089	0.094	0.089	0.007
3	1.084	1.153	1.087	1.042	1.036	0.117*	0.402	0.476	0.415	0.376	0.383	0.093*
4	0.436	0.433	0.425	0.433	0.452	-0.019	0.178	0.189	0.175	0.179	0.189	0.001
5	0.480	0.466	0.473	0.487	0.489	-0.023	0.070	0.055	0.066	0.079	0.087	-0.031*
6	0.934	1.003	0.907	0.914	0.912	0.092	0.120	0.171	0.107	0.136	0.113	0.058
7	0.702	0.687	0.683	0.699	0.735	-0.048*	0.114	0.113	0.109	0.115	0.124	-0.012
8	0.506	0.500	0.496	0.503	0.524	-0.024	0.133	0.136	0.129	0.133	0.143	-0.008
9	0.477	0.500	0.491	0.459	0.458	0.042	0.158	0.180	0.165	0.150	0.152	0.028
10	0.706	0.702	0.686	0.705	0.727	-0.025	0.110	0.128	0.103	0.107	0.119	0.009
11	0.546	0.534	0.529	0.546	0.574	-0.040*	0.148	0.161	0.145	0.150	0.158	0.003
12	0.622	0.612	0.615	0.617	0.642	-0.030	0.197	0.198	0.194	0.195	0.212	-0.014
13	1.065	0.911	0.922	1.094	1.134	-0.224*	0.241	0.204	0.239	0.260	0.278	-0.074
14	0.612	0.612	0.593	0.607	0.629	-0.017	0.191	0.209	0.180	0.186	0.205	0.004
15	0.719	0.782	0.705	0.694	0.691	0.092	0.126	0.177	0.141	0.121	0.123	0.053
16	0.646	0.622	0.629	0.647	0.687	-0.065*	0.196	0.193	0.188	0.197	0.215	-0.022
17	0.787	0.763	0.848	0.773	0.758	0.006	0.241	0.258	0.249	0.237	0.232	0.026
18	0.640	0.638	0.643	0.632	0.648	-0.009	0.121	0.119	0.114	0.127	0.137	-0.018
19	0.557	0.566	0.528	0.558	0.576	-0.010	0.154	0.177	0.136	0.155	0.173	0.003
20	0.453	0.462	0.444	0.438	0.464	-0.002	0.117	0.117	0.106	0.124	0.132	-0.014
21	na	na	na	na	na	na	na	na	na	na	na	na

Notes: · Bootstrap-based test:"*" signals rejection of "Ho: ΔThIk = 0" against"H1:ΔThIk $\neq$0"at 5% · columns a and g report the whole-period value of the index, columns b-e and h-m the by-period value, columns f and n the difference value between P4 and P1.

Figure 3: Theil Index by sector and period



Agglomeration is lowest for sector 2 "manufacture of food products, beverages and tobacco" and highest for sector 3 "manufacture of textiles, textile products, leather and footwear". From the between/within decomposition (Figure 3), it emerges that agglomeration is mainly due to within-countries differences. The lowest share of between-inequality (maximum of within) is for "manufacture of coke, refined petroleum products and nuclear fuel" (sector 6), while maximum between–inequality (minimum of within) is for "manufacture of wood and products of wood and cork, ..." (sector 4).[17] It comes with no surprise that sector 6 has the highest share of within concentration, this sector enjoys large scale economies but it is strategic for each country. Then, production is highly concentrated within countries but not at the European level.

As for the evolution of agglomeration between the first (1995-1997) and the last period (2004-2006), column $\Delta ThI^{T,k}$ in Table 4 reports the time difference of the total Theil Index, and column $\Delta ThI^{B,k}$reports the difference of its between component. Agglomeration (total) changes significantly only in 6 out of twenty sectors: sectors 1, 7, 11, 13, 16 exhibit a negative variation (agglomeration decreases), while sector 3 exhibits a positive variation. On the whole, sectoral agglomeration does not increase in the period considered, it is either stable or decreasing. This is an important result in line with similar analyses (Brulhart & Traeger 2005) which confutes the common belief of increasing agglomeration in manufacture. Furthermore, also across-countries inequality (between-agglomeration) is mainly stable along the period considered (variation of the Theil Index's between component).

[17] When the within contribution is higher than the between, concentration depends mainly upon an uneven distribution within countries; the across-countries distribution is relatively less unequal. To wit, if a sector were spread unevenly between countries but equally among the regions of each country, concentration would depend only upon across-countries diversity.

4.3 The Sign of the Contemporaneous Variation

In this section we study the contemporaneous variation of the TFI and the *between component* of the Theil Index $\left(ThI^B\right)$ in order to check its sign. We match the TFI with between-agglomeration, and not with overall agglomeration (ThI^T), because the TFI provides information only about between-countries trade costs. Then, it is correct to consider how between-countries agglomeration relates to trade freeness.

We consider the sequential evolution (P2/P1, P3/P2 and P4/P3) of trade freeness and between-agglomeration, we are interested in knowing when trade-freeness and between-agglomeration have changed in the same direction ($TFI \uparrow$ - $ThI^B \uparrow$ or $TFI \downarrow$ - $ThI^B \downarrow$) and when in the opposite direction ($TFI \downarrow$ - $ThI^B \uparrow$ or $TFI \downarrow$ - $ThI^B \uparrow$). The time-differences of TFI^k and $ThI^{B,k}$, as well as their matched product Y^k, are reported in Table 5 and plotted in Figure 4.[18]

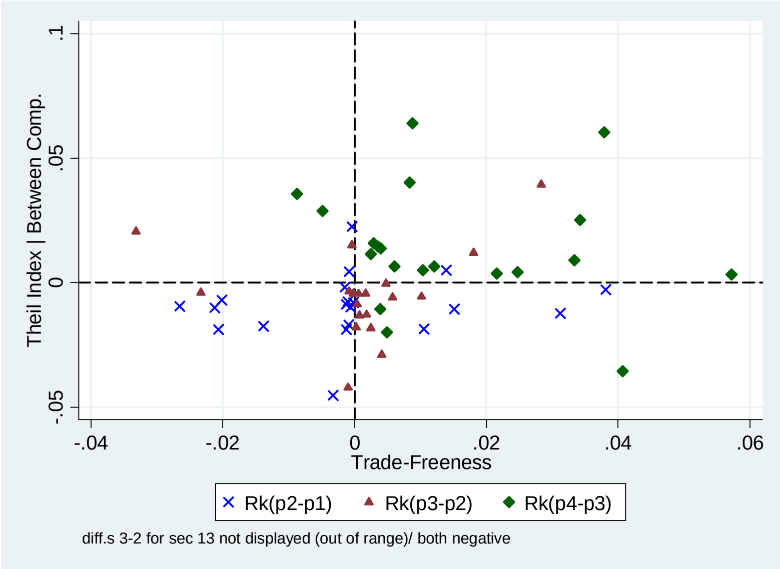
[18]Values plotted in figure 4 are those in column ΔTF^k and $\Delta ThI^{B,k}$ of table 5.

Table 5: TFI and ThI-between time-differences by sector and period.

	diff(P2-P1)			diff(P3-P2)			diff(P4-P3)		
sec	$\Delta TFI^{k}_{p2.p1}$	$\Delta ThI^{B,k}_{p2.p1}$	$Y^{k}_{p2.p1}$	$\Delta TFI^{k}_{p3.p2}$	$\Delta ThI^{B,k}_{p3.p2}$	$Y^{k}_{p3.p2}$	$\Delta TFI^{k}_{p4.p3}$	$\Delta ThI^{B,k}_{p4.p3}$	$Y^{k}_{p4.p3}$
1	-0.003	-0.045	+	-0.001	-0.042	+	0.005	-0.020	-
2	-0.001	0.004	-	0.001	-0.004	-	0.006	0.007	+
3	-0.020	-0.007	+	0.028	0.039	+	0.038	0.060	+
4	-0.001	-0.010	+	-0.001	-0.004	+	0.004	0.014	+
5	-0.001	-0.008	+	0.001	-0.013	-	0.004	-0.011	-
6	0.000	0.022	-	0.004	-0.029	-	0.009	0.064	+
7	-0.027	-0.009	+	0.010	-0.006	-	0.057	0.003	+
8	-0.021	-0.010	+	0.002	-0.004	-	0.012	0.007	+
9	-0.001	-0.002	+	0.000	0.015	-	0.004	0.015	+
10	0.031	-0.012	-	-0.023	-0.004	+	0.034	0.025	+
11	-0.001	-0.009	+	0.000	-0.004	+	0.003	0.016	+
12	-0.001	-0.017	+	0.005	0.000	-	0.022	0.004	+
13	0.011	-0.019	-	-0.098	-0.020	+	0.041	-0.036	-
14	-0.021	-0.019	+	0.006	-0.006	-	-0.005	0.029	-
15	0.038	-0.003	-	-0.033	0.021	-	-0.009	0.036	-
16	-0.014	-0.017	+	0.000	-0.009	-	0.025	0.004	+
17	0.014	0.005	+	0.018	0.012	+	0.033	0.009	+
18	0.015	-0.011	-	0.002	-0.013	-	0.010	0.005	+
19	-0.001	-0.019	+	0.002	-0.018	-	0.008	0.040	+
20	0.000	-0.008	+	0.000	-0.018	-	0.002	0.011	+
21	-0.01	na	na	0.01	na	na	0.05	na	na
			if + =14			if + =7			if + =15

Notes: · "$\Delta TFI^{k}_{pX.pY}$" is the trade-freeness difference value between period X and period Y, "$\Delta ThI^{B,k}_{pX.pY}$" is the Theil Index (Between Comp.) difference value between period X and period Y, "$Y^{k}_{pX.pY}$" is the product of the trade-freeness difference times the Theil Index difference.

Figure 4: TFI and ThI^B matched time-differences by sector.



For sectors in portion I (X-axis and Y-axis positive) and III (X-axis and Y-axis negative) of Figure 4, trade freeness and between-agglomeration evolve in the same direction: positive TFI changes are associated with positive ThI^B changes and vice versa. This is the case for 36 out of sixty pairs.[19]

Figure 4 is useful to gain a first insight, but we need to employ a statistical procedure to test the robustness of the sign. To wit, it could be that the sign of the observed differences is determined by few non-relevant observations, while there is not a statistically-significant information about the sign of the contemporaneous variation. This is what we aim to check now through a formal statistical procedure which allows us to gain information at the sectoral level.

We build our testing-strategy on the consideration that if trade freeness and between-agglomeration evolve in an opposite direction ($TFI \downarrow$ - $ThI^B \uparrow$ or $TFI \downarrow$ - $ThI^B \uparrow$), the product of their variation is negative, otherwise it is positive. In Table 5 we report ΔTFI^k and $\Delta ThI^{B,k}$ and the sign of their product Y^k for each period/sector combination. We check the sign of Y^k through hypothesis-testing based on bootstrap-generated distributions. The bootstrap distribution (Y_r^k) of Y^k comes from combining the bootstrap distributions of ΔTFI^k and $\Delta ThI^{B,k}$ obtained as discussed in sections 4.1 and 4.2. The Y_r^k values are generated as:

$$Y_r^k = \underbrace{\Delta TFI_r^k}_{\alpha} \cdot \underbrace{\Delta ThI_r^{B,k}}_{\beta} \tag{5}$$

where, as defined before, $\Delta TFI_r^k = TFI_{r,p}^k - TFI_{r,p-1}^k$ and $\Delta ThI_r^{B,k} = ThI_{r,p}^{B,k} - ThI_{r,p-1}^{B,k}$. The Y_r^k values are the product of bootstrap-generated values, $r = 1, ..., R$ where R is the total

[19] 20 pairs (one for each sector) for the differences between P4-P3, twenty for P3-P2 and twenty for P2-P1.

number of replications instructed for the bootstrap.[20] To test the sign of Y^k using the Y_r^k distribution we employ a non-parametric approach; an alternative approach (percentiles-based method) for the same objective is discussed in appendix 2. The non-parametric approach used here is based on the portion of positive cases over the total as defined by the following statistic:

$$J^k = \frac{1}{R} \cdot \sum_{i=1}^{R} \overset{+}{Y}_r^k \qquad (6)$$

where $\overset{+}{Y}_r^k = 1$ if $Y_r^k > 0$ and $\overset{+}{Y}_r^k = 0$ if $Y_r^k \leq 0$. $\overset{+}{Y}_r^k$ is a binary variable which has a Bernullian distribution, but J^k is binomially distributed. We actually run two tests based on the J^k statistic, the hypotheses system for each of the two tests is:

1. "$H_0 : J^k = 1/2$" against "$H_{1a} : J^k > 1/2$"; rejection area for $Z_1^k > z_\alpha$.
2. "$H_0 : J^k = 1/2$" against "$H_{1b} : J^k < 1/2$"; rejection area for $Z_1^k < -z_\alpha$.

z_α is the critical value of the standard-normal for a significance-level equal to α.[21] The null-hypothesis states that the portion of positive cases is not different from half: there is no clear information about the sign of the contemporaneous variation. Obviously, H_{1b} is admissible only when H_{1a} is not. The test-statistic is:

$$Z_1^k = \frac{J^k - p_0}{\sqrt{\frac{p_0(1-p_0)}{B}}}$$

where $p_0 = 1/2$ since we test "$H_0 : J^k = 1/2$". The results based on the Z_1^k test-statistic when $\alpha = 1\%$ are in Table 6 (Z_1^k columns).

[20] The Y_r^k distribution is not normal even in those sectors where ΔTF_r^k and $\Delta ThI_r^{B,k}$ are normally distributed. The product of two normal distributions is a non-normal distribution known as *Normal Product Distribution*.

[21] We use the standard-normal approximation of the Binomial distribution because R is sufficiently large.

Table 6: Test of the sign of the contemporaneous variation.

		Z_1^k=	diff(P2-P1)				diff(P3-P2)				diff(P4-P3)			
sec	HME	$H1a$	Y^k	J^k	Z_1^k	$PCmethod$	Y^k	J^k	Z_1^k	$PCmethod$	Y^k	J^k	Z_1^k	$PCmethod$
1		2/3	+	0.997	H1a	H1a	+	0.854	H1a	H1a	-	0.093	H1b	H1b
2		1/3	-	0.243	H1b	H1b	-	0.222	H1b	H1b	+	0.943	H1a	H1a
3		3/3	+	0.601	H1a	H1a	+	0.983	H1a	H1a	+	0.986	H1a	H1a
4		3/3	+	0.833	H1a	H1a	+	0.652	H1a	H1a	+	0.986	H1a	H1a
5		1/3	+	0.917	H1a	H1a	-	0.079	H1b	H1b	-	0.073	H1b	H1b
6		1/3	-	0.399	H1b	H1b	-	0.254	H1b	H1b	+	0.988	H1a	H1a
7	yes	2/3	+	0.949	H1a	H1a	-	0.315	H1b	H1b	+	0.671	H1a	H1a
8	yes	2/3	+	1	H1a	H1a	-	0.288	H1b	H1b	+	0.93	H1a	H1a
9		2/3	+	0.629	H1a	H1a	-	0.119	H1b	H1b	+	0.995	H1a	H1a
10	yes	2/3	-	0.034	H1b	H1b	+	0.631	H1a	H1a	+	0.998	H1a	H1a
11	yes	3/3	+	0.945	H1a	H1a	+	0.716	H1a	H1a	+	0.979	H1a	H1a
12	yes	2/2	+	0.87	H1a	H1a	-	0.532	Ho	Ho	+	0.766	H1a	H1a
13	yes		-	n.a.	n.a.	n.a.	+	n.a.	n.a.	n.a.	-	n.a.	n.a.	n.a.
14	yes	1/3	+	0.92	H1a	H1a	-	0.406	H1b	H1b	-	0.364	H1b	H1b
15		0/2	-	0.49	Ho	Ho	-	0.157	H1b	H1b	-	0.382	H1b	H1b
16	yes	2/2	+	0.988	H1a	H1a	-	0.516	Ho	Ho	+	0.708	H1a	H1a
17	yes	1/1	+	n.a.	n.a.	n.a.	+	n.a.	n.a.	n.a.	+	0.64	H1a	H1a
18	yes	1/3	-	0.119	H1b	H1b	-	0.397	H1b	H1b	+	0.539	H1a	Ho
19	yes	2/3	+	0.947	H1a	H1a	-	0.039	H1b	H1b	+	1	H1a	H1a
20			+	n.a.	n.a.	n.a.	-	n.a.	n.a.	n.a.	+	n.a.	n.a.	n.a.
21														
			Ho: 1/17; H1a: 12/17; H1b: 4/17				Ho: 2/17; H1a: 5/17; H1b: 10/17				Ho: 0/18; H1a: 14/18; H1b: 4/18			

Notes:· The "Z_1^k" columns report the outcome of the non-parametric test (section 4.3), the "PC method" columns report the outcome of the test based on the 45-55 interquantiles range (appendix 2). · The "Y^k" columns report the sign of the product $\Delta TFI_p^k * \Delta ThI_p^{B,k}$ (in table 5), column "J^k" the portion of positive observations over the total. · "Ho:Jk=1/2","H1a:Jk>1/2","H1b:Jk<1/2". · The "$Z_1^k = H1a$" column reports the number of cases in favour of $H1a$ by sector.

No rejection of the null hypothesis (H_0: no distinguishable sign) is for 3 out of fifty-two tests available.[22] When H_0 is rejected (49 cases), rejection in favor of a change in the same direction (H_{1a}) is for 31 out of forty-nine cases, rejection in favor of a change in the opposite direction (H_{1b}) is for 18 out of forty-nine cases. Consequently, trade freeness and between-agglomeration have changed more in the same direction (63% cases), but they have changed in the opposite direction too. As for evidence at the sectoral level, only sectors 3, 4, 11 keep constant their relationship throughout the time-span considered (trade freeness and between-agglomeration evolve in the same direction); the other sectors alternate between variations in the same and in the opposite direction. On the whole, a stable relationship between trade freeness and between-agglomeration does not emerge when one considers each sector separately.

5 Analysis of Trade Freeness and Agglomeration: HME sectors

As explained in section 2, New Trade Theory models which embed the HME predict that decreasing trade costs foster agglomeration because the HME itself is supposed to strengthen. If this happens, increases of trade freeness need to match with between-agglomeration positive variations. In this section we check whether or not this emerges in sectors which support the HME hypothesis. Then, first we check in which sectors the HME hypothesis holds, secondly we control whether the sign of the contemporaneous variation (as studied in section 4.3) is positive for those HME sectors.

5.1 Test of the HME hypothesis

In this subsection we test the HME hypothesis, we achieve this by verifying condition 1 in each sector through Behrens et al.'s (2005) approach. They develop three tests (all based on the same Z_{ij}^k statistic) and use Spearman's rank correlation coefficients; the tests are non-parametric and use observed production and demand shares. Here, we apply only one of their test (the most powerful) and Spearman's coefficients.
The statistic at the base of their testing-strategy is:

$$Z_{ij}^k = \underbrace{\left(\frac{\lambda_i^k}{\theta_i} - \frac{\lambda_j^k}{\theta_j}\right)}_{A} \cdot \underbrace{(\theta_i - \theta_j)}_{B} \tag{7}$$

where k stands for the sector, i and j are two (a pair) of the M countries considered, λ^k stands for the sectoral production share and θ for the overall demand share. The total number

[22] If the test were always executable, we would have 60 resuts (three time-diferences x twenty sectors). However, it is not in eight sector/time-difference combinations. We decide not to bootstrap when the observed sample contains more than 50% missing values, otherwise it would run on a sample with too little information.

of Z_{ij}^k equals the number of combinations of M countries taken by two. If $B > 0$ country i's demand share is larger than country j's, the HME requires A to be positive as well: $Z_{ij}^k > 0$ therefore supports the HME. If $B < 0$ country i's demand share is less than country j's, the HME requires A to be negative (the product of two negative quantities is positive), consequently $Z_{ij}^k > 0$ always supports the HME. Building on Z_{ij}^k, Behrens et al. (2005) define the Pairwise-Average Z-test, the Country-Average Z-test and the World-Average Z-test. The statistic of the Pairwise-Average Z-test, which is their most powerful in testing the HME hypothesis, is:[23]

$$S_1^k = \frac{\sum_i \sum_{j<i} \overset{+}{Z_{ij}^k}}{N}$$

where $\overset{+}{Z_{ij}^k} = 1$ if $Z_{ij}^k > 0$ and $\overset{+}{Z_{ij}^k} = 0$ if $Z_{ij}^k \leq 0$, $N = M(M-1)/2$ is the number of combinations. The test checks whether Z_{ij}^k is positive in more than half cases through S_1^k. S_1^k is the fraction of favorable cases over the total. The system of hypotheses is "$H_0 : S_1^k = 1/2$" against "$H_1 : S_1^k > 1/2$"; rejection of H_0 in favour of H_1 supports the HME hypothesis. S_1^k is binomially distributed. Given that the number of pairs (available combinations) is sufficiently large, we resort to the standard-normal approximation of the binomial distribution, which is:

$$Z_2^k = \frac{N \cdot S_1^k - N \cdot p + \frac{1}{2}}{\sqrt{N \cdot p \cdot (1-p)}}$$

where $N = M(M-1)/2$, $p = 1/2$ and $(1/2)$ is the continuity correction for the approximation of the discrete binomial to the continuous normal distribution. We also compute Spearman's rank-correlation coefficients between the series λ_i^k/θ_i and θ_i (Conover 1999). The coefficients are instrumental to testing the independency hypothesis between the two series; positive (and possibly significant) correlation values signal support of the HME hypothesis. The test results and Spearman's coefficients are in Table 7.[24]

Spearman's coefficients signal positive correlation (support the HME hypothesis) for 15 out of twenty-one sectors; however, only for four sectors the correlation is statistically significant at 5 percent. The strongest correlation is for sector 11 ("manufacture of fabricated metal products"). The Pairwise Z-test signals that the HME hypothesis holds in eleven out of twenty-one sectors (5% significance); strongest evidence is again for sector 11. Spearman's coefficients are positive for all these eleven sectors.[25]

[23] The strongest test of the HME is the Pairwise Z-test, the Country Average and the World Average follow. The pairwise test is the strongest because, differently from the other two, it does not sum across individual Z_{ij}^k values. Then, it cannot be the case of many small negative observations to be offset by some few positive ones.

[24] We report results only for the Pairwise-Average Z-test, results for the Country and World-Average tests are available upon request.

[25] The sectors which support the HME hypothesis are: sector 7 "manuf. of chemicals and chemical products", 8 "manuf. of rubber and plastic products", 10 "manuf. of basic metals", 11 "manuf. of fabricated metal products, except machinery and equipment", 12 "manuf. of machinery and equipment n.e.c", 13 "manuf. of office machinery and computers", 14 "manuf. of electrical machinery and apparatus n.e.c.", 16 "manuf. of medical, precision and optical instruments, watches and clocks", 17 "manuf. of motor vehicles,

Table 7: Test of the HME hypothesis.

sec	PW-Zt	SM-rh	No.CV+	sec	PW-Zt	SM-rh	No.CV+
1	0.50	-0.03	2/3	11	0.87**	0.83**	3/3
2	0.35	-0.42	1/3	12	0.78**	0.71**	2/2
3	0.45	-0.14	3/3	13	0.67**	0.41	na
4	0.57	0.18	3/3	14	0.60*	0.25	1/3
5	0.53	0.05	1/3	15	0.55	0.08	0/2
6	0.33	-0.38	1/3	16	0.66**	0.38	2/2
7	0.62*	0.26	2/3	17	0.76**	0.70**	1/1
8	0.73**	0.63*	2/3	18	0.69**	0.53	1/3
9	0.37	-0.33	2/3	19	0.62*	0.35	2/3
10	0.69**	0.45	2/3	20	0.56	0.15	na
				21	0.36	-0.37	na
mean	0.58						

Notes: •"PW-Zt" reports the statistic of the Pairwise-Av. Z-test, "SM-rh" reports Spearman's coefficients. •* $p < 0.05$, ** $p < 0.01$ • "No.CV+" reports the number of positive contemporaneous variations as resulting from the nonparametric test in table 6.

5.2 The HME and the evolution of trade freeness and agglomeration

At this point we check whether trade-freeness increases match with between-agglomeration's positive variations stably in the sectors which support the HME hypothesis.[26] In Table 7 and 6 (respectively in the columns "$No.CV+$" and "$Z_1^k = H1a$") we report the number of cases when the contemporaneous variation is statistically positive by sector (outcome of the sign-test); the test is available only for 9 of the eleven sectors which support the HME hypothesis.[27]

For sector 11 a positive contemporaneous variation emerges in three out of three time-differences available: the relationship holds and it is stable in this sector. For sectors 7, 8, 10, 19 in two out of three time-differences available: the relationship holds in the majority of cases but it is not stable. For sectors 12 and 16 in two out of two cases available, while for sectors 14 and 18 only in one case out of three available.

These findings show that a direct relationship between trade freeness and between-agglomeration emerges stably only in one HME sector out of nine when one adopts a restrictive criterion (three out of three cases available: sector 11 only) and in five HME sectors out of nine when one adopts a less restrictive criterion (at least in two out of three cases available: sectors 11, 7, 8, 10, 19). However, regardless of the criterion adopted, the analysis shows that the HME prescription of a direct relationship between trade freeness and agglomeration needs to be

trailers and semi-trailers", 18 "manuf. of other transport equipment", 19 "manuf. of furniture; manufacturing n.e.c".

[26] It is to remember that an inverse relationship between trade costs and agglomeration translates into a direct relationship between trade freeness and agglomeration; this last tested by looking at the sign of the contemporaneous variation.

[27] We have only one time-difference/sector combination available for the sign-test in sector 17, then we do not consider it in the HME assessment.

checked sector by sector and that it is not robust.

6 Conclusions

The main objective of this work was to verify the HME prescription about the relationship between trade costs and agglomeration by comparing our sectoral results with previous aggregate analyses which affirm that it holds. In the process to achieve this, we have developed a general study of the relationship between trade freeness and agglomeration. The strength and originality of our analysis lies in its sectoral focus. This has required us to device an original testing approach based on combined bootstrap-distributions.

Our analysis shows that the net evolution of trade freeness and agglomeration, as well as their interaction, needs to be studied sector by sector. On the whole, trade freeness seems to have increased while agglomeration shows a non-increasing trend in the period considered; but the net evolution differs across sectors. As for the interaction of trade freeness and between-agglomeration, they are more likely to move in the same direction but the sign of their contemporaneous variation is not stable by sector. This applies also (and to a larger extent) to sectors which support the HME hypothesis.

On the basis of our analysis, we therefore believe previous aggregate analyses not to reveal properly the fragility of the theoretical relationship between trade freeness and agglomeration in case of the HME, fragility which emerges in our sectoral study.

References

Anderson, J. E. & van Wincoop, E. (2003), 'Gravity with gravitas: a solution to the border puzzle', *American Economic Review* **93**(1), 170–192.

Anderson, J. & van Wincoop, E. (2004), 'Trade costs', *Journal of Economic literature* **42**(3), 691–751.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G. & Robert-Nicoud, F. (2005), *Economic geography and public policy*, Princeton University Press.

Behrens, K., Lamorgese, A., Ottaviano, G. & Tabuchi, T. (2005), Testing the home market effect in a multi-country world, Discussion Paper 2005/55, UCLouvain-CORE.

Behrens, K., Lamorgese, A., Ottaviano, G. & Tabuchi, T. (2009), 'Beyond the home market effect: Market size and specialization in a multi-country world', *Journal of International Economics* **79**(2), 259–265.

Brulhart, M. & Traeger, R. (2005), 'An account of geographic concentration patterns in Europe', *Regional Science and Urban Economics* **35**(6), 597–624.

Cafiso, G. (2011), 'Sectoral border effects and the geographic concentration of production', *Review of World Economics (Weltwirtschaftliches Archiv)* **147(3)**, 543–566.

Cairncross, F. (1997), *The death of distance*.

Cameron, A. & Trivedi, P. (2005), *Microeconometrics: methods and applications*, Cambridge University Press.

Chen, N. & Novy, D. (2011), 'Gravity, trade integration and heterogeneity across industries', *Journal of International Economics* **forthcoming**.

Combes, P., Mayer, T. & Thisse, J. (2008), *Economic geography*, Princeton University Press.

Conover, W. J. (1999), *Practical Nonparametric Statistics*, Wiley, New York.

Diciccio, T. J. & Romano, J. P. (1988), 'A review of bootstrap confidence intervals', *Journal of the Royal Statistical Society. Series B (Methodological)* **50**(3), 338–354.
URL: *http://www.jstor.org/stable/2345699*

Disdier, A. & Head, K. (2008), 'The puzzling persistence of the distance effect on bilateral trade', *The Review of Economics and Statistics* **90**(1), 37–48.

Haaland, J., Kind, H., Knarvik, K. & Torstensson, J. (1999), What determines the economic geography of Europe?, Discussion Paper 2072, CEPR.

Head, K. & Mayer, T. (2004), The empirics of agglomeration and trade, *in* J. V. Henderson & J. F. Thisse, eds, 'Handbook of Regional and Urban Economics', Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, chapter 59, pp. 2609–2669.

Head, K., Mayer, T. & Ries, J. (2002), 'On the pervasiveness of home market effects', *Economica* **69**(275), 371–390.

Head, K. & Ries, J. (2001), 'Increasing returns versus national product differentiation as an explanation for the pattern of US-Canada trade', *American Economic Review* **91**(4), 858–876.

Helpman, E. & Krugman, P. (1985), *Market structure and foreign trade*, MIT press Cambridge, MA.

Jacks, D. S. (2009), 'On the death of distance and borders: Evidence from the nineteenth century', *Economics Letters* **105**(3), 230–233.
URL: *http://econpapers.repec.org/RePEc:eee:ecolet:v:105:y:2009:i:3:p:230-233*

Melitz, M. J. & Ottaviano, G. I. P. (2008), 'Market size, trade, and productivity', *Review of Economic Studies* **75**(1), 295–316.
URL: *http://ideas.repec.org/a/bla/restud/v75y2008i1p295-316.html*

Niepmann, F. & Felbermayr, G. (2010), 'Globalisation and the Spatial Concentration of Production', *World Economy* **33**(5), 680–709.

Shapiro, S. & Francia, R. (1972), 'An approximate analysis of variance test for normality', *Journal of the American Statistical Association* **67**(337), 215–216.

Appendix 1. Data

Throughout the paper we consider twenty-one sectors of activity classified according to ISIC rev.3/NACE 1.1, these sectors are: -a- the aggregate for "Mining and Quarrying" (NACE: C, ISIC:10-14), -b- 18 subgroups of manufacture as partition of the "Total Manufacturing" aggregate (NACE: D, ISIC: 15-37; see Table 1 for the list of all the sectors), -c- the aggregate for "Electricity, Gas and Water Supply" (NACE: E, ISIC: 40-41), and -d- the aggregate for "Agriculture, Hunting and Forestry" (NACE: A, ISIC: 01-02). The countries comprised in the analysis are those in the EU-15 group: Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom. Data for Belgium and Luxembourg are recorded together for the so-called Belgium-Luxembourg Economic Union (BLEU). The time range is 1995-2006, whole-period figures are the average of all the yearly observations. We use values for four sequential sub-periods defined as follows: period 1 values are the average of 1995-1997 yearly figures, period 2 of 1998-2000, period 3 of 2001–2003 and period 4 of 2004-2006.

Agglomeration analysis
To calculate the sectoral Theil Index we use employment figures (number of employees) at the Nuts-2 regional level; data are extracted from the Eurostat Regio data set. We start with 207 Nuts-2 regions: Austria, 9 regions; Belgium, 10 regions; Germany, 38 regions; Denmark, 5 regions (5 deleted); Spain, 19 regions (8 deleted); Finland, 5 regions (1 deleted); France, 22 regions; Greece, 13 regions (1 deleted); Ireland, 2 regions; Italy, 21 regions; Luxembourg, 1 region (1 deleted); the Netherlands, 12 regions; Portugal, 5 regions; Sweden, 8 regions; United Kingdom, 37 regions. Some regions were deleted in case of too-many missing values. Employment data for sector 21 "Agriculture, Hunting and Forestry" were not available in the Eurostat Regio data set, we therefore did not consider this sector for the agglomeration analysis.

Trade Freeness
For the trade freeness indicator we use bilateral export among the EU-15 countries which constitute our sample, plus National Trade figures computed as "total production less total export". All figures are in current US dollars. Bilateral export, total production and total export are extracted from the OECD Stan database.

Production and Demand shares for the HME
To test the HME we use sectoral-production and aggregate-demand country shares. Production shares are calculated using sectoral value-added figures, we use Domestic Absorption to account for demand. Domestic absorption is computed as "production less export plus import". Value added, national production, export and import in US dollars are extracted from the OECD Stan database.

Bootsrap-based Test for the TFI variation

We test the hypothesis $\Delta TFI_p^k \neq 0$ (where $\Delta TFI_p^k = TFI_p^k - TFI_{p-1}^k$) by using bootstrap simulations to generate two distributions with R observations each: one of TFI_p^k and one of TFI_{p-1}^k values. Each bootstrap runs $R = 1000$ replications. The same procedure is executed K times, one for each sector k. The procedure is described here sequentially, we omit the superscript k in what follows:

1. We start with two observed samples of $TFI_{ij,p}$ values of size N_p $(p = 1, 2)$; one for each period considered for the time difference.

2. The bootstrap generates R samples $(r = 1, ..., R)$ of size N_p $\left(N_p^r\right)$ of bootstrap-generated $TFI_{ij,p}$ values. This is done twice: once for p and once for $p - 1$.

3. For each sample r, we calculate the average of the N_p^r bootstrap-generated $TFI_{ij,p}$ values: TFI_p^r. So that we have two distributions (one for p and one for $p - 1$) of R values TFI_p^r.

4. We calculate the R differences $\Delta TFI_p^r = TFI_p^r - TFI_{p-1}^r$, by so doing we have for each sector k a distribution (R observations) of ΔTFI_p^r.

5. The ΔTFI_p^r distribution is used to generate standard errors to test the hypothesis $\Delta TFI_p \neq 0$.

We check normality of the ΔTFI_p^r distribution through the Shapiro-Francia Test (Shapiro & Francia 1972); normality is not rejected only in sectors 3, 7, 15. Given non-normality we resort to Bias-Corrected and Accelerated Confidence Intervals (BCA-CI) to define rejection areas. Sectors 13 and 20 are not available for bootstrap-simulation given the high number of missing values in the observed N_p sample.
For a discussion of the different confidence intervals for hypothesis testing available in this context, see Diciccio & Romano (1988). We refer the reader to Cameron & Trivedi (2005) chapter 11 for more information about the Bootstrap.

Appendix 2. A percentiles-based Test of the Sign

The test of the sign (discussed in section 4.3) checks whether the portion of positive observations is statistically larger than half for deciding about the sign of the Y^k statistic. In this context, the word *statistically* makes the difference. Indeed, if we were not looking for a statistically-significant result, one might simply check what is the percentage of positive observations over the total (R), and conclude that Y^k is positive if more than 50% observations are positive. This reduces to observe the median of the distribution. The issue with this approach is that the median might be just one position away from zero. Then, a conclusion based only on the sign of the median could lead to a non-robust statement about the sign of Y^k. We propose now an alternative way to decide about the sign of the Y^k statistic.
In the spirit of hypothesis-testing based on Percentiles Confidence Intervals [Cameron & Trivedi (2005) section 11.2.7], one may define buffers around the median which guarantee to decide about H_0 at a certain robustness-level. We define these buffers γ as follows: when γ increases, the rejection area decreases; we get more restrictive on rejection of H_0. We opt for this definition because it guarantees that more evidence is required to opt for either sign of the statistic when larger buffers are used. We consider three standard levels: $\gamma = 1\%$, $\gamma = 5\%$ and $\gamma = 10\%$, and define respectively the rejection areas in the following Table 8.

Table 8: Rejection Areas

H_0: J_k= 1/2	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$
alternatives ↓	rejection area ↓	>	rejection area ↓	>	rejection area ↓
H_{1a}: J_k> 1/2	$p49 > 0$		$p45 > 0$		$p40 > 0$
H_{1b}: J_k< 1/2	$p51 < 0$		$p55 < 0$		$p60 < 0$
H_1: J_k≠ 1/2	$p49.5 > 0 \mid p50.5 < 0$		$p47.5 > 0 \mid p52.5 < 0$		$p45 > 0 \mid p55 < 0$

The rationale behind the definition of the rejection areas for H_0 is that when at least $(50 + \gamma)\%$ observations are positive (negative), one can assume that Y^k is positive (negative) at a robustness-level equal to $\gamma\%$. On the contrary, one concludes that the sign is not distinguishable (at a robustness-level equal to $\gamma\%$) if there is not enough evidence.[28] We run the two tests discussed in section 4.3 using $\gamma = 5\%$; the two tests are:

1. "$H_0 : J^k = 1/2$" against "$H_{1a} : J^k > 1/2$"; rejection area for $45p > 0$.

2. "$H_0 : J^k = 1/2$" against "$H_{1b} : J^k < 1/2$"; rejection area for $55p < 0$.

Results based on this approach are in Table 6. Except for sector 18/diff(P4-P3), rejection of the null-hypothesis for the same alternative is exactly in the same cases as those selected by the Z_1^k statistic ($\alpha = 1\%$).

[28] It is to notice that the robustness-level γ is inversely related to the significance level α used in standard hypothesis testing as applied in section 4.3. In the former, the lower is γ the more likely the rejection of H_0 is. On the contrary, in conventional hypothesis testing, the lower is α, the less likely the rejection of H_0 is.