EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1116 / NOVEMBER 2009

SECTORIAL BORDER EFFECTS IN THE EUROPEAN SINGLE MARKET

AN EXPLANATION THROUGH INDUSTRIAL CONCENTRATION

by Gianluca Cafiso

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1116 / NOVEMBER 2009

SECTORIAL BORDER EFFECTS IN THE EUROPEAN SINGLE MARKET

AN EXPLANATION THROUGH INDUSTRIAL CONCENTRATION [1]

by Gianluca Cafiso [2]

In 2009 all ECB publications feature a motif taken from the €200 banknote.

1 The opinions expressed in this paper are solely those of the author and do not necessarily reflect the views of the European Central Bank or their respective institutions.

2 University of Catania, Economics Department, Corso Italia 55, Catania – Italy; e-mail: gcafiso@unict.it

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Website
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

The statement of purpose for the ECB Working Paper Series is available from the ECB website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract

The purpose of this paper is to explain the relation between the Border Effect and industrial concentration. This is achieved by founding this relation on the Home Market Effect and testing the robustness of this foundation through an application to the European Single Market. A sectorial Gravity Equation is estimated using different econometric estimators, in particular we discuss a recently suggested technique for the estimation of log-linear CES models. Overall, our findings suggest a steady relation between the Border Effect and industrial concentration. Besides, the analysis of industrial concentration through a synthetic index provides us with valuable insights into the structure of the European industry.

Keywords: Trade, Border Effect, Industrial Concentration, Home Market Effect, European Single Market.

JEL Classification: F10, F12, F15.

Non-Technical Summary

This paper studies trade and industrial concentration patterns among the fifteen countries which constituted the European Union before the enlargement in May 2004. It flows into that strand of research which is committed to explaining the Border Effect estimated in many gravity equations. The purpose is to check the relation between the Border Effect estimated by industrial sector and industrial concentration, a relation that we found on the Home Market Effect. Among the empirical works which test different explanations of the Border Effect, we build our analysis on Chen (2004) because she is the first to detect a relation of this kind. We extend her analysis by clarifying the theoretical linkage between the BE and industrial concentration, and then we test it again by using a richer model, more accurate data exclusively for the EU and different econometric techniques. Overall, our findings strongly support the relation discussed. Moreover, the analysis of industrial concentration provides us with valuable insights into the structure of the European industry.

The border effect is a well-known concept in international trade which serves as a measure of trade integration among trade partners (McCallum 1995). We estimate it through a gravity equation which uses sectorial trade data and other control variables, we use figures for the 1995-2003 period. Concentration is quantified through indices which we discuss. For the purpose of the paper, we use the Theil index to quantify absolute concentration since we need to account for economies of scale as the main determinants of trade. Nonetheless, we compute relative concentration as well and compare it with absolute. The analysis of concentration through the Theil index allows us to discuss the between/within pattern of industrial concentration and to check its trend through a bootstrap test.

We put much effort in evaluating the different econometric estimators available in order to find the most appropriate one. In particular, we compare the econometric estimator discussed by Silva and Tenreyro (2006), known as Poisson Pseudo Maximum Likelihood, with the more common OLS and TOBIT. We provide an explanation of why, given the nature of our dataset, there is not much difference when one uses any of the three above-mentioned estimators.

I. Introduction.

The degree of trade integration among partners is of interest to trade economists. It is so because research in international trade looks for the causes of limited trade integration. A frequently-used measure of trade integration is the Border Effect (BE), this accounts for the lack of integration in a group of different economies due to the presence of border-linked trade costs (McCallum 1995). Many estimations of the aggregate BE are available for the European Single Market, but few are at the level of specific industrial sectors; we believe that Chen's (2004) are the most remarkable.[1]

Chen estimates the BE for 78 different industrial sectors by means of a gravity equation. The novelty is her attempt to explain the BE by means of *trade costs* variables and *behavioural responses to trade costs* variables. Among the *behavioural responses to trade costs* variables that Chen considers, there is an industrial concentration index describing the geographic agglomeration of productive activities; at our knowledge, she is the first to have tried an explanation of this kind for the BE. At first sight the link between concentration and the BE is not straightforward, but if one considers the relation between the Gravity Approach and the New Economic Geography, then it will get clearer. A concentration index mirrors the behavioural response to a trade-costs variable because it reflects the outcome of firms' location decisions given the presence of trade costs. Indeed, on the basis of the New Economic Geography models, firms settle in a way to minimize trade costs (Ottaviano and Thisse 2004). However, firms' location decision inevitably influences trade exchanges among countries. It is so because in any country affected by the firms' move, the ratio between internal to external trade changes. In this perspective, a Concentration Index is explicative of the BE and changes of industrial concentration over time must therefore influence the BE. This is a relevant improvement in the understanding of the BE and this sheds light on the relation between observed trade flows and industrial concentration patterns.

Even though the originality of Chen's analysis is unquestionable, we felt that the relation between the BE and spatial concentration needed more investigation to overcome some flaws in Chen's.[2] The aim of this paper is therefore to enhance the explanation of the BE through industrial concentration by means of: a) the discussion of the theoretical linkage between the BE and Industrial Concentration; b) a concentration index computed using European employment data; c) the estimation of the gravity equation through different

---

[1] Among the others who have looked for an explanation of the Border Effect, we recall Parsley and Wei (2001), and Evans (2003).

[2] There are some specific points that we aim to improve with respect to Chen's analysis. First, her use of an industrial concentration index (Ellison and Glaeser 1997) computed with American industrial data but used to explain trade patterns within the EU. Second, the estimator that Chen uses; we discuss this in section III.A.

econometric techniques using yearly observations for the 1995-2003 period. Moreover, in appendix I we discuss relative concentration and compare the results with those obtained using absolute concentration, while in appendix II we describe the dataset used and discuss the CIF/FOB pair and industry-specific ratios as proxies of bilateral shipment costs.

II. Theoretical Underpinnings.

As regards the relation between Industrial Concentration (IC) and the Border Effect (BE), Chen affirms that "a low value of the index indicates that the industry in question is not reliant on a specific geographic location, whereas industries which require to produce in specific locations display high values. If some firms are not attached to any specific location (a low value of the index), we expect that they choose their location of production so as to minimise cross-border transaction costs, and as a result, border effects could be magnified. The size of the border effect is therefore expected to be inversely related to the EG index" (Chen 2004, page 206). This rationale recalls the so-called Home Market Effect (HME) as described by Krugman (1980) and Helpman and Krugman (1985). Indeed, we believe that the HME coherently causes the inverse relation between IC and the BE, in the rest of this section we explain this.

In New Economic Geography models, firms maximize profits by settling close to their consumers (to wit, on the same border side) to avoid the border-linked trade costs (see Ottaviano and Thisse 2004, introduction to section 3). The country with the highest consumption of an industry's good will run a trade surplus of that good because it hosts a more than proportionate (with respect to domestic consumption) share of firms: this is the Home Market Effect (Krugman 1980). The surplus is allocated abroad to those countries in which domestic production is not enough to cover domestic consumption. The BE accounts for the size of this surplus relative to domestic consumption, it quantifies how much *Export towards the Representative Partner* and *National Trade* differ given the occurrence of border-linked trade costs. Trade costs make foreign goods more expensive than domestic, for this reason domestic consumption is skewed towards domestic goods (consumers are not indifferent and consumption is said to be Home Biased).[4] Then, interest falls upon the magnitude of the BE in a group of economies supposed to be very integrated in order to weigh the distortion caused by border-linked trade costs.

[4] A remarkable analysis of this issue is in Obstefeld & Rogoff (2000), paragraph 2: "The Puzzle of Home Bias in Trade".

According to this mechanism, the higher the industrial concentration is, the smaller the Border Effect becomes; on the contrary, the lower the industrial concentration, the larger the Border Effect. We now make this relation clearer through a two-case example:

1st case - *High concentration, Small BE.*

Imagine to split the European countries between two groups. Countries in group A demand a low quantity of a certain good, while countries in group B demand a large quantity. In this case producers concentrate more than proportionally in the group of countries which demands more (group B), so serving local demand through local production. Export from group B to group A is relatively high, because in A the demand has to be served through exports. Consequently, the BE (think of it as the ratio of National Trade over Export) is *relatively small* for those countries where the production takes place.

2nd case – *Low concentration, Large BE.*

Let us now think that every country demands the same quantity. In this case, producers settle equally between the two groups and among the countries within each group, therefore there is not concentration. If this is the case, the ratio between National Trade and Export towards the representative partner is high for all the countries and consequently the BE is *relatively large*.

To sum up, we might say that the estimated Border Effect (BE) is a function of Industrial Concentration (IC) and IC is a function of trade costs (TC):

- $BE_s = f(IC_s)$ and $f' < 0$,
- $IC_s = g(TC_s)$; s stands for a specific sector.

How trade-costs determine agglomeration (the function g) depends upon the specific model considered, but the statement that the larger country has more firms is generally supported - under specific conditions- by many New Economic Geography models both in a 2- and multy-country setting (see Combes et al. 2009).

We now look for an empirical proof of the abovementioned inverse relationship. To accomplish this, we proceed into three consecutive stages: 1st) we discuss the most appropriate estimator for the gravity equation given our data sample and comment the Border Effect estimates for the different sectors and time periods considered; 2nd) we compute a Concentration Index per each industrial sector, and 3rd) we eventually check the relation between the Border Effect and industrial concentration.

III. Estimation of the Border Effect

The aim of this section is twofold. First, we discuss the econometric features characterizing the estimation of the gravity equation in order to detect which estimator can provide us with consistent estimates. Second, we discuss the Border Effects' estimates by sector (section III.B) and explain why we prefer the estimation yielded by the Poisson Pseudo Maximum Likelihood Estimator. The estimations in section III.B are performed using a dataset described in details in appendix II. In a nutshell, the data are bilateral exports and production data (the latter for the national trade observations) for the 1995-2003 period; the data are grouped in 20 industrial sectors which we consider throughout our analysis (Table 1).

Table 1. List of sectors.

Sec 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

III.A. Econometric Considerations.

The estimation of a gravity equation can be performed with different estimators and techniques (see Baldwin 2006), the most used estimator is the Ordinary Least Squares (OLS). When the analysis is carried out with sectorial data, the less the data are aggregated, the higher the odds of zero values of the dependent variable is. In this case the implementation of OLS is not to be immediate given the bias generated when the dependent variable is censored. Moreover, since the gravity equation is a log-linear reduced form of an utility maximization problem (where the utility function is a CES) there is need to tackle the issues raised by Silva and Tenreyro (2006). In this section we discuss alternative estimators of a log-linear gravity equation, we start by considering the Tobit as applied by Chen (2004), and we compare it to OLS and to the estimator proposed by Silva and Tenreyro (2006).

Chen estimates the sectorial BE through a cross-section gravity equation, the equation she estimates is:

$$ex_{ij,k} = \alpha_0 + \alpha_i + \alpha_j + \beta_1 y_{i,k} + \beta_2 y_j + \beta_3 contig_{ij} + \beta_4 dist_{ij} + \beta_5 wv_k + \sum_{k=1}^{K} \gamma_k NT_k + \varepsilon_{ij,k}, \qquad (1)$$

where the dependent variable $ex_{ij,k}$ is the logarithm of export from country *i* to country *j* of the k^{th} industrial good, while the covariates are: $y_{i,k}$, the log exporter's production of the k^{th} industrial good; y_j, the log partner's GDP; $dist_{ij}$, the log geographical distance between the i^{th} exporter and the j^{th} partner; wv_k, the weight-to-value ratio (for more information about this variable, see appendix II); $\gamma_k NT_k$, industry-specific National Trade dummies which catch domestic trade; α_i and α_j, exporter-specific and partner-specific dummies which implement the fixed-effects estimator; α_0, the intercept term.

Since the theoretical model is a multiplicative form (see Anderson and van Wincoop 2003), Chen follows the common practice of estimating the relation in its log-linear form. She implements the Tobit Maximum-Likelihood estimator since there are zero-values of the dependent variable (5% of the total) in her dataset.[5] However, given the inconsistency of the Tobit estimator when non-normal and/or heteroskedastic errors occur (this is always the case in trade flows data) and that the bias of OLS is increasing in the percentage of censored observations (under specific conditions this result is due to Goldberg 1981 and Ruud 1986), the choice between Tobit and OLS is to take by weighing the pros and cons of Tobit with respect to OLS given Tobit's likely inconsistency. Indeed, with such little percentage of censored observations and heteroskedastic or/and non-normally distributed residuals, it is likely that the bias of the OLS estimator is less than the Tobit's.[6] Then OLS might still be the right choice.

Besides, Silva and Tenreyro (2006) demonstrate that heteroskedasticity causes the OLS estimates of the coefficients in levels to be biased when the theoretical model is a CES; the bias is due to the logarithmic transformation of the dependent variable. They indicate as solution the use of a Non-Linear Estimator. As regards non-linear estimators, they check the

[5] As Wooldridge (2002) explains, the censoring of the dependent variable can be either a recording problem of the data or derive from the solution of an optimization process. In the second case, we might imagine that for some agents the optimal choice is the corner solution y = 0. Wooldridge (2002) calls this kind of response variable a corner solution outcome and he affirms that "for corner solution outcomes, it makes more sense to call the resulting model a corner solution model"(page 518).

[6] Usually, the percentage of zero observations in Tobit estimations is higher than 30%, for example see Tobin (1958).

suitability of the Non-Linear Least Squares estimator as used in a trade application by Frankel and Wei (1993). They explain why this turns out inefficient as it ignores the heteroskedasticity which is typical of bilateral trade data, and suggest the Poisson Pseudo Maximum Lixelihood (PPML) estimator with a heteroskedasticity-robust covariance matrix. In the following section, we use the PPML estimator and compare it with the more classical Tobit and OLS.

III.B. Estimation of the Border Effect by Industrial Sector.

When one estimates the Border Effect (BE), she needs to pay attention to the estimator used as well as to the controls included in the equation. Indeed, we need to use a *three-way error components model* (Egger and Pfaffermayr 2003) instead of a more common two-way error components model. This is so because the latter accounts for the longitudinal dimension, and this does not allow the estimation of the BE.[7] By using Cluster and Panel Analysis notation, we can express our model as follows:

$$y_{ijt} = \alpha_{c\cdot\exp} + \alpha_{c\cdot par} + \gamma_{year} + x'_{ijt}\beta + \varepsilon_{ijt}, \qquad (2)$$

where $i \in C\cdot\exp$ (exporters' clusters), $j \in C\cdot par$ (partners' clusters), $i \neq j$ except for the National Trade observations for which $i = j$. The model in eq.(2) is known as Cluster Dummy Variables Model (see Cameron and Trivedi 2005); each y_{ijt} falls simultaneously in two different clusters, one for the exporter and one for the partner. The operational version of eq. (2) is the following gravity equation, which is an inter-temporal modified version of Chen's eq.(1) :

$$ex_{ijt,k} = \alpha_0 + \alpha_i + \alpha_j + \alpha_t + \beta_1 y_{it,k} + \beta_2 y_{jt} + \beta_3 contig_{ij} + \beta_4 dist_{ij} + \beta_5 r_{ij,k} + \sum_{k=1}^{25}\gamma_k NT_k + \varepsilon_{ijt,k}. \qquad (3)$$

The dependent variable $ex_{ijt,k}$ is the logarithm of country *i* 's export to country *j* of the k^{th} industrial good in the t^{th} year, while the covariates are: $y_{it,k}$, the log exporter's production; y_{jt}, the log partner's GDP; $dist_{ij}$, the log geographical distance between exporter *i* and partner *j*; $r_{ij,k}$, the pair-specific C.I.F/F.O.B ratio for the k^{th} industrial sector (we use this variable instead of Chen's wage-to-value ratio, see appendix II for more information about this variable); $\gamma_k NT_k$, the industry-specific National Trade dummy which catch the NT observations; α_i and α_j, the exporter and partner-specific dummy variables; α_t and α_0, the year-specific dummies and the intercept term.

---

[7] An interesting discussion about the technical aspects of this kind of estimations is in Andrews et al. (2006).

Our interest is focused on the γ coefficients which are the unique period estimates of the BE. The PPML, Tobit and OLS estimation of eq.(3) are in Table 2, the value of the BE is plotted in Chart 1. By comparing the three different estimations, we point out that:

- The difference between the estimated coefficients and standard errors is trivial across the Tobit and the OLS estimation. We expected this given the little percentage of censored observations.
- As regards the coefficients, the comparison of the Poisson -from one side- with the Tobit and the OLS estimates –from the other- reveals that only for 2 sectors out of 20 the difference is larger than 0.5 (sector 14 and 16); however, it is never higher than 0.6. In all the three estimations the coefficients are correctly signed and the difference in magnitude is not large except for the C.I.F./F.O.B. ratio.
- The ranking of the estimated BEs is the same between the OLS and the Tobit, it is slightly different between the Poisson and the Tobit-OLS estimations. Nonetheless, the three sectors with the highest and lowest BE are still the same across the different estimators used.
- Only for the PPML estimator the null hypothesis of the RESET test (Ramsey 1969) is not rejected; the Normality test of the Tobit residuals rejects the normality assumption so confirming the bias of the Tobit estimation.[8]

On the basis of what discussed in section III.A and the points listed above, we deem as more reliable the PPML estimator.[9] We therefore endorse the PPML estimator and we only discuss PPML estimates in the rest of the paper; when OLS estimates are reported too, they are meant as a robustness check.

[8] In the Tobit estimation normality is tested through a conditional moment test, the null hypothesis is that the disturbances have a normal distribution. The test implements the bootstrap method described by Drukker (2002).

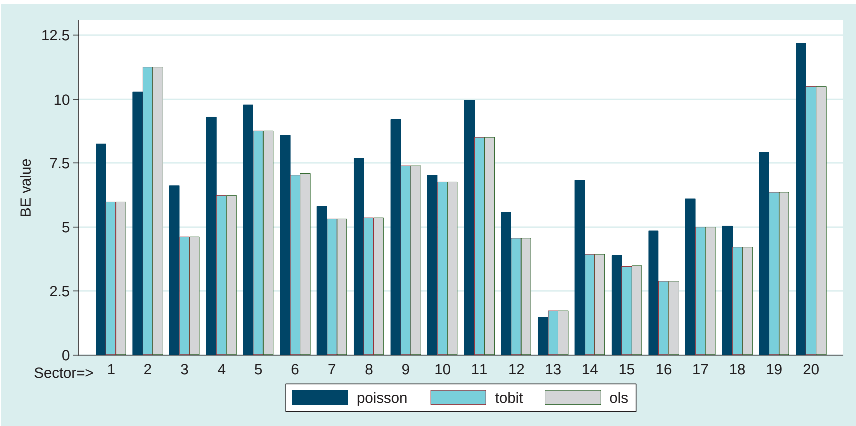
[9] Reasons in favour of the PPML estimator: 1) the likely bias of the Tobit due to non-normally distributed residuals as it emerges from the test, 2) the bias of OLS given Silva-Tenreyro critique, e) the outcome of the RESET test.

Table 2. Border Effect by estimator, 1-period analysis

	Poisson		Tobit		OLS	
Dependent	flow	s.e. robust	Log(flow+1)	s.e. robust	Log(flow+1)	s.e. robust
Production_s	0.92	(0.01)**	0.72	(0.01)**	0.72	(0.01)**
GDP	0.99	(0.17)**	1.11	(0.18)**	1.12	(0.18)**
Distance	-0.85	(0.03)**	-0.99	(0.02)**	-0.99	(0.02)**
CIF/FOB_s	-0.35	(0.10)**	-2.19	(0.07)**	-2.18	(0.07)**
Contiguity	0.35	(0.03)**	0.37	(0.03)**	0.37	(0.03)**
NT_1	2.11	(0.06)**	1.79	(0.12)**	1.79	(0.12)**
NT_2	2.33	(0.05)**	2.42	(0.13)**	2.42	(0.12)**
NT_3	1.89	(0.06)**	1.53	(0.15)**	1.53	(0.15)**
NT_4	2.23	(0.05)**	1.83	(0.13)**	1.83	(0.13)**
NT_5	2.28	(0.05)**	2.17	(0.13)**	2.17	(0.13)**
NT_6	2.15	(0.05)**	1.95	(0.12)**	1.96	(0.12)**
NT_7	1.76	(0.06)**	1.67	(0.14)**	1.67	(0.14)**
NT_8	2.04	(0.05)**	1.68	(0.12)**	1.68	(0.12)**
NT_9	2.22	(0.05)**	2.00	(0.13)**	2.00	(0.13)**
NT_10	1.95	(0.05)**	1.91	(0.12)**	1.91	(0.12)**
NT_11	2.30	(0.04)**	2.14	(0.13)**	2.14	(0.13)**
NT_12	1.72	(0.05)**	1.52	(0.13)**	1.52	(0.13)**
NT_13	0.39	(0.17)*	0.55	(0.19)**	0.55	(0.19)**
NT_14	1.92	(0.05)**	1.37	(0.13)**	1.37	(0.13)**
NT_15	1.36	(0.09)**	1.24	(0.14)**	1.25	(0.14)**
NT_16	1.58	(0.07)**	1.06	(0.13)**	1.06	(0.13)**
NT_17	1.81	(0.05)**	1.61	(0.14)**	1.61	(0.14)**
NT_18	1.62	(0.05)**	1.44	(0.15)**	1.44	(0.15)**
NT_19	2.07	(0.05)**	1.85	(0.13)**	1.85	(0.13)**
NT_20	2.50	(0.04)**	2.35	(0.16)**	2.35	(0.16)**
Obs	27725		27725		27725	
Pseudo R^2	0.95		0.33			
R^2					0.77	
RESET test	0.10		0.00		0.00	
Normality Test			0.00			

- Output of PPML, Tobit and OLS estimation.
- Dependent variable in level for the PPML estimator, dependent variable in log. For the Tobit and OLS estimator.
- The coefficient estimate for “NT_X” is the Border Effect estimate for sector X (X=1,…,20).
- P-value both for the RESET and Normality test.
- Robust standard errors in parentheses.
- Exporter, Partner and Year dummies included.
- * significant at 5%; ** significant at 1%.

Chart 1. Border Effect by estimator, 1-period analysis



Note: Antilog transformation applied to the estimated values.

The estimations in Table 2 provide us with a unique value of the BE, we now look at the evolution of the BE by industrial sector in order to detect which sectors have become more integrated overtime. The BE estimation for every year and each industrial sector provides us with 180 estimated coefficients (20 sectors × 9 years), too many to spot any significant time shift. As an alternative, we prefer to estimate 4-year mean values of the BE which we can later match with the Concentration Index. Consequently, the BE estimate for the first period averages the values for 1995-1998, the estimate for the second averages the values for 2000-2003; the 1999 gap is meant to better catch any shift between the two periods. The estimates are reported in Table 3 and plotted in Chart 2.

The highest BE is in sector 20 ("Electricity, Gas and Water Supply"), while the lowest is in sector 13 ("Manufacture of Office Machinery and Computer"). If we think about these industries in the European context, the ranking makes sense. Indeed, the "Gas, Electricity and Water" industry in Europe is more oriented towards domestic than foreign demand, while the "Manufacture of Office Machinery and Computer" industry is very much export-oriented.

At a first sight integration deepens in all the sectors overtime, the 2nd period BE is less than the 1st period BE. We have computed the linear-restriction Wald test to check whether or not the estimated values are statistically different across the two periods. The Wald statistics are reported in Table 3 ("$\Delta_{p1/p2}$ – test" column) , the null hypothesis of a statistically non-significant difference is rejected 16 out of 20 times at the 0.05 significance level. The outcome of the test suggests that integration deepens in the period 1995-2003, a period which is indeed marked by the effect of relevant integration policies such as the start of the Single European Market in 1993 and the introduction of the Euro in 1999.

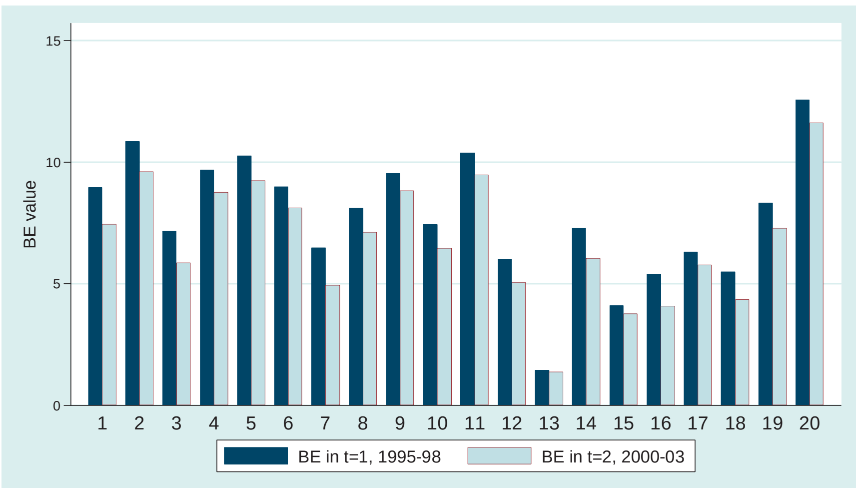
In the next section we discuss concentration of production through a synthetic index. The index matches the BE estimates discussed in this section, and it will be later included in the gravity equation to check the relation depicted in section I.

Table 3. Border Effect, 2-period analysis

Production_s	0.92	(0.01)**				
GDP	0.91	(0.17)**				
Distance	-0.86	(0.03)**				
CIF/FOB_s	-0.37	(0.11)**				
Contiguity	0.35	(0.03)**				
BE estimates in period 1 (1995-98)			*BE estimates in period 2 (2000-03)*			$\Delta_{p1/p2}$ - test
NT_s1_p1	2.19	(0.06)**	NT_s1_p2	2.01	(0.09)**	-0.18*
NT_s2_p1	2.38	(0.05)**	NT_s2_p2	2.26	(0.05)**	-0.12**
NT_s3_p1	1.97	(0.06)**	NT_s3_p2	1.77	(0.08)**	-0.2**
NT_s4_p1	2.27	(0.05)**	NT_s4_p2	2.17	(0.05)**	-0.1**
NT_s5_p1	2.33	(0.05)**	NT_s5_p2	2.22	(0.05)**	-0.11**
NT_s6_p1	2.20	(0.06)**	NT_s6_p2	2.09	(0.07)**	-0.11_
NT_s7_p1	1.87	(0.07)**	NT_s7_p2	1.60	(0.08)**	-0.27**
NT_s8_p1	2.09	(0.05)**	NT_s8_p2	1.96	(0.05)**	-0.13**
NT_s9_p1	2.25	(0.05)**	NT_s9_p2	2.18	(0.05)**	-0.07*
NT_s10_p1	2.01	(0.06)**	NT_s10_p2	1.86	(0.06)**	-0.15*
NT_s11_p1	2.34	(0.05)**	NT_s11_p2	2.25	(0.05)**	-0.09**
NT_s12_p1	1.79	(0.05)**	NT_s12_p2	1.62	(0.05)**	-0.17**
NT_s13_p1	0.37	(0.19)	NT_s13_p2	0.32	(0.40)	-0.05_
NT_s14_p1	1.98	(0.06)**	NT_s14_p2	1.80	(0.07)**	-0.18**
NT_s15_p1	1.41	(0.11)**	NT_s15_p2	1.33	(0.15)**	-0.08_
NT_s16_p1	1.68	(0.08)**	NT_s16_p2	1.41	(0.10)**	-0.27*
NT_s17_p1	1.84	(0.05)**	NT_s17_p2	1.75	(0.06)**	-0.09_
NT_s18_p1	1.70	(0.06)**	NT_s18_p2	1.47	(0.08)**	-0.23**
NT_s19_p1	2.12	(0.06)**	NT_s19_p2	1.99	(0.06)**	-0.13**
NT_s20_p1	2.53	(0.05)**	NT_s20_p2	2.45	(0.05)**	-0.08**
Observations	24411					
Pseudo R2	0.95					

- Output of PPML estimation.
- The coefficient estimate for “NT_sX_pY” is the Border Effect estimate for sector X (X=1,…,20) in period Y (Y=1,2).
- “$\Delta_{p1/p2}$ - test” reports the difference in value between the 2nd and 1st period BE. It also reports the Wald Restriction test whose Ho is “difference between period 1 and 2 is equal to zero”. * indicates rejection of Ho at 5%, ** indicates rejection of Ho at 1%, _ indicates no rejection.
- Robust Standard Errors in parenthesis.
- Exporter, Partner and Year dummies included.
- * significant at 5%; ** significant at 1%.

Chart 2. Border Effect, 2-period analysis



Note: Antilog transformation applied to the estimated values.

IV. Concentration of Industrial Activity.

Productive activities can be spread in different ways throughout a geographical space. Usually, the distribution is uneven and it is possible to observe patterns, such as agglomeration of some activities in specific locations. The causes of agglomeration are different and many models explain such patterns (Baldwin et al. 2003). For the purpose of our analysis, we need to measure the degree of concentration of productive activities through a synthetic index (namely, how much each industrial sector considered is far away from an even distribution over the geographical space), index which we later use to explain the BE in the next section.

As aforementioned, Chen uses the values of the Ellison-Glaeser (EG) Index computed using US data and reported in Ellison and Glaeser (1997) to test the relation between industrial concentration and the BE.[10] Differently from Chen, we think that it is a priority to compute the concentration index with data for the European industries. Ideally, it would be the EG index computed using European plant-level data. However, these data are not available to us and we therefore prefer to use the Theil index since it can be computed with the publicly available data on European employment (Eurostat Regio dataset, see appendix II) for every sector for which we estimate the BE.[11]

Concentration can be measured in Absolute or Relative terms. An industry is concentrated in absolute terms if the bulk of production takes place only in few locations, while it is concentrated in relative terms if its geographic distribution differs from the average spread of productive activities (all other industries considered) among locations. The two measures coincide for a group of geographic units of identical size, but they do not if geo-units differ in size. High relative concentration of an industry implies a high degree of specialisation, while this is not necessarily the case with high absolute concentration. Depending on what we focus on, one measure or the other is the appropriate one. If one studied comparative advantage and specialisation, this clearly would concern relative concentration; but since we are interested in scale economies and trade, then the relevant measure is the absolute concentration of production (Haland et al. 1999).

---

[10] In the group of concentration indices, the EG delivers a higher degree of accurateness because it controls for differences in the size distribution of plants and for differences in the size of geographic areas. However, this accurateness is at a cost since the EG Index requires to compute the Herfindahl index of plant shares (Ellison and Glaeser 1997, page 899). Then, if the activity variable used is employment, data on the number of plants and on employment in each plant need to be available (Ellison and Glaeser 1997, page 906).

[11] It is to bear in mind that the Theil index does not control for the size distribution of plants as the EG does, this delivers a different description of the concentration pattern. There are many concentration indices available to describe the agglomeration of activity across geographic space, good reviews of these indices with applications to Europe are Aiginger and Davies (2004), Cutrini (2006). Combes and Overman (2003) list seven properties which concentration indices should fulfil (page 13). They acknowledge that no measure meets all those criteria simultaneously and that the choice of one index implies neglecting some criteria.

We compute the *Theil index* of Absolute Geographic Concentration. This index is obtained through the formula of the Generalised Class of Entropy Indices when the sensitivity parameter α is equal to 1(see, Brulhart and Traeger 2005).[12] Its formula is:

$$GE(1)_s = \frac{1}{R} \times \sum_{r=1}^{R} \left[\left(\frac{x_{sr}}{\overline{x}_s} \right) \times \log \left(\frac{x_{sr}}{\overline{x}_s} \right) \right] \quad 0 \leq GE(1)_s \leq \log R \tag{4}$$

where $\overline{x}_s = \frac{1}{R} \sum_{r=1}^{R} x_{sr}$ and x_{sr} is activity x (employment, production, value added, etc.) in region r (r = *1,…,R*) in the industrial sector *s* (*s* = *1, …, S*); where each region belongs exclusively to a country *c* (*c* = *1,…,C*). The index has a different formula if one wants to calculate relative concentration, in that case the calculation is not across-sectors independent. A property of the Generalized Class of Entropy Indices is that they can be decomposed additively to tell how much measured concentration derives from within or between groups diversity:

$$GE(\alpha)_s = GE\underline{w}(\alpha)_s + GE\underline{b}(\alpha)_s$$

where the groups can be either countries or macro-regions; we apply this decomposition to gain further insights into the concentration patter.[13]

We use regional employment to compute the Theil index for the twenty industrial sectors which we study. We consider 191 regions (Nuts 2) belonging to 13 EU countries (countries are the groups used for the within/between decomposition). We compute the index both for a unique period of 9 years (1995-2003) and for two sub-periods of 4 years each (1995-1998 and 2000-2003) to assess variations in the concentration pattern.[14] The sectors considered are the same used for the estimation of the border effect (listed in Table 1), detailed information about the dataset and the manipulations enforced are available in appendix II.

The values of the concentration index by sector for the 1-period only analysis, including the decomposition, are reported in Table 4. The sector most concentrated is “Manufacture of Textiles, textile products, leather and footwear” (sector 3) while the least concentrated is “Manufacture of food products, beverages and tobacco” (sector 2). From the between/within decomposition, it emerges that concentration is mainly due to agglomeration within countries. The highest share of within contribution -minimum of between- is for “manuf. of coke, refined petroleum products and nuclear fuel” (sector 6), while the least -maximum of

[12] If $\alpha = 1$ one obtains the Theil Index, while if $\alpha = 2$ one gets another concentration index named Half Square Coefficient of Variation, The more positive the sensitivity parameter is, the more sensitive the index is to activity differences at the top of the distribution, the less is, the more sensitive it is at the bottom of the distribution.
[13] The within component is calculated as the index itself but restricting to observations only within the group.
[14] The employment figures used are an average of the yearly observations within the period.

between- is for "manuf. of wood and of products of wood and cork, …" (sector 4).[15] It comes with no surprise that sector 6 experiences the highest share of within concentration. Indeed, this sector enjoys very large scale economies but it is strategic for each country. Then, production is highly concentrated within countries but not at the European level. On the contrary, one can guess about sector 4 that given its features, a resource based-industry, it is more concentrated in geo-areas which overcome national boundaries.

Table 4. Absolute Concentration, 1-period analysis

Sector Number	*TH_o value*	*TH_o Rank*	*TH_w. value*	*TH_w. Perc.*	*TH_b Value*	*TH_b Perc.*
1	0.69	*6th*	0.58	84.2%	0.11	15.8%
2	0.31	*20th*	0.23	73.1%	0.08	26.9%
3	1.09	*1st*	0.69	63.2%	0.40	36.8%
4	0.43	*19th*	0.25	59.1%	0.18	40.9%
5	0.46	*17th*	0.40	86.7%	0.06	13.3%
6	0.99	*3rd*	0.88	89.1%	0.11	10.9%
7	0.68	*8th*	0.58	84.9%	0.10	15.1%
8	0.49	*16th*	0.37	75.2%	0.12	24.8%
9	0.49	*15th*	0.34	68.4%	0.15	31.6%
10	0.71	*5th*	0.60	84.4%	0.11	15.6%
11	0.54	*14th*	0.40	74.1%	0.14	25.9%
12	0.6	*10th*	0.42	70.8%	0.17	29.2%
13	1.04	*2nd*	0.70	67.2%	0.34	32.8%
14	0.59	*12th*	0.42	71.2%	0.17	28.8%
15	0.69	*7th*	0.57	82.5%	0.12	17.5%
16	0.6	*11th*	0.43	72.1%	0.17	27.9%
17	0.84	*4th*	0.60	71.3%	0.24	28.7%
18	0.63	*9th*	0.53	83.6%	0.10	16.4%
19	0.56	*13th*	0.41	73.6%	0.15	26.4%
20	0.46	*18th*	0.34	74.7%	0.12	25.3%

- "TH_o" stands for Overall Theil Index, "TH_w" stands for Within component of the Overall Theil Index, "TH_b" stands for Between component of the Overall Theil Index. "Perc" stands for percentage of the Overall Theil Index.
- Note that TH_o = TH_w + TH_b.
- "Rank" provides the ranking from the most to the least absolute-concentrated sector.

The index values in Table 4 are for the unique period, however it is interesting to check if the index has changed over time. For this reason, we report the index value in period 1 (1995-1998) and in period 2 (2000-2003) in Table 5; the values are plotted in Chart 3.

At a first sight, it emerges that the within and between components are quite stable over time and the index itself does not seem to change much. Nonetheless, only a rigorous statistical procedure might correctly indicate whether or not the variation of the index is significant. With this purpose we have implemented a Z-test based on a bootstrap procedure to check

[15] When the within contribution is higher than the between, concentration mainly depends by an unequal distribution within countries, while the across-countries distribution is relatively less unequal. To wit, if a sector were located unevenly among countries but equally spread among the regions of each country, then the between contribution would play an exclusive role in determining concentration.

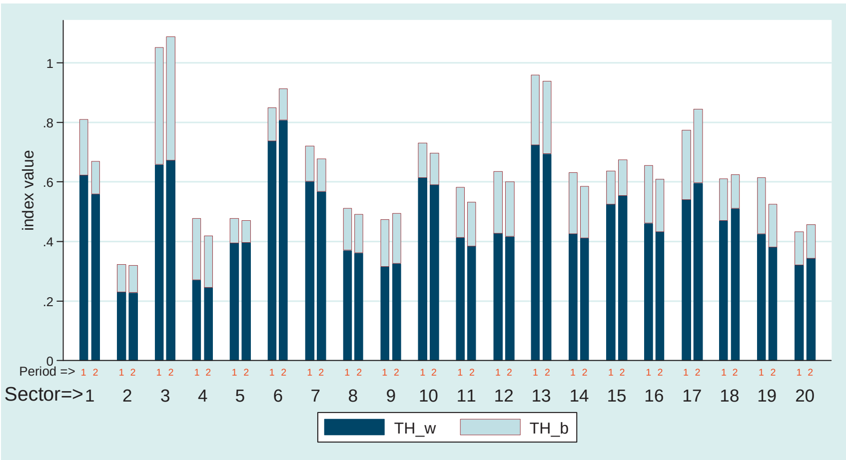
the variation of the index (this testing strategy has been suggested by Biewen 2001).[16] For seven sectors out of twenty-five the difference is statistically significant (cells with asterisk in Table 5 significant). The concentration pattern is stable for the majority of the sectors considered, but concentration decreases in all the sectors for which the variation is statistically significant (sector 1, 7, 8, 12, 14, 16 and 19). This outcome suggests a non-increasing concentration trend in "mining and quarrying" and in "manufacture" as a whole.[17]

Table 5. Absolute Concentration, 2-period analysis

Sec Num ↓	Overall Theil Index			Sec Num ↓	Overall Theil Index		
Period →	p1	p2	$\Delta_{p1/p2}$ - test	Period →	p1	p2	$\Delta_{p1/p2}$ - test
1	0.81	0.67	-0.14 **	11	0.58	0.53	-0.05_
2	0.32	0.32	0.00_	12	0.64	0.60	-0.03 *
3	1.05	1.09	+0.04_	13	0.96	0.94	-0.02_
4	0.48	0.42	-0.06_	14	0.63	0.58	-0.05 *
5	0.48	0.47	-0.01_	15	0.64	0.67	+0.04_
6	0.85	0.91	+0.06_	16	0.65	0.61	-0.05 *
7	0.72	0.68	-0.04 *	17	0.77	0.84	+0.07_
8	0.51	0.49	-0.02 *	18	0.61	0.62	+0.01_
9	0.47	0.49	+0.02_	19	0.61	0.53	-0.09 *
10	0.73	0.70	-0.04_	20	0.43	0.46	+0.03_

- "$\Delta_{p1/p2}$ – test" column reports the difference between the index value in period 1 and 2. It reports also the outcome of the bootstrap test for which Ho is "No significant variation". * indicates rejection of Ho at 5%, ** indicate rejection of Ho at 1%, _ indicates no rejection.

Chart 3. Absolute Concentration, 2-period analysis



Note: For each sector and period, "TH_w" is the within component, while "TH_b" is the between component of the Theil index; the whole height of each bar comes from the sum of the two components and it is equal to the overall amount of the index.

[16] The bootstrap is based on 2500 replications, a block-wise re-sampling (by country) is instructed to adjust for the assumption that the disturbances attached to each observation are iid draws (see Brulhart and Traeger 2005, page 607).

[17] As regards manufacture, this finding emerges in Brulhart and Traeger (2005) as well (page 614).

V. The Relation between Industrial Concentration and the Border Effect.

At this point of our analysis, we can eventually check whether or not the relation between Industrial Concentration (IC) and the Border Effect (BE) holds as we have discussed in section I: the IC index is an explicative of the BE and the relation is inverse. We adopt the same procedure used by Evans (2003) and Chen (2004) which consists in explaining the BE through IC simultaneously at its estimation. Instead of twenty industry-specific NT dummies, we use two variables: a unique NT dummy which does not differentiate across industrial sectors and a sector-specific interaction-term. The estimation output is in Table 6.

Table 6. Border Effect and Abs. Concentration, 1 and 2-period analysis

	Poisson				*OLS*			
	1-period analysis		*2-period analysis*		*1-period analysis*		*2-period analysis*	
Production_s	0.94	(0.01)**	0.94	(0.01)**	0.72	(0.01)**	0.71	(0.01)**
GDP	0.86	(0.25)**	0.80	(0.24)**	1.11	(0.18)**	1.12	(0.18)**
Distance	-0.87	(0.03)**	-0.87	(0.03)**	-0.98	(0.02)**	-0.98	(0.02)**
CIF/FOB_s	-0.33	(0.11)**	-0.36	(0.12)**	-2.3	(0.07)**	-2.33	(0.07)**
Contiguity	0.34	(0.03)**	0.34	(0.03)**	0.38	(0.03)**	0.38	(0.03)**
NT	2.58	(0.05)**			2.37	(0.11)**		
IT	-0.88	(0.05)**			-0.99	(0.16)**		
NT_p1			2.71	(0.06)**			2.65	(0.18)**
NT_p2			2.55	(0.07)**			2.31	(0.17)**
IT_p1			-1.01	(0.08)**			-1.38	(0.27)**
IT_p2			-0.95	(0.09)**			-0.96	(0.27)**
Observations	27725		24411		27725		24411	
R^2					0.77		0.77	
Pseudo R^2	0.94		0.94					

- Output of PPML and OLS estimation.
- As for the 1-period analysis: NT stands for National Trade dummy and IT stands for Interaction Term. As regards the 2-period analysis: NT_p1 stands for National Trade dummy for the 1st period, NT_p2 stands for National Trade dummy for the 2nd period, IT_p1 stands for Interaction term for the 1st period, IT_p2 stands for Interaction term for the 2nd period.
- Exporter, Partner and Year dummies included.
- Robust standard errors in parentheses.
- * significant at 5%; ** significant at 1%.

We present results both for the PPML and OLS estimator, and for the one and two periods analysis. OLS results are to prove the robustness of the relation with respect to the estimator used, but we prefer the PPML estimates for the reasons discussed in section III.B. On the basis of the estimation output, the relation between IC and the BE holds as expected: the higher the concentration, the lower the BE (the negative sign of the estimated interaction terms). This finding is robust with respect to all the specifications reported in Table 6. The Theil index ranges from 1.09 –sector 3- to 0.31 –sector 2- (see Table 4), then for sector 3 the BE is on average (2.58 – (0.88 × ln 1.09))= 2.50, while for sector 2 the BE is (2.58 – (0.88 × ln0.31)) = 3.61 (the values used are from the first column in Table 6).

VI. Conclusions.

In this paper we have studied the relation between the Border Effect and Industrial Concentration. Our aim was to clarify the theoretical linkage between the Border Effect and industrial concentration (we have provided an explanation founded on the Home Market Effect) and to test this relation through an analysis which uses exclusively European data. For this purpose we have estimated a gravity model with different econometric estimators, our findings turn out robust with respect to the different techniques. The robustness of the results casts doubts about the relevance of Silva and Tenreyro's critique to the researcher who has an applied non-theoretical focus. On the whole, the empirical evidence discussed in this paper strongly supports our explanation of the BE through Industrial Concentration founded on the Home Market Effect.

We believe that a so-much robust and clear evidence in favour of the relation discussed depends upon restricting the analysis to a group of homogenous countries which belong to a single market that tends to be a political union. It is likely that the inclusion of trade flows from more heterogeneous countries would have troubled somehow the relation depicted. This should be object of further study to better understand the evolution of trade patterns between the west and the east of the world in an era of strong reallocation of production.

Appendix I. Relative Concentration.

The concentration index discussed in section IV -eq(4)- is for absolute concentration. In this appendix we show the computation of the Theil index for relative concentration. We discuss relative concentration for two reasons: first, because we want to check what happens to the relation between the BE and industrial concentration when one uses Relative concentration, secondly, because by comparing absolute with relative concentration, one can gain important insights into the features of an industrial sector.[18]

The formula of the Theil index when one wants to calculate relative concentration is:

$$RGE(1)_s = \sum_{r=1}^{R} \left[\left(\frac{TY_r}{TTY} \right) \times \left(\frac{\tilde{x}_{r,s}}{\tilde{x}_s} \right) \times \ln \left(\frac{\tilde{x}_{r,s}}{\tilde{x}_s} \right) \right] \qquad (5)$$

where:

- x_{sr} is activity x in region r ($r = 1,\ldots,R$) in industrial sector s ($s = 1, \ldots, S$),
- $TY_r = \sum_{s=1}^{S} x_{r,s}$, region r's total employment (all sectors),
- $TTY = \sum_{r=1}^{R} \sum_{s=1}^{S} x_{r,s}$, total employment (all regions, all sectors),
- $\tilde{x}_{r,s} = \frac{x_{r,s}}{TY_r}$ and $\tilde{x}_s = \frac{TY_s}{TTY}$

Differently from absolute concentration, the computation of the index requires values from all the sectors, and not only for the sector to which the index is referred. It is so because the index controls for the average spread of all the other activities to which the specific industry is compared; this implies that the panel of data used needs to be balanced. The index values are in Table 7, to ease comparison the values of the Absolute concentration index are reported as well (same values as in Table 4).

As expected, the ranking of the sectors for concentration changes, however it does not change much; the correlation coefficient between Absolute and Relative concentration is 0.76. Sector <u>3</u> is the 5th most relative-concentrated, while it is the first absolute-concentrated. Sector <u>13</u> keeps its position as the 2nd most concentrated both in absolute and relative terms. Sector <u>6</u> is the most relative-concentrated, while it is the 3rd most absolute-concentrated. A remarkable difference is only for sector <u>4</u> which goes from the 19th (absolute concentration) to the 10th position (relative concentration).[19]

[18] Haaland et al. (1999) suggest that when some industries are highly ranked in terms of relative concentration but not in absolute terms, this might imply that those industries are important in few smaller countries. A likely explanation is specialization according to comparative advantage. Alternatively, other industries could be highly concentrated in absolute terms but low ranked in relative concentration, then one could deduce that those industries have a bias towards localization in larger countries.

[19] For an example of the information that one can draw by comparing absolute with relative concentration we refer the reader to Brulhart and Traeger (2005) section 4.

Table 7 . Relative versus Absolute concentration, 1-period analysis

Sec Num	R_TH	Rank R_TH	A_TH	Rank A_TH	Sec Num	R_TH	Rank R_TH	A_TH	Rank A_TH
1	0.62	*3*	0.69	*6*	11	0.07	*20*	0.54	*14*
2	0.11	*17*	0.31	*20*	12	0.10	*19*	0.60	*10*
3	0.38	*5*	1.09	1	13	0.68	*2*	1.04	*2*
4	0.20	*10*	0.43	*19*	14	0.12	*16*	0.59	*12*
5	0.15	*14*	0.46	*17*	15	0.31	*8*	0.69	*7*
6	0.74	*1*	0.99	*3*	16	0.19	*11*	0.60	*11*
7	0.23	*9*	0.68	*8*	17	0.36	*6*	0.84	*4*
8	0.11	*18*	0.49	*16*	18	0.42	*4*	0.63	*9*
9	0.18	*12*	0.49	*15*	19	0.14	*15*	0.56	*13*
10	0.36	*7*	0.71	*5*	20	0.18	*13*	0.46	*18*

- "R_TH" stands for Theil Index for Relative Concentration, "A_TH" stands for Theil Index for Absolute Concentration.
- Pearson Correlation Coefficient between R_TH and A_TH is 0.76
- The "Rank R_TH" and "Rank A_TH" columns order the sectors for the magnitude of the concentration index.

There is not much variation between the two indices because we consider geo-units which do not differ extremely in size. If we had considered countries as a whole, then the two indices would have differed much more because the spread of aggregate employment would have been much more unequal.[20]

In Table 8 we report the estimation of the relation between the border effect and Relative concentration for the 1-period analysis. The estimations in Table 8 are as those in Table 6 where we used absolute concentration; once again we report the OLS results as a robustness check. From the estimation output it emerges that the relation holds even when one uses relative concentration, however at a lower extent. Indeed, for absolute concentration the interaction term was -0.88 (PPML 1-period estimation in Table 6) while for relative concentration it is -0.59 (see Table 8). We guess that the relation softens because relative concentration describes localization patterns which are not directly linked to economies of scale and trade (see section IV). However, it still holds steady given the high correlation between the Relative and Absolute index.

[20] To wit, the difference in aggregate employment between Germany and Portugal is much more than the one between a Nuts-2 region of Germany and a Nuts-2 region of Portugal. For a detailed discussion on how the concentration indices vary according to the size of the geo-units, we refer the interested reader to Brulhart and Traeger (2005) and more generally to the "Modifiable Area Unit Problem" literature.

Table 8. Border Effect and Relative Concentration, 1-period analysis

	Poisson		*OLS*	
		s.e. robust		s.e. robust
Production_s	0.96	(0.01)**	0.72	(0.01)**
GDP	0.84	(0.30)**	1.11	(0.18)**
Distance	-0.88	(0.03)**	-0.98	(0.02)**
CIF/FOB_s	-0.33	(0.11)**	-2.30	(0.07)**
Contiguity	0.34	(0.03)**	0.38	(0.03)**
NT	2.17	(0.05)**	1.89	(0.06)**
IT	-0.59	(0.07)**	-0.53	(0.16)**
Observations	27725		27725	
R-squared			0.77	

- Output of PPML and OLS estimations for the 1-period analysis.
- "NT" stands for National Trade dummy, "IT" stands for Interaction Term.
- Robust standard errors in parentheses.
- Exporter, Partner and Year dummy included.
- * significant at 5%; ** significant at 1%.

Appendix II. Data Description.

TRADE DATA for the estimation of the Border Effect

Throughout the paper we consider twenty industrial sectors organised according to ISIC rev. 3/NACE 1.1, these sectors are: *1)* the aggregate for "Mining and Quarrying" (NACE: C, ISIC: 10-14), 2) "Electricity, Gas and Water Supply" (NACE: E, ISIC: 40-41), 3) 18 subgroups of manufacture activities as partition of the "Total Manufacturing" aggregate (NACE: D, ISIC: 15-37; see Table 1 for the list of all the sectors considered). The time range is 1995-2003 both for a data availability issue and to correctly match the trade data with the employment data available for the computation of the concentration indices. The countries comprised in the analysis are Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.[21]

Bilateral exports among the 15 European countries considered (needed for the gravity equation) are extracted from the OECD-Stan Dataset (Bilateral Trade Flows dataset). These data are integrated with the National Trade (NT) observations for every country-sector combination (the NT observations are for the estimation of the Border Effect) and they amount to sectorial production less total sectorial export (to wit, the amount of sectorial production consumed within the country). Sectorial production and total sectorial export are extracted from the OECD-Stan dataset as well (Industrial Analysis section). GDPs in Purchasing Power Parities (regressor in the gravity equation) are extracted from the Penn World Table 6.2. All trade figures, originally in nominal values, are converted in real terms by using the Producer Price Index from OECD Economic Outlook 2007. Weighted distances

[21] Data for Belgium and Luxembourg are recorded together for the so-called Belgium-Luxembourg Economic Union (BLEU).

between and within countries (regressor in the gravity equation) are from the CEPII-Distance dataset (Clair et al. 2004).[22]

Since our benchmark paper is Chen (2004), we wanted to estimate a gravity equation as close as possible to hers. However, her equation includes a variable named Wage-to-Value (wv_k) which was not possible to reproduce in our dataset (see eq.(1)). This is the ratio between *the weight in kilos* and *the value in currency units* of a certain trade flow which reflects the diversity in trade costs borne by different goods; diversity which we deem wrong to neglect. There are two problems with the wage-to-value ratios used by Chen: first, to our knowledge the weight of trade flows is not available in any public dataset, second, it does not differentiate trade costs across pairs.[23] The formula of the weight-to-value ratio used by Chen (2004) is:

$$wv_k = \left[\frac{\sum_j \sum_i weight - ex_{ij,k}}{\sum_j \sum_i value - ex_{ij,k}} \right], \qquad (6)$$

where the numerator is the weight of the k^{th} industrial good ($k = 1, \ldots, K$) exported from the i^{th} to j^{th} country ($i = 1, \ldots, N$ and $j = 1, \ldots, N$) while the denominator is the monetary value of the same trade flow.

Since it was not possible to calculate the Wage-to-Value ratios, we use the C.I.F/F.O.B ratios to account for different trade costs across industrial goods and pairs.[24] The ratio of the C.I.F. over the F.O.B. value provides an approximate measure of the trade costs in which all goods incur. In a recent paper David Hummels and Volodymyr Lugovskyy (2006) analyse accurately the utility of these ratios as indicators of trade costs. They start affirming that "since data on transportation costs are difficult to obtain, in their absence many researchers have turned to indirect measures of transportation costs constructed using matched partner C.I.F./F.O.B. ratios" (page 69), their overall conclusion is that those ratios are poor indicators of trade costs but that, in absence of any other alternative "the matched partner data may be useful as a rough control variable for aggregate bilateral transportation costs" (page 84). Since we had no other alternative, we decided to use them anyway.

We compute the ratios by matching export and import figures by sector. In computing the annual mean value of the ratios, we follow Hummels and Lugovskyy (2006) in ruling out the ratios higher than 2.[25] The highest and lowest value of the annual mean C.I.F./F.O.B. ratios

[22] For more information about the measurement of distance and how it can affect the estimation of BEs, see Head and Mayer (2002).

[23] Chen (2004) uses data which she obtains directly from the European Commission. Unfortunately, It is not possible to find any direct reference to that dataset through the on-line publications of the European Commission agencies.

[24] C.I.F. is the abbreviation for "Cost of Insurance and Freight" while F.O.B. means "Free On Board".

[25] It was not possible to compute any C.I.F.,/F.O.B. ratio for the National Trade observations, then for those observations we set the ratio equal to 1. Alternatively, we could have set it equal to the lowest value computed for

with the correspondent pair for each industrial sector are reported in Table 9. In the same table, the 5th column says if the two countries in the pair share a common border, while the last column in the table reports the decile of the distance distribution in which each pair falls.

We propose an original way to read the information provided by the C.I.F./F.O.B. ratios. First, suppose that the C.I.F./F.O.B. ratios were an unbiased measure of bilateral trade costs. When the trade costs summed up in the C.I.F./F.O.B. ratios depend mainly on distance, the pairs of countries with the lowest value of the ratio should fall in the lowest decile of the distance distribution, while those with the highest value should fall in the highest part of the distance distribution. This perfectly happens for the industrial sector 3 ("Textiles, textile products, leather and footwear") for which the pair with the lowest value of the C.I.F./F.O.B. ratio is France-BelgiumLuxembourg (the distance between France and Belgium-Luxembourg falls in the 1st decile) while the pair with the highest value is Finland-Spain (the distance between Finland and Spain falls in the 10th decile), as well as for sector 12 ("manuf. of machinery and equipment n.e.c.") and 10 ("manuf. of basic metals"). On the contrary, one can imagine that for those pairs for which this does not happen, the trade costs summed up in the C.I.F./F.O.B. ratios do not depend upon distance; this is the case for the majority of the sectors considered.

EMPLOYMENT DATA for the computation of the Concentration Indices

The activity variable used to compute the concentration index is employment. Our figures have a sectorial and geographic dimension (for the list of sectors see Table 1). These data come from the Eurostat Regio dataset (Structural Business Statistics subsection). The Data are at the Nuts-2 level for the period 1995-2003. Observations were available from 1995 to 2006, but we limited extraction to 2003 to select a sample which matches the trade data used for the BE estimation. We started 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.

The most relevant problem with this dataset is the high occurrence of missing values. Indeed, a reliable computation of Absolute Concentration requires a balanced panel of data (namely, the same number of observations for every sector), while this is a requisite for the Relative Concentration Index since its computation requires values from all the sectors.

---

international transactions, but in that case we would have assumed that for a given pair (that one with the lowest value of the ratio) there was no difference between national and international transactions.

Table 9. Maximum and minimum C.I.F./F.O.B. ratios by sector

Industry Number	cif/fob Min	cif/fob Max	Pair of Countries	Contig.	Decile	Industry Number	cif/fob Min	cif/fob Max	Pair of Countries	Contig.	Decile
1	1.02		NL - FR	No	2	11	1.00		PT – NL	No	8
		2	AT - IR	No	7			1.84	IR – FI	No	9
2	1.00		SP – FR	Yes	4	12	1.00		DE – NL	No	1
		1.65	AT – IR	No	7			1.89	GR – IR	No	10
3	1.00		FR - BLEU	yes	1	13	1.00		DE – GE	Yes	1
		1.84	FI –SP	No	10			1.88	AT – DE	No	3
4	1.00		SP – FR	yes	4	14	1.00		IT – SP	No	5
		1.85	GR – IR	no	10			1.68	IR – AT	No	7
5	1.00		IT – PT	No	8	15	1.00		BLEU – AT	No	3
		1.66	GR - BLEU	No	9			1.78	BLEU – UK	No	1
6	1.00		AT - GE	yes	2	16	1.00		FI – GE	No	6
		1.92	IT - DE	No	4			1.64	GR – UK	No	10
7	1.00		IR - NL	No	3	17	1.00		GR – IR	No	10
		1.70	DE - IT	No	6			1.98	IR – SW	No	6
8	1.00		UK – GR	No	10	18	1.04		AT – SW	No	5
		1.68	GR - IR	No	10			1.68	BLEU – AT	No	3
9	1.00		SP - FR	Yes	4	19	1.00		FR – GE	Yes	2
		1.72	NL – PT	No	8			1.98	IR – BLEU	No	3
10	1.01		IT - AT	Yes	2	20	1.00		FR – AT	No	4
		1.74	PT - GR	No	10		1.00		NL – FI	No	7
							1.00		NL – GR	No	9
							1.00		IT – SP	No	5
								1.66	GE – AT	Yes	2

Austria (AT), Belgium-Luxembourg (BLEU), Germany (GE), Denmark(DE), Spain (SP), Finland (FI), France (FR), Greece (GR), Ireland (IR), Italy (IT), Netherlands (NL), Portugal (PT), Sweden (SW), United Kingdom (UK).

To fill up the missing values, the following operations have been sequentially implemented:

1. Cubic Spline Interpolation over the annual observations,
2. Computation of unique values through the average of the within period observations. For the 1-period analysis, the value comes from averaging the yearly observations for the entire 1995-2003 period. On the contrary, for the 2-period analysis, the first period value comes from averaging the yearly observations for 1995-1998, while the second period value comes from averaging the yearly observations for 2000-2003.
3. If the higher-level geo-unit employment figure is available (i.e. employment for the Nuts-1 units), we split this for the within subunits (i.e. the Nuts-2 regions whose values we need) by using as weight the number of local units for each Nuts-2 region over the Nuts-1 total.[26]
4. Deletion of those Nuts-2 regions for which the missing value could not be filled-in (see values in brackets above).

[26] In the Eurostat-Regio dataset, the local units are the firms where the number of employees is recorded. Then, the number of local units is the number of firms in a specific Nuts-2 region.

Bibliographical References.

• Aiginger K. and S. W. Davies (2004), "Industrial Specialisation and Geographic Concentration: two sides of the same coin? Not for the European Union," Journal of Applied Economics, 7(2): 231-248.

• Anderson J. E. and E. van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", American Economic Review, 93(1): 170-192.

• Andrews M., T. Schank and R. Upward (2006), "Practical fixed-effects estimation methods for the three-way error-components model," Stata Journal, 6(4).

• Baldwin R., R. Forslid, P. Martin, G. Ottaviano and F. Robert-Nicoud (2003), Economic Geography and Public Policy, Princeton University Press.

• Baldwin R. (2006), "The Euro's Trade Effects", ECB, Working Paper n° 594.

• Biewen M. (2001), "Bootstrap inference for inequality, mobility and poverty measurement", Journal of Econometrics, 108: 317-342.

• Brulhart M. and R. Traeger (2005), "An account of geographic concentration patterns in Europe", Regional Science and Urban Economics, 35: 597-624.

• Cameron A. C. and P. K. Trivedi (2005), Microeconometrics, Cambridge University Press.

• Chen N. (2004), "Intra-national versus international trade in the European Union: why do national borders matter?", Journal of International Economics, 63: 93-118.

• Clair G., G. Gaulier, T. Mayer and S. Zignago (2004), "Notes on CEPII's distances measures", CEPII.

• Combes P-P. and H. G. Overman (2003), "The Spatial Distribution of Economic Activities in the EU", CEPR, Working Paper n° 3999.

• Combes P-P., T. Mayer and J. F. Thisse (2009), Economic Geography, Princeton University Press.

• Cutrini E. (2006), "The Balassa Index meets the Dissimilarity Theil Index: A Decomposition Methodology for Location Studies", UPM-DE, Quaderni di Ricerca n° 274.

• Drukker D. M. (2002), "Bootsrapping a conditional moment test for normality after tobit estimation", Stata Journal, 2(2): 125-139.

• Egger P. and M. Pfaffermayr (2003), "The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects", Journal Empirical Economics, 28.

• Ellison G. and E. L. Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", Journal of Political Economy, 115(5): 889-927.

• Evans C. L. (2003), "The Economic Significance of National Border Effects", American Economic Review, 93(4): 1291-1312.

• Frankel J. and S. Wei (1993), "Trade Blocs and Currency Blocs", NBER, Working Paper n° 4335.

• Goldberger A. S. (1981), "Linear Regression after Selection", Journal of Econometrics, 15*:* 357–366.

• Haaland J. I., H. J. Kind and K-H. Ulltveit-Moe (1999), "What Determines the Economic Geography of Europe?", CEPR, Discussion Paper n° 2072.

• Head K. and T. Mayer (2002), "Illusory Border Effects: Distance Mis-measurement inflates estimates of the Home Bias in trade", CEPII, Working Paper n° 2002-01.

• Helpman E. and P. Krugman (1985), Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy, MIT Press.

• Hummels D. and V. Lugovskyy (2006), "Are Matched Partner Trade Statistics a Usable Measure of Transportation Costs?", Review of International Economics, 14(1): 69–86.

• Krugman P. (1980), "Scale Economies, Product Differentiation, and The Pattern of Trade", American Economic Review, 70: 950-59.

• McCallum J. (1995), "National Borders Matter: Canada-U.S. Regional Trade Patterns", American Economic Review, 85(3):615-623.

• Ottaviano G. I. P. and J. F. Thisse (2004), "Agglomeration and Economic Geography", in Handbook of Regional and Urban Economics.

• Obstfeld M. and K. Rogoff (2000), "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?", NBER, Working Paper n° 7777.

• Parsley D. C. and S. Wei (2001), "Explaining the border effect: the role of exchange rate Variability, shipping costs, and geography", Journal of International Economics, 55: 87-105.

• Ramsey J. B. (1969), "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", Journal of the Royal Statistical Society, 31: 350–371.

• Ruud P. A. (1986), "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution", Journal of Econometrics, 32: 157–187.

• Silva S. and S. Tenreyro (2006), "The Log of Gravity", The Review of Economics and Statistics, 88(4): 641–658.

• Tobin J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26: 24–36.

• Wooldridge J. M. (2002), Econometric Analysis of Cross-Section and Panel Data, MIT Press.

European Central Bank Working Paper Series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.europa.eu).

1077 "The reception of public signals in financial markets – what if central bank communication becomes stale?" by M. Ehrmann and D. Sondermann, August 2009.

1078 "On the real effects of private equity investment: evidence from new business creation" by A. Popov and P. Roosenboom, August 2009.

1079 "EMU and European government bond market integration" by P. Abad and H. Chuliá, and M. Gómez-Puig, August 2009.

1080 "Productivity and job flows: heterogeneity of new hires and continuing jobs in the business cycle" by J. Kilponen and J. Vanhala, August 2009.

1081 "Liquidity premia in German government bonds" by J. W. Ejsing and J. Sihvonen, August 2009.

1082 "Disagreement among forecasters in G7 countries" by J. Dovern, U. Fritsche and J. Slacalek, August 2009.

1083 "Evaluating microfoundations for aggregate price rigidities: evidence from matched firm-level data on product prices and unit labor cost" by M. Carlsson and O. Nordström Skans, August 2009.

1084 "How are firms' wages and prices linked: survey evidence in Europe" by M. Druant, S. Fabiani, G. Kezdi, A. Lamo, F. Martins and R. Sabbatini, August 2009.

1085 "An empirical study on the decoupling movements between corporate bond and CDS spreads" by I. Alexopoulou, M. Andersson and O. M. Georgescu, August 2009.

1086 "Euro area money demand: empirical evidence on the role of equity and labour markets" by G. J. de Bondt, September 2009.

1087 "Modelling global trade flows: results from a GVAR model" by M. Bussière, A. Chudik and G. Sestieri, September 2009.

1088 "Inflation perceptions and expectations in the euro area: the role of news" by C. Badarinza and M. Buchmann, September 2009.

1089 "The effects of monetary policy on unemployment dynamics under model uncertainty: evidence from the US and the euro area" by C. Altavilla and M. Ciccarelli, September 2009.

1090 "New Keynesian versus old Keynesian government spending multipliers" by J. F. Cogan, T. Cwik, J. B. Taylor and V. Wieland, September 2009.

1091 "Money talks" by M. Hoerova, C. Monnet and T. Temzelides, September 2009.

1092 "Inflation and output volatility under asymmetric incomplete information" by G. Carboni and M. Ellison, September 2009.

1093 "Determinants of government bond spreads in new EU countries" by I. Alexopoulou, I. Bunda and A. Ferrando, September 2009.

1094 "Signals from housing and lending booms" by I. Bunda and M. Ca'Zorzi, September 2009.

1095 "Memories of high inflation" by M. Ehrmann and P. Tzamourani, September 2009.

1096 "The determinants of bank capital structure" by R. Gropp and F. Heider, September 2009.

1097 “Monetary and fiscal policy aspects of indirect tax changes in a monetary union” by A. Lipińska and L. von Thadden, October 2009.

1098 “Gauging the effectiveness of quantitative forward guidance: evidence from three inflation targeters” by M. Andersson and B. Hofmann, October 2009.

1099 “Public and private sector wages interactions in a general equilibrium model” by G. Fernàndez de Córdoba, J.J. Pérez and J. L. Torres, October 2009.

1100 “Weak and strong cross section dependence and estimation of large panels” by A. Chudik, M. Hashem Pesaran and E. Tosetti, October 2009.

1101 “Fiscal variables and bond spreads – evidence from eastern European countries and Turkey ” by C. Nickel, P. C. Rother and J. C. Rülke, October 2009.

1102 “Wage-setting behaviour in France: additional evidence from an ad-hoc survey ” by J. Montornés and J.-B. Sauner-Leroy, October 2009.

1103 “Inter-industry wage differentials: how much does rent sharing matter?” by P. Du Caju, F. Rycx and I. Tojerow, October 2009.

1104 “Pass-through of external shocks along the pricing chain: a panel estimation approach for the euro area” by B. Landau and F. Skudelny, November 2009.

1105 “Downward nominal and real wage rigidity: survey evidence from European firms ” by J. Babecký, P. Du Caju, T. Kosma, M. Lawless, J. Messina and T. Rõõm, November 2009.

1106 “The margins of labour cost adjustment: survey evidence from European firms ” by J. Babecký, P. Du Caju, T. Kosma, M. Lawless, J. Messina and T. Rõõm, November 2009.

1107 “Interbank lending, credit risk premia and collateral” by F. Heider and M. Hoerova, November 2009.

1108 “The role of financial variables in predicting economic activity” by R. Espinoza, F. Fornari and M. J. Lombardi, November 2009.

1109 “What triggers prolonged inflation regimes? A historical analysis.” by I. Vansteenkiste, November 2009.

1110 “Putting the New Keynesian DSGE model to the real-time forecasting test” by M. Kolasa, M. Rubaszek and P. Skrzypczyński, November 2009.

1111 “A stable model for euro area money demand: revisiting the role of wealth” by A. Beyer, November 2009.

1112 “Risk spillover among hedge funds: the role of redemptions and fund failures” by B. Klaus and B. Rzepkowski, November 2009.

1113 “Volatility spillovers and contagion from mature to emerging stock markets” by J. Beirne, G. M. Caporale, M. Schulze-Ghattas and N. Spagnolo, November 2009.

1114 “Explaining government revenue windfalls and shortfalls: an analysis for selected EU countries” by R. Morris, C. Rodrigues Braz, F. de Castro, S. Jonk, J. Kremer, S. Linehan, M. Rosaria Marino, C. Schalck and O. Tkacevs.

1115 “Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors” by L. Alessi, M. Barigozzi and M. Capasso, November 2009.

1116 “Sectorial border effects in the European single market: an explanation through industrial concentration” by G. Cafiso, November 2009.