Sectoral border effects and the geographic concentration of production

Gianluca Cafiso

► To cite this version:

Gianluca Cafiso. Sectoral border effects and the geographic concentration of production. Review of World Economics, 2011, 147 (3), pp.543-566. 10.1007/s10290-011-0091-0 . hal-00673676

HAL Id: hal-00673676

<https://hal.science/hal-00673676>

Submitted on 24 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sectoral Border Effects and the Geographic Concentration of Production

*Gianluca Cafiso**

Abstract

The object of this paper is to study the relationship between the border effect and the geographic concentration of production. We explain this relationship through the home market effect and test the robustness of this explanation by using an analysis that considers the European single market. A sectoral gravity equation is estimated with different econometric estimators; in particular, we discuss a recently suggested estimator for log-linear CES models. Overall, our findings suggest a steady relationship between the border effect and the concentration of production. Furthermore, the analysis of concentration through a synthetic index provides us with valuable insights into the structure of the European industry.

Keywords: Trade, Border Effect, Geographic Concentration,
Home Market Effect, European Single Market.
JEL Codes: F10, F12, F15.

December 2010

* University of Catania - Economics Department, 95129 Corso Italia 55, Catania - Italy. Mail to: gcafiso@unict.it. This work is partially drawn from my PhD thesis at the University of Pavia. I thank Paolo Epifani and Rodolfo Helg for their guidance and useful comments. I acknowledge financial support from the Italian Ministry of Education and Research by way of a PhD scholarship.

I. Introduction

The degree of integration among trade partners is of interest to trade economists because research in international trade looks for the causes of limited trade integration (Head and Mayer 2000, Nitsch 2000). One frequently used measure of trade integration is the border effect (BE), which is estimated using a gravity equation (McCallum 1995). The BE accounts for the difference between *national trade* and *trade with foreign partners*, whereas in the case of full integration, there is no difference between the two. At an early stage, research focused on how to correctly estimate the BE (Anderson and van Wincoop 2003); later studies concentrated more on its explanation. Among the papers that attempt to explain the BE, we believe that Chen (2004) provides the most comprehensive analysis.

The theoretical model in Anderson and van Wincoop (2003) shows that the BE is caused by the combined effect of the elasticity of substitution (between varieties produced in different countries) and the tariff equivalent of border costs.¹ The elasticity of substitution component of the BE is usually overlooked because it concerns individuals' preferences, whereas interest focuses on impediments to trade. As for the border costs component, Chen (2004) points out that some specific border related factors, such as exchange rate volatility and tariff and non tariff barriers, are not explicative of the BE. Next, she turns to *trade cost* variables and *behavioural response to trade cost* variables. Among the behavioural response to trade cost variables, Chen includes an index of geographic concentration of production activities. She concludes that concentration explains the BE. Our interest focuses on the relationship between concentration and the BE.

A concentration index mirrors behavioural responses to trade costs because it reflects the outcomes of firms' location decisions given the presence of trade costs. Indeed, new economic geography (NEG) models suggest that firms settle in ways that minimise trade costs (Ottaviano and Thisse 2004). Firms' location decisions inevitably influence trade exchanges among countries; the ratio of internal to external trade changes in any country affected by firms' decisions. From this perspective, a concentration index may explain the BE, and concentration changes should influence the BE. This is a relevant improvement in our understanding of the BE.

The contribution of this paper lies in its clarification of the relationship between the BE and the concentration of production. We provide an explanation for this relationship, verify

¹ In the extreme case in which there are no border-related trade costs, a BE can still emerge when nationals are biased towards domestic products (high elasticity of substitution). In this case, the BE functions more as an indicator of home-biased preferences than an indicator of trade integration with respect to border-related trade costs. In empirical analyses, however, it is difficult to assess the relevance of the elasticity of substitution; more on this phenomenon can be found in Hummels (2001).

whether it emerges in an enhanced analysis that corrects some relevant points in Chen's analysis and test whether our explanation of the relationship is supported by the data.² Moreover, the analysis of geographic concentration provides us with valuable insights into the structure of the European industry and its evolution.

The paper is organised as follows: Section II explains why the relationship between the BE and concentration emerges; Section III estimates the BE by industrial sector and discusses the relevant econometrics; Section IV studies the concentration of production in the EU to compute an index for use in the analysis; Section V verifies whether the relationship emerges in our enhanced analytical framework; Section VI applies our explanation of the relationship to the data; and Section VII draws conclusions. The appendix describes the data used.

II. Explaining the Relationship between Concentration and the Border Effect

Profit maximisation in NEG models requires firms to settle close to consumers in order to reduce trade costs (Ottaviano and Thisse 2004). Our analysis considers border-related trade costs and posits that firms settle on the same offside of the border as consumers to avoid such costs. The country with the highest consumption of an industry's goods will therefore run a trade surplus in those goods because it hosts a more than proportional share of firms (with respect to its domestic consumption): this is the home market effect (HME) (Krugman 1980, Helpman and Krugman 1985). The surplus is allocated abroad to those countries in which domestic production is not sufficient to cover domestic consumption. The BE accounts for the size of this surplus relative to domestic consumption; it quantifies how much *National Trade* and *Export to the Representative Partner* differ. According to this mechanism, the higher the concentration is, the smaller the BE becomes, and vice versa. The HME seems to coherently explain the inverse relationship between the BE and concentration. From this perspective, the BE is endogenous with respect to firms' location decisions. We clarify this relationship through a two-case example.

1st case - High concentration, Small BE.

Imagine splitting the European countries into two groups. Countries in group A demand a low quantity of a certain good, whereas countries in group B demand a large quantity. In this case, producers concentrate more than proportionally in the countries that demand more (group B), thus serving local demand through local production. Export from group

² We improve Chen's (2004) analysis as follows: i) we employ a concentration index computed using European employment data, whereas Chen uses an index (Ellison and Glaeser 1997) computed with American industrial data to explain trade patterns within the EU; ii) we use a more appropriate estimator for the gravity equation considered. In addition to what Chen does, we explain why the relationship between the BE and concentration emerges and develop a study of concentration functional to the analysis.

B to group A is relatively high because in A the demand has to be filled through imports. 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.

Now imagine 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 no concentration. If this is the case, the ratio of national trade to export is high for all of the countries and, consequently, the BE is *relatively large*.

In sum, the BE is a function of the concentration of production (CP) and of other factors $\vec{z}$; whereas, the CP is a function of trade costs (TC), concentration of demand (CD) and other factors ($\vec{x}$).

- $BE_k = f(CP_k, \vec{z})$ where $\partial BE_k / \partial CP_k < 0$.
- $CP_k = g(TC_k, CD_k, \vec{x})$ where $\partial CP_k / \partial TC_k < 0$, $\partial CP_k / \partial CD_k > 0$.

Because decreasing trade costs strengthen the HME, the derivative of CP with respect to trade costs is negative. CP implies that a higher level of trade exchanges is needed to serve those markets where production does not take place (the relationship); CP occurs because the demand is concentrated as prescribed by the HME (explanation of the relationship). If this explanation is valid, CP depends upon CD, and the BE should therefore be related to CD as well; Section VI considers this in more detail.

III. Estimation of the Border Effect

The aim of this section is twofold. First, it reviews the potential estimators for the gravity equation (Section III.A). Second, it discusses the BE estimates by sector and explains why the poisson pseudo maximum likelihood (PPML) estimator is preferable (Section III.B). The estimations in Section III.B use a data set described in the appendix. In a nutshell, the data set consists of bilateral exports and production data for the 1995–2003 period for the EU countries before the 2004 enlargement; the data are grouped in 20 industrial sectors that we consider throughout our analysis (Table 1).

Table 1. List of sectors.

<i>Sec Number</i>	<i>Nace 1.1/ ISIC rev 3</i>	<i>Sector Name</i>
1	C, 10_14	Mining and quarrying
2	D, 15_16	M. of food products, beverages and tobacco
3	D, 17_19	M. of Textiles, textile products, leather and footwear
4	D, 20	M. of wood and of products of wood and cork, except furniture; M. of articles of straw and plaiting materials
5	D, 21_22	M. of pulp, paper and paper products; publishing and printing
6	D, 23	M. of coke, refined petroleum products and nuclear fuel
7	D, 24	M. of chemicals and chemical products
8	D, 25	M. of rubber and plastic products
9	D, 26	M. of other non-metallic mineral products
10	D, 27	M. of basic metals
11	D, 28	M. of fabricated metal products, except machinery and equipment
12	D, 29	M. of machinery and equipment n.e.c.
13	D, 30	M. of office machinery and computers
14	D, 31	M. of electrical machinery and apparatus n.e.c.
15	D, 32	M. of radio, television and communication equipment and apparatus
16	D, 33	M. of medical, precision and optical instruments, watches and clocks
17	D, 34	M. of motor vehicles, trailers and semi-trailers
18	D, 35	M. of other transport equipment
19	D, 36	M. of furniture; manufacturing n.e.c
20	E, 40_41	Electricity, gas and water supply

- "M." stands for "Manufacture"

III.A. Econometric Considerations

The estimation of a gravity equation can be performed with different estimators and techniques (Baldwin and Taglioni 2006, Silva and Tenreyro 2006); the most commonly used estimator is the ordinary least squares (OLS). When the analysis is carried out with sectoral data, the less the data are aggregated, the higher the probability of a zero value-dependent variable. In this case, the use of OLS cannot be immediate given the bias generated when the dependent variable is censored. Moreover, because the gravity equation is a log-linear reduced form of a utility maximisation problem (where the utility function is a CES), the issue raised by Silva and Tenreyro (2006) is relevant. This section discusses alternative estimators of a log-linear gravity equation, beginning with a consideration of the Tobit as applied by Chen (2004) and then comparing it to OLS and to the estimator proposed by Silva and Tenreyro (2006).

Chen estimates the BE through a gravity equation that uses sectoral data: ³

$$ex_{ij,k} = \alpha_i + \alpha_j + \beta_1 y_{i,k} + \beta_2 y_j + \beta_3 adj_{ij} + \beta_4 dis_{ij} + \beta_5 wv_k + \beta_6 NT + \varepsilon_{ij,k} \quad (1)$$

³ This is the equation that Chen (2004) estimates to generate the output in Table 1 (columns 4, 5, 6) of her paper; this produces a unique estimate of the BE. To estimate the BE by sector (Table 2 in Chen's paper), she uses a gravity equation for each sector, a strategy not followed in our analysis.

where the dependent variable $ex_{ij,k}$ is the logarithm of export from country i to country j of good k , whereas the covariates are as follows: $y_{i,k}$, the log exporter's production of good k ; y_j , the log partner's GDP; adj_{ij} , a control for geographic proximity between exporter i and partner j ; dis_{ij} , the log geographic distance between exporter i and partner j ; wv_k , the weight-to-value ratio (for more information about this variable, see the appendix); NT , a national trade dummy that quantifies the border effect; α_i and α_j , exporter and partner fixed-effects that account for the multilateral resistance terms (Baldwin and Taglioni 2006).

Because the theoretical model is a multiplicative form (Anderson and van Wincoop 2003), Chen follows the common practice of estimating the log-linear relation. She applies the Tobit estimator (Tobin 1958) because the dependent variable is censored to zero (5% of the total) in her data set.⁴ However, given the inconsistency of the Tobit estimator when non-normal and/or heteroskedastic errors occur (Pagan and Vella 1989) and that the bias of OLS is increasing in the percentage of censored observations (Goldberg 1981, Ruud 1986), the choice between Tobit and OLS must entail weighing the pros and cons of Tobit with respect to OLS given Tobit's likely inconsistency. Indeed, given such a small percentage of censored observations and heteroskedastic and/or non-normally distributed residuals, it is likely that the bias of the OLS estimator is less than the Tobit's; OLS might still be the right choice.⁵ Nonetheless, Silva and Tenreyro (2006) demonstrate that heteroskedasticity causes the OLS estimates of the coefficients in level to be biased when the theoretical model is a CES; the bias is due to the logarithmic transformation of the dependent variable. They indicate the use of the PPML estimator with a heteroskedasticity-robust covariance matrix as a solution (Cameron and Trivedi 2005). In the following section, we apply the PPML estimator as suggested by Silva and Tenreyro, and compare it with the Tobit and OLS.

When estimating the BE, one must choose carefully the estimator to use as well as the fixed-effects to include in the gravity equation. The BE is estimated by a dummy that captures the national trade observations ($ex_{ii,kt}$); as a consequence, it is only possible to include exporter- and partner-specific fixed-effects, not pair fixed-effects.⁶ Fixed-effects account for the multilateral resistance terms (Anderson and van Wincoop 2003); Baldwin and Taglioni (2006) argue that the fixed-effects' dimensions should match those of the data set used. On this

⁴ 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 optimisation process. In the second case, we might imagine that for some agents the optimal choice is the corner solution $y = 0$; Wooldridge calls this kind of response variable "corner solution outcome" (page 518).

⁵ Usually, the percentage of zero observations in Tobit estimations is higher than 30%; for example, see Tobin (1958).

⁶ See Cafiso (2008) Section IV.A for further discussion.

ground, if one uses a data set with the pair, sector and year dimensions and wants to estimate the BE, one should include exporter/sector/year and partner/sector/year dummies. This is computationally problematic because there may be thousands of dummies, as in the present case. We therefore decided to include exporter/year and partner/year dummies (sectoral dimension left out) in the following estimations to make the analysis feasible.⁷

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

In this section, we discuss the estimation of the BE by sector using the PPML, Tobit and OLS estimators. We use a modified version of Chen's equation (1), which includes the temporal dimension and estimates all of the sectoral BEs simultaneously in one regression:

$$ex_{ij,kt} = \alpha_{i,t} + \alpha_{j,t} + \beta_1 y_{i,kt} + \beta_2 y_{j,t} + \beta_3 adj_{ij} + \beta_4 dis_{ij} + \beta_5 cfr_{ij,k} + \sum_{k=1}^K \gamma_k NT_k + \varepsilon_{ij,kt} \quad (2)$$

The dependent variable $ex_{ij,kt}$ is the logarithm of country i 's export to country j of good k in year t , while the covariates are as follows: $y_{i,kt}$, the log exporter's production of good k in year t ; $y_{j,t}$, the log partner's GDP in year t ; adj_{ij} , a control for geographic proximity between exporter i and partner j ; dis_{ij} , the log geographic distance between exporter i and partner j ; $cfr_{ij,k}$, the pair-specific CIF/FOB ratio for good k ; NT_k , the industry-specific national trade dummies that quantify the BE; $\alpha_{i,t}$ and $\alpha_{j,t}$, the exporter/year and partner/year fixed-effects.

Our interest is focused on the γ_k coefficients, which are the unique period estimates of the BE. The PPML, Tobit and OLS estimations of eq.(2) are in Table 2; the value of the BE is plotted in Figure 1. By comparing the three different estimations, we note the following.

- The difference between the estimated coefficients and standard errors is negligible across the Tobit and OLS estimations. We expected this, given the small percentage of censored observations.
- As regards the BE estimates, the comparison of the Poisson with the Tobit and the OLS estimates reveals that the maximum standard deviation is 0.25 for Sector 14, whereas the

⁷ In the case of "exporter/sector/year + partner/sector/year" fixed-effects, we need to include 3,952 dummies. It was not possible to generate any estimation output with this pattern of fixed-effects, so we preferred "exporter/year + partner/year" fixed-effects. Nonetheless, we attempted the use of "exporter/sector/year + partner/sector/year" fixed-effects in a smaller sample (only 5 sectors, from 27,725 to 6,860 observations) in order to compare the estimation output. The estimation output is alike, our results are therefore robust regardless of the the fixed-effects pattern used. A possible explanation is that there is no so much heterogeneity in the real data even if theoretically the MRTs need to vary by each dimension.

minimum is 0.02 for Sector 10; the across-sectors standard deviation is 0.347 for the PPML estimation. In all three estimations, the coefficients are correctly signed, and the difference in magnitude is small but for the CIF/FOB ratio.

- The ranking of the estimated BEs is the same between the OLS and the Tobit but slightly different between the Poisson and the Tobit/OLS estimations. The group of the four sectors with the highest and the three with the lowest BEs is constituted by the same elements across the different estimations.
- The null hypothesis of the RESET test (Ramsey 1969) is rejected in all instances except that of the PPML estimator. The normality test (Drukker 2002) rejects the normality of the Tobit residuals; this result confirms the bias of the Tobit estimation.⁸

On the basis of the discussion of Section III.A and the points listed above, we deem the PPML estimator most reliable.⁹ We therefore endorse the PPML estimator and discuss exclusively its estimates throughout the remainder of the paper; when OLS estimates are reported, they serve only as a robustness check.

Figure 1. BE by estimator, 1-period analysis

Note: Antilog transformation applied to the estimated values.

⁸ In the Tobit estimation, normality is tested through a conditional moment test, and the null hypothesis is that the disturbances have a normal distribution.

⁹ Reasons in favour of the PPML estimator: 1) the likely bias of the Tobit due to non-normally distributed residuals (as from the test) and Silva-Tenreiro's critique, 2) the bias of OLS given the occurrence of zero-trade observations and Silva-Tenreiro's critique, 3) the outcome of the RESET test.

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

POISSON $EXP_{ij,kt}$			TOBIT $\log(EXP_{ij,kt} + 1)$			OLS $\log(EXP_{ij,kt} + 1)$		
PROD _{i,kt}	0.92	(0.01)**	PROD _{i,kt}	0.71	(0.01)**	PROD _{i,kt}	0.71	(0.01)**
GDP _{j,t}	0.58	(0.05)**	GDP _{j,t}	0.91	(0.05)**	GDP _{j,t}	0.79	(0.03)**
DISTANCE _{ij}	-0.86	(0.03)**	DISTANCE _{ij}	-0.98	(0.02)**	DISTANCE _{ij}	-0.98	(0.02)**
CIF/FOB _{ij,k}	-0.37	(0.11)**	CIF/FOB _{ij,k}	-2.34	(0.07)**	CIF/FOB _{ij,k}	-2.32	(0.07)**
CONTIG _{ij}	0.35	(0.03)**	CONTIG _{ij}	0.38	(0.03)**	CONTIG _{ij}	0.38	(0.03)**
NT ₂₀	2.50	(0.04)**	NT ₂	2.43	(0.12)**	NT ₂	2.43	(0.13)**
NT ₂	2.33	(0.04)**	NT ₂₀	2.36	(0.16)**	NT ₂₀	2.36	(0.16)**
NT ₁₁	2.30	(0.04)**	NT ₅	2.17	(0.13)**	NT ₅	2.17	(0.13)**
NT ₅	2.28	(0.04)**	NT ₁₁	2.13	(0.13)**	NT ₁₁	2.13	(0.13)**
NT ₄	2.23	(0.05)**	NT ₉	2.00	(0.13)**	NT ₉	2.00	(0.13)**
NT ₉	2.22	(0.04)**	NT ₆	1.95	(0.12)**	NT ₆	1.95	(0.12)**
NT ₆	2.15	(0.05)**	NT ₁₀	1.90	(0.12)**	NT ₁₀	1.91	(0.12)**
NT ₁	2.11	(0.06)**	NT ₁₉	1.85	(0.13)**	NT ₁₉	1.85	(0.13)**
NT ₁₉	2.06	(0.05)**	NT ₄	1.83	(0.13)**	NT ₄	1.83	(0.13)**
NT ₈	2.04	(0.05)**	NT ₁	1.79	(0.12)**	NT ₁	1.79	(0.12)**
NT ₁₀	1.95	(0.05)**	NT ₇	1.67	(0.14)**	NT ₇	1.67	(0.14)**
NT ₁₄	1.91	(0.05)**	NT ₈	1.67	(0.12)**	NT ₈	1.67	(0.12)**
NT ₃	1.89	(0.05)**	NT ₁₇	1.60	(0.14)**	NT ₁₇	1.60	(0.14)**
NT ₁₇	1.81	(0.05)**	NT ₃	1.52	(0.15)**	NT ₃	1.52	(0.16)**
NT ₇	1.76	(0.06)**	NT ₁₂	1.52	(0.13)**	NT ₁₂	1.52	(0.13)**
NT ₁₂	1.72	(0.05)**	NT ₁₈	1.43	(0.15)**	NT ₁₈	1.43	(0.15)**
NT ₁₈	1.62	(0.05)**	NT ₁₄	1.36	(0.13)**	NT ₁₄	1.36	(0.13)**
NT ₁₆	1.57	(0.07)**	NT ₁₅	1.23	(0.14)**	NT ₁₅	1.23	(0.14)**
NT ₁₅	1.36	(0.09)**	NT ₁₆	1.05	(0.13)**	NT ₁₆	1.05	(0.13)**
NT ₁₃	0.38	(0.17)*	NT ₁₃	0.53	(0.19)**	NT ₁₃	0.53	(0.19)**
observations	27,725		observations	27,725		observations	27,725	
pseudo R ²	0.95		pseudo R ²	0.32		pseudo R ²		
R ²			R ²			R ²	0.77	
Reset Test	0.148		Reset Test	0.00		Reset Test	0.00	
Normality Test			Normality Test	0.00		Normality Test		

- 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 BE estimate for sector X (X=1,...,20).

- P-value both for the RESET and the normality test. - Robust standard errors in parentheses. - BE ordered from larger to smaller.

- Fixed-Effects for MRTs included: Exporter/Year + Partner/Year. - * significant at 5%; ** significant at 1%.

The estimations in Table 2 provided us with a unique value of the BE. We now consider the BE evolution by industrial sector in order to detect which sectors have become more integrated over time. As an alternative to the year-by-year estimation of the BE, which provides too many values, we estimate 4-year mean values. Consequently, the BE estimate for the first period averages the values for 1995–1998, and the estimate for the second averages the values for 2000–2003; the 1999 gap serves to highlight any shift between the two periods. The estimates are reported in Table 3 and plotted in Figure 2.

The highest BE is in Sector 20 ("Electricity, Gas and Water Supply"), whereas the lowest is in Sector 13 ("Manufacture of Office Machinery and Computer"). Considering 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, whereas the

“Manufacture of Office Machinery and Computer” industry is very much export-oriented.¹⁰ At a first glance, integration deepens in all the sectors overtime; the 2nd period BE is less than the 1st period BE. We computed the linear-restriction Wald test to determine whether the estimated values are statistically different across the two periods. The test results are in Table 3 (column 6), and the null hypothesis (statistically non-significant difference) is rejected in sixteen out of twenty sectors. The outcome of the test suggests that integration deepens in 1995–2003, a period that was indeed marked by the effects 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 the concentration of production through a synthetic index. The index matches the BE estimates discussed in this section, and it will be included later in the gravity equation to check the relation discussed in Section I.

Table 3. Border Effect, 2-period analysis

	1	2	3	4	5	6
PROD _{i,kt}	0.92	(0.01)**	CIF/FOB _{ij,k}	-0.40	(0.11)**	
GDP _{j,t}	0.57	(0.05)**	CONTIG _{ij}	0.34	(0.03)**	
DISTANCE _{ij}	-0.86	(0.03)**				
BE estimates in period 1 (1995-98)			BE estimates in period 2 (2000-03)			$\Delta_{p1/p2}$ - test
NT_1	2.18	(0.06)**	NT_1	2.01	(0.09)**	-0.17*
NT_2	2.38	(0.05)**	NT_2	2.26	(0.05)**	-0.12**
NT_3	1.97	(0.06)**	NT_3	1.76	(0.08)**	-0.21**
NT_4	2.26	(0.05)**	NT_4	2.17	(0.05)**	-0.09**
NT_5	2.32	(0.05)**	NT_5	2.22	(0.05)**	-0.10**
NT_6	2.19	(0.06)**	NT_6	2.09	(0.06)**	-0.10
NT_7	1.86	(0.07)**	NT_7	1.60	(0.08)**	-0.26**
NT_8	2.09	(0.05)**	NT_8	1.96	(0.05)**	-0.13**
NT_9	2.25	(0.05)**	NT_9	2.18	(0.05)**	-0.07*
NT_10	2.00	(0.06)**	NT_10	1.86	(0.06)**	-0.14*
NT_11	2.33	(0.05)**	NT_11	2.25	(0.05)**	-0.08**
NT_12	1.79	(0.05)**	NT_12	1.62	(0.05)**	-0.17**
NT_13	0.35	(0.19)	NT_13	0.32	(0.40)	-0.03
NT_14	1.98	(0.06)**	NT_14	1.80	(0.07)**	-0.18**
NT_15	1.40	(0.11)**	NT_15	1.33	(0.15)**	-0.07
NT_16	1.68	(0.08)**	NT_16	1.41	(0.10)**	-0.27*
NT_17	1.83	(0.05)**	NT_17	1.76	(0.06)**	-0.07
NT_18	1.69	(0.06)**	NT_18	1.47	(0.08)**	-0.22**
NT_19	2.11	(0.06)**	NT_19	1.98	(0.06)**	-0.13**
NT_20	2.52	(0.05)**	NT_20	2.45	(0.05)**	-0.07*
Observations	24411	Pseudo R2	0.95			

- Output of PPML estimation.

- The coefficient estimate for “NT_{sX}pY” is the BE 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 H0 is “difference between period 1 and 2 is equal to zero”. * indicates rejection of H0 at 5%, ** indicates rejection of H0 at 1%.

- Robust Standard Errors in parenthesis. - * significant at 5%; ** significant at 1%.

- Fixed-Effects for MRTs included: Exporter/Year + Partner/Year.

¹⁰ An anonymous referee pointed out that the inclusion of two non-manufacturing sectors (“Mining and Quarrying” and “Electricity, Gas and Water supply”) could bias the estimation of the BE for the manufacturing sectors. To address this issue, we ran the estimations in Table 2 by ruling out the two non-manufacturing sectors. The BE estimates and ranking remain unchanged, thus confirming that the two non-manufacturing sectors do not introduce any bias but enrich our analysis.

Figure 2. Border Effect, 2-period analysis

Note: Antilog transformation applied to the estimated values.

IV. Geographic Concentration of Production in Europe

The distribution of production is uneven throughout the geographic space. It is usually possible to observe patterns, such as agglomeration of specific activities in some locations. The causes of agglomeration are different, and there are several models to explain such patterns (Baldwin et al. 2005). For the purposes of our analysis, the CP must be measured through a synthetic index (namely, how far each industrial sector is from an even distribution over the geographic space), which will be used to explain the BE in the next section.

As discussed above, Chen (2004) uses the values of the Ellison–Glaeser (EG) Index computed using US data and reported in Ellison and Glaeser (1997) to test the relationship between concentration and the BE.¹¹ In contrast with Chen, we believe that computing the concentration index with data for the European industries is preferable.¹² Ideally, the EG index would be computed using European plant-level data. However, these data are not available to us; therefore, we use the Theil index because it can be computed with the publicly available data on European employment (Eurostat’s Regio database, see the appendix) for every sector for which the BE is estimated.¹³

¹¹ In the group of concentration indices, the EG delivers a higher degree of accuracy because it controls for differences in the sizes of plants and for differences in the sizes of geographic areas. However, this accuracy is at a cost because the EG index requires computation of 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 must be available (Ellison and Glaeser 1997, page 906).

¹² We computed the correlation between the Theil index for the EU and the Ellison–Glaeser index for the US. This is 0.09, a very low value which signals no correlation. Such low value is likely to depend on the aggregation of the sectors in Ellison and Glaeser (1997) to match our sectors (from 459 to 20 sectors); for the scope of our analysis we use simple averaging. Moreover, Ellison and Glaeser use employment figures dated 1987 while we use 1995-2003 figures.

¹³ The Theil index does not control for the distribution of plants’ size as the EG does, thus delivering a different description of the concentration pattern. Several concentration indices can be used to account for agglomeration.

Concentration can be measured in *absolute* or *relative* terms. An industry is concentrated in absolute terms if the bulk of production takes place in only a few locations, whereas 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 coincide if geo-units differ in size. High relative concentration of an industry implies a high degree of specialisation, but this is not necessarily the case with high absolute concentration. Depending on the research interest, one measure or the other will be appropriate. A study of comparative advantage and specialisation would concern relative concentration, but because we are interested in scale economies and trade, the relevant measure is the absolute concentration of production (Haaland et al. 1999).

We compute the *Theil index* for absolute concentration. This index is obtained through the formula of the generalised class of entropy indices where the sensitivity parameter α is set equal to 1 (Brulhart and Traeger 2005).¹⁴ Its formula is as follows:

$$GE(1)_k = \frac{1}{R} \times \sum_{r=1}^R \left[\left(\frac{x_{kr}}{\bar{x}_k} \right) \times \log \left(\frac{x_{kr}}{\bar{x}_k} \right) \right] \quad (3)$$

where $0 \leq GE(1)_k \leq \log R$, $\bar{x}_k = \frac{1}{R} \sum_{r=1}^R x_{kr}$ and x_{kr} is activity x (employment, production, value added, etc.) in region r ($r = 1, \dots, R$) in the industrial sector k ($k = 1, \dots, K$); where each region belongs exclusively to a country c ($c = 1, \dots, C$). One property of the generalised class of entropy indices is that they can be decomposed additively to specify how much measured concentration derives from within or between group diversity:

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

where the groups can be either countries or macro-regions. We apply this decomposition to gain further insights into the concentration patter.¹⁵

We use regional employment to compute the Theil index for the twenty sectors that we study.¹⁶ We consider 191 regions (Nuts-2) belonging to 13 EU countries (countries are the

Good reviews of these indices (with applications to Europe) are Aiginger and Davies (2004), Cutrini (2006). Combes and Overman (2004) list seven properties which concentration indices should satisfy. They acknowledge that no measure meets all those criteria simultaneously and that the choice of one index implies neglecting some criteria.

¹⁴ If $\alpha = 1$, we obtain the Theil Index, whereas if $\alpha = 2$, we get the 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 positive the sensitivity parameter, the more sensitive the index at the bottom of the distribution. The index has a different formula for calculating relative concentration, see Cafiso (2009).

¹⁵ The within component is calculated as the index itself but restricted to observations only within the group.

¹⁶ The use of employment figures may bias concentration values when there are relevant productivity differences across regions and sectors; this is why some authors prefer using value-added figures (Behrens et al. 2005). In our

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.¹⁷ The sectors considered are the same ones used for the estimation of the BE (listed in Table 1); detailed information about the data set are available in the appendix. The values of the concentration index by sector for the 1-period analysis, including the decomposition, are reported in Table 4.

The most concentrated sector 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—and therefore minimum of between—is for “Manufacture of coke, refined petroleum products and nuclear fuel” (Sector 6), while the lowest share of within contribution—maximum of between—is for “Manufacture of wood and of products of wood and cork, ...” (Sector 4).¹⁸ It comes as no surprise that Sector 6 experiences the highest share of within concentration. Indeed, this sector enjoys large scale economies but it is strategic for each country. Production is highly concentrated within countries but not at the European level. However, one can guess that, given the features of Sector 4 and the fact that it is a resource-based industry, it is more concentrated in geo-areas which overcome national boundaries.

analysis, we preferred employment because this was available at the Nuts-2 regional level (necessary for the between/within decomposition of the Theil index) for the sectors, countries and years to match with the trade data. Besides, we reckon that such differences are not particularly relevant in our analysis since we consider a homogenous set of countries such as the EU-15 group.

¹⁷ The employment figures used are an average of the yearly observations within the period.

¹⁸ When the within contribution is higher than the between, concentration mainly depends on an unequal distribution within countries, while the across-countries distribution is relatively more equal. To wit, if a sector was located unevenly among countries but equally spread among the regions of each country, then concentration would be caused exclusively by across-country diversity.

Table 4. Concentration of Production, 1-period analysis

Sector Name	Sector Number	TH_O	TH_W		TH_B	
		Value	Value	Perc.	Value	Perc.
M. of Textiles, textile products, leather and footwear	3	1.09	0.69	63.20%	0.4	36.80%
M. of office machinery and computers	13	1.04	0.7	67.20%	0.34	32.80%
M. of coke, refined petroleum products and nuclear fuel	6	0.99	0.88	89.10%	0.11	10.90%
M. of motor vehicles, trailers and semi-trailers	17	0.84	0.6	71.30%	0.24	28.70%
M. of basic metals	10	0.71	0.6	84.40%	0.11	15.60%
Mining and quarrying	1	0.69	0.58	84.20%	0.11	15.80%
M. of radio, television and communication equipment and apparatus	15	0.69	0.57	82.50%	0.12	17.50%
M. of chemicals and chemical products	7	0.68	0.58	84.90%	0.1	15.10%
M. of other transport equipment	18	0.63	0.53	83.60%	0.1	16.40%
M. of machinery and equipment n.e.c.	12	0.6	0.42	70.80%	0.17	29.20%
M. of medical, precision and optical instruments, watches and clocks	16	0.6	0.43	72.10%	0.17	27.90%
M. of electrical machinery and apparatus n.e.c.	14	0.59	0.42	71.20%	0.17	28.80%
M. of furniture; manufacturing n.e.c	19	0.56	0.41	73.60%	0.15	26.40%
M. of fabricated metal products, except machinery and equipment	11	0.54	0.4	74.10%	0.14	25.90%
M. of rubber and plastic products	8	0.49	0.37	75.20%	0.12	24.80%
M. of other non-metallic mineral products	9	0.49	0.34	68.40%	0.15	31.60%
M. of pulp, paper and paper products; publishing and printing	5	0.46	0.4	86.70%	0.06	13.30%
Electricity, gas and water supply	20	0.46	0.34	74.70%	0.12	25.30%
M. of wood and of products of wood and cork, except furniture; M. of articles of straw and plaiting materials	4	0.43	0.25	59.10%	0.18	40.90%
M. of food products, beverages and tobacco	2	0.31	0.23	73.10%	0.08	26.90%

- "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. - Sectors are listed from the most to the least concentrated. - "M." of stands for "Manufacture."

The index values in Table 4 cover the entire period considered. To illustrate changes in the index over time, we report the index value for Period 1 (1995–1998) and Period 2 (2000–2003) in Table 5; the values are plotted in Figure 3.

At a first glance, it becomes apparent that the within and between components are quite stable over time (not reported in Table 5 but plotted in Figure 3), and the index itself does not change significantly. We implement a bootstrap-based test to check the variation of the index (testing strategy suggested by Biewen 2002).¹⁹ In ten out of twenty sectors, the difference is statistically significant (cells with asterisk in Table 5): concentration decreases in Sectors 1, 4, 7, 8, 11, 12, 14, 16 and 19, but it increases in Sector 3. This outcome signals a non-increasing concentration trend in manufacturing as a whole.²⁰

¹⁹ The bootstrap is based on 5,000 replications. Because the bootstrap distribution is not normal for some sectors, rejection areas are defined through the *Bias Corrected and Accelerated Confidence Interval*.

²⁰ As regards manufacturing, this finding also emerges in Brulhart and Traeger (2005; page 614).

The index values computed in this section are included in the gravity equation estimated in the next section. This will serve to verify the relationship between the BE and the geographic concentration of production.

Table 5. Concentration of Production, 2-period analysis

Sec Num ↓ Period →	Overall Theil Index			Sec Num ↓ Period →	Overall Theil Index		
	p1	p2	$\Delta_{p1/p2}$ - test		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 H0 is "No significant variation". * indicates rejection of H0 at 5%.

Figure 3. Concentration of Production, 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.

V. Concentration of Production and the Border Effect

At this point, we can eventually determine whether the relationship between the concentration of production (CP) and the BE emerges in our enhanced analytical framework. As discussed in Section I, we expect the CP index to explain the BE and their relationship to be inverse. Insight is gained from the scatter-plot in Figure 4. The most concentrated sectors do appear to have a lower BE; the negative coefficient of the trend line makes this clear.

We adopt the same procedure used by Evans (2003) which consists in explaining the BE through CP simultaneously to its estimation. Instead of industry-specific NT dummies, we use two variables: a unique NT dummy which does not differentiate across industrial sectors

and an interaction-term made with the sectoral values of the CP index. The regression is as follows:

$$ex_{ij,kt} = \alpha_{i,t} + \alpha_{j,t} + \beta_1 y_{i,kt} + \beta_2 y_{j,t} + \beta_3 adj_{ij} + \beta_4 dis_{ij} + \beta_5 cfr_{ij,k} + \beta_6 NT + \beta_7 NT \times IC_k + \varepsilon_{ij,kt} \quad (4)$$

The sign and significance of the β_7 coefficient to the interaction term indicates whether sectors with a specific CP display larger or smaller BEs. We present results for both the PPML and OLS estimators and for the 1- and 2-period analyses. OLS results are presented to prove the robustness of the relationship regardless of the estimator used; we prefer the PPML estimates for the reasons discussed in Section III.B. The estimation output is in Table 6.

Figure 4. Relation between the BE and CP

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

	Poisson				OLS			
	1-period analysis		2-period analysis		1-period analysis		2-period analysis	
PROD _{i,k,t}	0.94	(0.01)**	0.94	(0.01)**	0.72	(0.01)**	0.71	(0.01)**
GDP _{j,t}	0.57	(0.05)**	0.57	(0.05)**	0.79	(0.03)**	0.79	(0.03)**
DISTANCE _{ij}	-0.87	(0.03)**	-0.87	(0.03)**	-0.98	(0.02)**	-0.98	(0.02)**
CIF/FOB _{ij,k}	-0.34	(0.11)**	-0.37	(0.11)**	-2.3	(0.07)**	-2.33	(0.07)**
CONTIG _{ij}	0.34	(0.03)**	0.34	(0.03)**	0.38	(0.03)**	0.38	(0.03)**
NT	2.57	(0.05)**			2.38	(0.11)**		
IT	-0.88	(0.05)**			-1.00	(0.16)**		
NT ₁			2.70	(0.06)**			2.65	(0.18)**
NT ₂			2.54	(0.07)**			2.33	(0.17)**
IT ₁			-1.01	(0.08)**			-1.37	(0.27)**
IT ₂			-0.94	(0.08)**			-0.99	(0.27)**
Observations	27725		24411		27725		24411	
R-squared					0.78		0.78	
R-pseudo	0.94		0.94					

- Output of PPML and OLS estimation. - Fixed-Effects for MRTs included: Exporter/Year + Partner/Year.

- 1-period analysis: NT stands for National Trade dummy and IT stands for Interaction Term; 2-period analysis: NT₁ and IT₁ are the National Trade dummy and the Interaction Term for the 1st period, NT₂ and IT₂ are the 2nd period National Trade dummy and Interaction term. - Robust standard errors in parentheses. - * significant at 5%; ** significant at 1%.

On the basis of the estimation output, the relationship between CP and the BE holds as expected: the higher the concentration, the lower the BE (the negative sign of the interaction

term). This finding is robust with respect to all the specifications reported in Table 6.²¹ By using the estimated coefficients for the NT dummy (2.57) and for the interaction term (-0.88) in Table 6, one can predict how much the BE would be, on average, in an industry with a specific level of geo-concentration. For example, for an industry with an index of 1.09 (Sector 3 in Table 4), the BE would be equal to 2.49 [= 2.57 - (0.88 × ln 1.09)], whereas for an industry with an index of 0.31 (Sector 2 in Table 4), the BE would be equal to 3.60 [= 2.57 - (0.88 × ln 0.31)].²²

VI. Concentration of Demand and the Border Effect

In the previous section, we showed that the BE and the concentration of production emerge as related in our enhanced analytical framework (as in the Chen 2004 study). In this section, we study whether the previous relationship emerges because of the HME. As discussed in Section II, the HME predicts that the country with the highest consumption share of an industry good will run a trade surplus of that good because it hosts a more than proportional (with respect to domestic consumption) share of firms. We can therefore picture a causal relationship in which domestic demand is the trigger and concentration of production is the outcome. It makes sense then to check the relationship between the CP and the CD and, if this holds, to use CD as an explicative of the BE in the same manner that CP was used in Section V. If the BE is indeed related to CD, this will support our explanation based on the HME. The analysis is carried out as follows: we first calculate CD by industrial sector, then we check the correlation between CP and CD, and finally, we use CD as an explication of the BE in the same manner as CP in the regressions in Table 6.

Concentration of demand is computed using *Domestic Absorption*. Domestic absorption is an indicator of actual consumption; it is computed as “*Production less Export plus Import*”.²³ We use the Theil Index for absolute concentration (see Eq.(3)), as we did for production in

²¹ We tested the relationship further by using relative concentration; the premise holds, but slightly less conclusively. Results are available upon request.

²² The predicted values are valid under a ceteris paribus condition. Indeed, the estimated BE values (those in Table 3) are different because other conditions differ across sectors (e.g., technical barriers to trade, internationalisation costs).

²³ To compute the CD index as accurately as done for CP in section IV, we need domestic absorption figures at the Nuts-2 level. However, export and import figures are available only at the country level. Then, we have generated Nuts-2 level observations by splitting the national figures through regional weights. We compute the weights using the number of local units at the basis of the employment data set (used to compute concentration of production). This process could artificially boost correlation between CP and CD, unfortunately, no alternative was available. More information about the data generating process is in the appendix.

Section IV. Its value by industrial sector is in Table 7, and the between/within decomposition of the index is also reported.²⁴

Table 7. Concentration of Demand, 1-period analysis

<i>Sector Name</i>	<i>Sector Number</i>	TH_O	TH_W		TH_B	
		value	Value	perc	Value	perc
M. of Textiles, textile products, leather and footwear	3	0.78	0.61	78%	0.17	22%
M. of coke, refined petroleum products and nuclear fuel	6	0.64	0.53	84%	0.1	16%
M. of basic metals	10	0.63	0.49	77%	0.14	23%
M. of office machinery and computers	13	0.55	0.44	80%	0.11	20%
M. of other transport equipment	18	0.53	0.4	74%	0.14	26%
M. of pulp, paper and paper products; publishing and printing	5	0.52	0.46	88%	0.06	12%
M. of electrical machinery and apparatus n.e.c.	14	0.52	0.33	64%	0.19	36%
M. of chemicals and chemical products	7	0.48	0.39	81%	0.09	19%
M. of fabricated metal products, except machinery and equipment	11	0.45	0.33	72%	0.12	28%
M. of radio, television and communication equipment and apparatus	15	0.45	0.37	82%	0.08	18%
M. of medical, precision and optical instruments, watches and clocks	16	0.45	0.32	72%	0.13	28%
M. of motor vehicles, trailers and semi-trailers	17	0.44	0.3	68%	0.14	32%
M. of rubber and plastic products	8	0.43	0.33	77%	0.1	23%
M. of machinery and equipment n.e.c.	12	0.41	0.31	75%	0.1	25%
M. of furniture; manufacturing n.e.c	19	0.39	0.32	83%	0.07	17%
M. of other non-metallic mineral products	9	0.36	0.23	64%	0.13	36%
M. of wood and of products of wood and cork, except furniture; M. of articles of straw and plaiting materials	4	0.31	0.21	67%	0.1	33%
Mining and quarrying	1	0.29	0.22	75%	0.07	25%
Electricity, gas and water supply	20	0.29	0.23	80%	0.06	20%
M. of food products, beverages and tobacco	2	0.27	0.19	72%	0.08	28%

- "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 $TH_O = TH_W + TH_B$. - Sectors are listed from the most to the least concentrated. - "M." stands for "Manufacture"

Sectors 3 and 2 emerge as the most and least concentrated sectors, as they are for production (compare with Table 4). Once again, within concentration plays a major role. As in Haaland et al. (1999), if the relationship between CP and CD holds (and we suppose that it does based on the HME), their correlation should be high. The linear correlation between CP and CD is displayed in Table 8 both for the overall index and the between/within components. The correlation is indeed high but not close to 1, leaving divergence between the distribution of demand and production.²⁵ A histogram of CD and CP by sector is in Figure 5.

²⁴ The results in this section are only for the 1-period analysis. Construction of Domestic Absorption figures for two different periods requires an even deeper data manipulation while the 2-period analysis itself does not provide any additional insight.

²⁵ Proof of the link between CP and CD which goes beyond correlation can be found in Haaland et al. (1999).

Table 8. Correlations between CP and CD

Domestic Absorption ↓	Industrial Production		
	Overall	Between	Within
Overall	0.74		
Between		0.51	
Within			0.74

Figure 5. Concentration of Production vs. Concentration of Demand

Notes: the Theil index plotted is normalized through the formula “Normalized TI = $1 - \exp(-TI)$ ” to allow comparability between the index values of CD and CP.

Eventually, we use CD to explain the BE. Here, the interaction term is for the CD index; the estimation output is in Table 9. Concentration of demand is as explicative of the BE as CP (compare with Table 6); the estimated coefficients are of the same magnitude and robust with respect to the Poisson and OLS estimators. For an industry with an index of 0.78 (Sector 3 in Table 7), the BE would be, on average, 2.73 [= $2.48 - (1.02 \times \ln 0.78)$], whereas for an industry with an index of 0.27 (Sector 2 in Table 7), the BE would be, on average, 3.81 [= $2.48 - (1.02 \times \ln 0.27)$].

The relationship between the BE and the concentration of demand is an important finding. If we exclude any bias in the process of generating the domestic absorption observations at the Nuts-2 level, this result supports our explanation of the BE through the HME, as discussed in Section II.

Table 9. BE explained by Concentration of Demand

	Poisson		OLS	
PROD _{i,kt}	0.95	(0.01)**	0.72	(0.01)**
GDP _{j,t}	0.57	(0.06)**	0.79	(0.03)**
DISTANCE _{ij}	-0.88	(0.03)**	-0.98	(0.02)**
CIF/FOB _{ij,k}	-0.34	(0.11)**	-2.3	(0.07)**
CONTIG _{ij}	0.33	(0.03)**	0.38	(0.03)**
NT	2.48	(0.05)**	2.27	(0.12)**
IT	-1.02	(0.08)**	-1.13	(0.24)**
Observations	27725		27725	
R-pseudo	0.94			
R-squared			0.78	

- Output of PPML and OLS estimation, 1-period analysis
 - NT stands for National Trade dummy and IT stands for Interaction Term.
 - Robust standard errors in parentheses. - * significant at 5%; ** significant at 1%.
 - Fixed-Effects for MRTs included: Exporter/Year + Partner/Year.

VII. Conclusions

In this paper, we have studied the relationship between the border effect and the geographic concentration of production. Our aim was to clarify this relationship and explain why it emerges (Section II). We have provided an explanation based on the home market effect that coherently supports the inverse relationship between the BE and concentration. Our empirical analysis indicates a stable inverse relationship between the BE and concentration (Section V), and that the HME can indeed be an explanation for this relationship (Section VI). We have also discussed alternative estimators of a gravity equation and weighted their pros and cons as tools for estimating the BE (Section III). The robustness of our results casts doubt on the relevance of Silva and Tenreyro's critique to researchers applying a non-theoretical focus.

The test of the relationship between the BE and concentration of production has required an analysis of geographic concentration of the group of countries considered. We have gained useful insights into the pattern of concentration by industrial sectors and how concentration has evolved over time. On the basis of our results, we exclude an increase of concentration of manufacturing in the EU in the period 1995–2003.

We believe that such robust and clear evidence supporting the relationship discussed depends on restricting the analysis to a group of homogenous countries that belong to a single market that also tends to be a political union. It is likely that the inclusion of trade flows from more heterogeneous countries would somehow complicate the relationship between the BE and the concentration of production. This should be the object of further study to better understand the evolution of trade patterns between the western and eastern parts of the world in an era of strong reallocation of production.

Bibliographical References

- Aiginger, K. and Davies, S. (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 van Wincoop, E. (2003), 'Gravity with gravitas: a solution to the border puzzle', *American Economic Review* 93(1), 170–192.
- Baldwin, R., Forslid, R., Martin, P., Ottaviano, G. and Robert-Nicoud, F. (2005), *Economic geography and public policy*, Princeton University Press, Princeton (NJ).
- Baldwin, R. and Taglioni, D. (2006), Gravity for dummies and dummies for gravity equations, Discussion Paper 5850, Centre for Economic Policy Research, London.
- Behrens, K., Ottaviano, G. I., Lamorgese, A. R. and Tabuchi, T. (2005), Testing the home market effect in a multi-country world, Working Paper 2005/55, Centre for Operations Research and Econometrics, Louvain.
- Biewen, M. (2002), 'Bootstrap inference for inequality, mobility and poverty measurement', *Journal of Econometrics* 108(2), 317–342.
- Brulhart, M. and Traeger, R. (2005), 'An account of geographic concentration patterns in Europe', *Regional Science and Urban Economics* 35(6), 597–624.
- Cafiso, G. (2008), The Euro's influence upon trade: Rose Effect versus Border Effect, Working Paper 941, European Central Bank, Frankfurt am Main.
- Cafiso, G. (2009), Sectorial border effects in the European single market: an explanation through industrial concentration, Working Paper 1116, European Central Bank, Frankfurt am Main.
- Cameron, A. and Trivedi, P. (2005), *Microeconometrics: methods and applications*, Cambridge University Press, Cambridge (UK).
- Chen, N. (2004), 'Intra-national versus international trade in the European Union: why do national borders matter?', *Journal of International Economics* 63(1), 93–118.
- Combes, P. and Overman, H. (2004), 'The spatial distribution of economic activities in the European Union', in J. V. Henderson and J. F. Thisse (Eds.), *Handbook of regional and urban economics*. Vol. 4. Amsterdam: Elsevier.
- Cutrini, E. (2006), The Balassa Index Meets the Dissimilarity Theil Index: a Decomposition Methodology for Location Studies, Working Paper 274, Università Politecnica delle Marche, Ancona.
- Drukker, D. (2002), 'Bootstrapping a conditional moments test for normality after Tobit estimation', *Stata Journal* 2(2), 125–139.
- Ellison, G. and Glaeser, E. (1997), 'Geographic concentration in US manufacturing industries: a dartboard approach', *Journal of political economy* 105(5), 889–927.

- Evans, C. (2003), 'The economic significance of national border effects', *The American Economic Review* **93**(4), 1291–1312.
- Goldberger, A. (1981), 'Linear regression after selection', *Journal of Econometrics* **15**(3), 357–366.
- Haaland, J., Kind, H., Knarvik, K. and Torstensson, J. (1999), What determines the economic geography of Europe?, Discussion Paper 2072, Centre for Economic Policy Research, London.
- Head, K. and Mayer, T. (2000), 'Non-Europe: the magnitude and causes of market fragmentation in the EU', *Review of World Economics* **136**(2), 284–314.
- Head, K. and Mayer, T. (2002), Illusory border effects: distance mismeasurement inflates estimates of home bias in trade, Working Paper 2002-01, Centre d' Études Prospectives et d'Informations Internationales, Paris.
- Helpman, E. and Krugman, P. (1985), *Market structure and foreign trade*, MIT press, Cambridge (MA).
- Hummels, D. (2001), Toward a Geography of Trade Costs, Working Paper, Purdue University, West Lafayette (IN).
- Hummels, D. and Lugovskyy, V. (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', *The American Economic Review* **70**(5), 950–959.
- Mayer, T. and Zignago, S. (2006), Notes on cepii's distances measures, Centre d' Études Prospectives et d'Informations Internationales, Paris.
- McCallum, J. (1995), 'National borders matter: Canada-US regional trade patterns', *The American Economic Review* **85**(3), 615–623.
- Nitsch, V. (2000), 'National borders and international trade: evidence from the European Union', *Canadian Journal of Economics* **33**(4), 1091–1105.
- Ottaviano, G. and Thisse, J. F. (2004), Agglomeration and economic geography, in J. V. Henderson and J. F. Thisse (Eds.), *Handbook of regional and urban economics*. Vol. 4. Amsterdam: Elsevier.
- Pagan, A. and Vella, F. (1989), 'Diagnostic tests for models based on individual data: a survey', *Journal of Applied Econometrics* **4**(S1), S29–S59.
- Ramsey, J. (1969), 'Tests for specification errors in classical linear least-squares regression analysis', *Journal of the Royal Statistical Society. Series B (Methodological)* **31**(2), 350–371.
- Ruud, P. (1986), 'Consistent estimation of limited dependent variable models despite misspecification of distribution', *Journal of Econometrics* **32**(1), 157–187.
- Silva, S. and Tenreyro, S. (2006), 'The log of gravity', *Review of Economics and Statistics* **88**(4), 641–658.

Tobin, J. (1958), 'Estimation of relationships for limited dependent variables', *Econometrica: Journal of the Econometric Society* **26**(1), 24–36.

Wooldridge, J. (2002), *Econometric analysis of cross section and panel data*, MIT press, Cambridge (MA).

Appendix - Data Description

TRADE DATA for the estimation of the Border Effect

The analysis comprises twenty sectors of activity 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- the aggregate for “Electricity, Gas and Water Supply” (NACE: E, ISIC: 40-41), 3- 18 subgroups of manufacture as partitions 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 for both a data availability issue and to correctly match the trade data with the employment data available for the concentration index. The countries considered are Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.²⁶

Bilateral exports among the 15 European countries considered (needed for the gravity equation) are extracted from the OECD Stan database (Bilateral Trade Flows section). These data are integrated with the national trade (NT) observations for every country-sector-year combination (the NT observations are for the estimation of the BE) and they amount to sectoral production less total sectoral export (to wit, the amount of sectoral production consumed within the country). Sectoral production and total sectoral export are extracted from the OECD Stan database 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 and production figures in current US dollars are converted in real terms by using the producer price index from the OECD *Economic Outlook* 2007. Weighted distances between and within countries are from the CEPII Distance database (Mayer and Zignago 2006).²⁷

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 weight-to-value (wv_k) which was not possible to reproduce for our sample.²⁸ Instead of the weight-to-value ratios, we use the CIF/FOB ratios to account for different trade costs across industrial goods and pairs.²⁹ We compute the ratios by matching export and import figures by sector. We follow

²⁶ Data for Belgium and Luxembourg are recorded together for the so-called Belgium-Luxembourg Economic Union (BLEU).

²⁷ For more information about the measurement of distance and how it can affect the estimation of BEs, see Head and Mayer (2002).

²⁸ 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.

²⁹ CIF is the abbreviation for “Cost of Insurance and Freight” while FOB means “Free On Board”. Hummels and Lugovskyy (2006) study the accuracy of these ratios as indicators of trade costs. They affirm that “since data on transportation costs are difficult to obtain, in their absence many researchers have turned to indirect measures of

Hummels and Lugovskyy (2006) in ruling out the ratios higher than 2.³⁰ An original interpretation of the CIF/FOB ratios used in this paper is in Cafiso (2009), appendix II .

DATA for the Concentration of Production Index

Concentration of production is computed by using employment figures which have the sectoral and geographic dimension (for the list of sectors see Table 1). These data come from the Eurostat's Regio database (Structural Business Statistics section). 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; the United Kingdom, 37 regions.

This data set is characterized by the high occurrence of missing values. A correct computation of absolute concentration would require a balanced-panel of data (namely, the same number of geo-units for every sector). To get as close as possible to a balanced data set, we fill up the missing values through the following actions:

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. 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.³¹
4. Deletion of those Nuts-2 regions for which the missing value could not be filled-in (see values in brackets above).

transportation costs constructed using matched partner CIF/FOB 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). We therefore decided to use them.

³⁰ It was not possible to compute any CIF/FOB ratio for the national trade observations, then we set the ratio equal to 1 for those observations .

³¹ In the Eurostat's Regio database, 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.

DATA for the Concentration of Demand Index

Concentration of demand is computed using domestic absorption (DA). DA is generated as “Production less Export plus Import”. To make concentration of demand comparable to concentration of production, we needed production, export and import data by Nuts-2 region, ISIC rev. 3 sector and year. However, export and import figures are available only at the national level. We therefore constructed DA figures at the Nuts-2 level by dividing national DA through regional weights. National production, export and import in US dollars are extracted from the OECD Stan database.

We use the amount of local units in each Nuts-2 region (as reported in the Eurostat’s Regio database) to generate weights; no alternatives were available because of the sectoral dimension. As already mentioned, the local units in the Eurostat’s Regio database 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. We acknowledge that this process might boost correlation between CD and CP, however, this was the only way to generate regional DA figures. It is to notice that the correlation between the two is hampered anyway: i) employment does not depend only upon the number of firms but also upon the size of each firm; ii) the distribution of national DA by sector is different from the distribution of national employment by sector.