Document de travail

Spatial Localization in Manufacturing: A Cross-Country Analysis

N° 2009-07
Avril 2009

Stefania Vitali
Chair of Systems Design, ETH Zurich

Mauro Napoletano
OFCE-DRIC

Giorgio Fagiolo
Sant'Anna School of Advanced Studies

Observatoire Français des Conjonctures Économiques
69, Quai d'Orsay 75340 Paris Cedex 07
Tel : 01 44 18 54 00 Fax : 01 45 56 06 15
E-mail: ofce@ofce.sciences-po.fr Web: http://www.ofce.sciences-po.fr

Spatial Localization in Manufacturing: A Cross-Country Analysis

Stefania Vitali* Mauro Napoletano† Giorgio Fagiolo‡

March 27, 2009

Abstract

This paper employs a homogenous firms database to investigate industry localization in European countries. More specifically, we compare, across industries and countries, the predictions of two of the most popular localization indices, i.e., the Ellison and Glaeser index (Ellison and Glaeser, 1997) and the Duranton and Overman index (Duranton and Overman, 2005). We find that, independently from the index used, localization is a pervasive phenomenon in all countries studied, but the degree of localization is very uneven across industries in each country. Furthermore, we find that the two indices significantly diverge in predicting the intensity of the forces generating localization within each industry. Finally, we perform a cross-sectoral analysis of localized industries. We show that, in all countries, localized sectors are mainly "traditional" sectors (like jewelery, wine, and textiles) and sectors where scale economies are important. However, once one controls for countries' industrial structures science-based sectors turn out to be the most localized ones.

Keywords: Industry Localization; Manufacturing Industries; Localization Indices; Spatial Concentration; Spatial correlation; Cross-country studies

JEL classification: R12, R3

*Chair of Systems Design, ETH Zurich. Mail Address: Chair of Systems Design, Kreuzplatz 5 ETH-Zentrum, KPL F27.2, CH-8032 Zurich, Switzerland. Tel: +41 44 632 41 02. Fax: +41 44 632 18 80. E-mail: `svitali@ethz.ch`.

†Observatoire Français des Conjonctures Economiques, Departement for Research on Innovation and Competition, and Sant'Anna School of Advanced Studies. Mail Address: Observatoire Français des Conjonctures Economiques, Department for Research on Innovation and Competition, 250 rue Albert Einstein, 06560 Valbonne, France. Tel: +33 4 93 95 43 75. Fax: +33 4 93 65 37 98. E-mail: `mauro.napoletano@ofce.sciences-po.fr`

‡Sant'Anna School of Advanced Studies. Mail Address: Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, I-56127 Pisa, Italy. Tel: +39 050 883343. Fax: +39 050 883344.
Email: `giorgio.fagiolo@sssup.it`

1 Introduction

This paper investigates the empirical location patterns of manufacturing industries in six European countries: Belgium, France, Germany, Italy, Spain, and the UK. Drawing on a comprehensive source covering data on European manufacturing firms, we simultaneously perform both a cross-country and a cross-sector analysis of industry spatial-localization patterns employing two of the most popular localization measures: the "Ellison and Glaeser Index"(E&G index henceforth, see Ellison and Glaeser, 1997) and the "Duranton and Overman Index" (D&O index henceforth, see Duranton and Overman, 2005). Our main goal is to provide a common empirical framework where, thanks to the harmonized source of data employed, one might be able to compare predictions of different indices across different countries in a homogeneous way. Indeed, as we argue in more detail below, existing empirical studies on industry localization have almost entirely focused on studying how different sectors were localized in a given country, according to a single index. Results are therefore hardly comparable, due to the inherent heterogeneity in data collection and definitions of variables (e.g., firm size).

The analysis of firms' location has attracted the attention of economists for a very long time (see e.g. Marshall, 1920). More recently, a relevant body of theoretical research in the "New Economic Geography" literature (see e.g. Krugman, 1991; Fujita et al., 1999) has been aimed at explaining what might be considered the basic stylized fact of economic geography, i.e. that firms look more clustered in space than any theory of comparative advantage would predict.

Beside these theoretical contributions, a good deal of empirical research has investigated localization in manufacturing industries (see among others, Ellison and Glaeser, 1997; Maurel and Sédillot, 1999; Barrios et al., 2005; Lafourcade and Mion, 2007; Duranton and Overman, 2005). All these works, mainly focusing on single countries, confirm the expectation that firms are generally clustered in space. However, they also find huge variability in the degree of localization across industrial sectors. A common characteristic of these studies is the use of some measure of the degree of firms' clustering in space ("localization indices"). Despite very similar methodological approaches, however, the literature has so far been quite heterogeneous in terms of the variety of the measures employed (see e.g. Combes and Overman, 2004b, for a survey).

For example, Krugman (1991) and Haaland et al. (1999) have proposed indices based on the "Location Quotient" or "Balassa-Index". These indices measure localization of an industry in "excess" with respect to what would be predicted by the overall presence of economic activities into specific areas. By measuring excess localization, these indices control for the overall tendency of manufacturing industries to agglomerate in specific areas due to exogenous factors, e.g. population. Nevertheless, they do not provide any null hypothesis against which evaluate the "absolute" degree of clustering of an industry.

Moreover, they do not control for the level of industry concentration, which is important to allow cross-industry comparisons and to avoid spurious measurements. Indeed, the concentration of industry activity into a specific area could simply reflect the presence in the area of a single plant accounting for most of the activity of the industry, rather than the genuine presence of several industry plants in the same area.[1] The contributions in Ellison and Glaeser (1997); Dumais et al. (2002); Maurel and Sédillot (1999) overcome all the foregoing problems, by providing localization indices which simultaneously: i) control for the overall concentration of manufacturing; ii) control for industry concentration, iii) provide a null hypothesis against which verifying the presence of localization. In particular, the index in Ellison and Glaeser (1997) tests the presence of localization driven by the combination of sector-specific spillovers and natural advantage of specific areas, against the null hypothesis of localization driven by random firm-specific choices. These indices represent key advances in the measuring of localization. Nevertheless, by construction, they require an ex-ante partitioning of the geographical space (e.g. a country) into smaller units (e.g. regions, departments). In other words, points on a map (corresponding to the location of business units) are transformed into units in "boxes" (cf. Duranton and Overman, 2005; Combes and Overman, 2004b). The division of the space into sub-units has the advantage of making the computational problems involved in the measurement of localization easier. However, it also introduces possible biases in the analysis. First, comparisons among countries are difficult, as the areas of spatial sub-units may significantly vary across different countries. Second, and relatedly, comparisons become difficult also *within* countries across different spatial scales (e.g. departments vs. regions). Finally, clusters of firms located at the borders of neighboring regions and/or spanning over the area covered by a single region are treated in the same way as clusters in two very distant regions. More precisely, as pointed out in Arbia (2001) and Lafourcade and Mion (2007), indices requiring the division of the space into smaller units are only able to capture "spatial concentration" of industrial activity into some areas. They are not able to measure instead "true agglomeration", that is the degree of spatial correlation between firms' location choices. Localization indices that tackle the foregoing problems are those proposed in Moran (1950) and Duranton and Overman (2005). In particular, the latter index does not require *any* ex-ante partition of the geographical space into subunits. This is because the index relies on the empirical distribution of distances across firms, computed by locating firms on the basis of their postal codes (more on that in Section 3).

This paper is an attempt at improving upon the foregoing literature along two dimensions. *First*, we simultaneously perform an investigation of industry localization in *several* EU countries by exploiting a firms database *homogeneous* across European countries.

Using harmonized data allow us to detect cross-country localization patterns that are

[1] See however Bottazzi et al. (2007) for some skeptical remarks on the need for controlling for industry lumpiness.

not influenced by the different ways to measure firms' size and location in the different countries under investigation. Differently from the literature, however, we use *firms* rather than *plants* data in our investigation. This could induce an upward bias in the estimation of the number of localized industries, as all the production units belonging to the same firm are all concentrated in the same area (the headquarters). To check for this possible bias, we perform a more detailed analysis of localization, using different approaches to detect and locate firm's plants in space. The results of these robustness analyses show that localization is not overestimated by using firms rather than plants data (more on that in Section 5).

Second, we depart from the standard practice of analyzing localization employing a single index, by performing a comparative study of different localization indices (E&G and D&O indices) to estimate localization patterns. In particular, we compare the two indices with respect to their predictions about: i) the number of sectors localized; ii) the intensity of industry localization forces; iii) the types of localized sectors. The choice about the two indices to compare is deliberate. Indeed, as briefly explained above, the E&G index and the D&O index lie at two opposite extremes in the treatment of geographical space. It turns out that one might expect significant divergences in the predictions about industry localization characteristics.

Previous attempts in the same direction can be found in Barrios et al. (2005) and Lafourcade and Mion (2007). However, differently from Barrios et al. (2005, 2008), we consider a larger number of (size-heterogenous) countries. Furthermore, to account for spatial features of the data we employ the D&O index rather than the Moran index (Moran, 1950). This choice has been made because, as discussed at more length in Arbia (2001), the Moran index cannot entirely capture the observed variability in spatial permutations. For example, the Moran index measures the extent to which an industry is spread across L neighboring regions. As a result, it is not able to distinguish between the case of an industry which is spatially distributed from West to East and the case wherein the same industry is distributed from West to South. Similarly it cannot account for the actual distribution of firms within the contiguous regions where clusters are observed (we will go back to these points in Section 3).

In line with previous studies, we find that localization is a pervasive phenomenon in all countries studied albeit, in each country, the degree of localization is very unevenly distributed across sectors. Furthermore, we show that countries display significant variability in the share of localized sectors. Finally, our exercises suggest that in all six countries localized industries are mainly "traditional" ones (jewelery, wine, textiles, etc.), as well as those where scale economies are important. This outcome mainly reflects the historical evolution of countries' industrial structures. Once we control for this factor, science based industries become those where localization is more pervasive. We also detect significant differences across the localization measures we employ. On the one hand, the two indices make the same predictions about: i) the share of countries' localized sectors; ii) the unevenness of localization forces across sectors; iii) the type of localized industries. On the other hand, the two indices markedly differ in predicting the intensity of the forces underlying localization within each industry, both across and within countries.

These results have theoretical and empirical implications. First, the presence of cross-country differences (given the index) in the share of localized sectors hints to the presence of factors (e.g., natural advantages, institutions) that make localization in general more pervasive in some EU countries than in others. Furthermore, the fact that the same type of industries are localized across the different countries suggests the prevalence of sector-specific localization drivers vs. cross-sectoral ones. Finally, our results provide evidence about the empirical importance of the spatial scale in measuring localization phenomena. More in detail, the spatial scale seems not to be important for detecting the *presence* of localization forces into an industry. Conversely, using different spatial scales can make a crucial difference in the *type* of localization phenomena one is observing. More precisely, cross-index differences in localization intensity for a significant fraction of industries indicate that localization forces operate quite differently across industries. In some sectors, localization forces lead to clusters of firms characterized by strong spatial correlation in firms' location choices, possibly spanning over the territory defined by any possible spatial subunit. In others, spatial correlation is much weaker and localization reflect only the spatial concentration of business units is some ex-ante defined areas.

The rest of the paper is organized as follows. Section 2 presents the database used in the analysis. Section 3 describes the localization indices employed. We begin with the Ellison and Glaeser index and then we move to the Duranton and Overman index. Section 4 is devoted to the presentation and discussion of our results on the empirical analysis of localization in EU countries. Furthermore, Section 5 checks the robustness of the results obtained in the previous section, with particular emphasis on the possible biases due to using firms rather than plants as the object of analysis. Finally, Section 6 concludes.

2 Data

The empirical analysis below is based on three different data sources. The data on firms are from the Orbis dataset of the Bureau Van Dijk, 2006 release (cf. www.bvdep.com/en/ORBIS). From this extensive dataset we have extracted information about location (i.e., postal codes), employment, and industrial classification of firms in six European countries (Belgium, France, Germany, Italy, Spain and the United Kingdom) for the period 2004-2006. These countries were selected partly out of choice and partly out of necessity. On the one hand, we wanted to focus on those countries that have already been the object of single-country studies in the relevant literature.[2] On the other hand, we were constrained by the availability of reliable and detailed firm-specific data that could be matched with the information of firms location (see also below). Indeed, only for a relatively small number of European countries we were able to collect information on sufficiently large samples disaggregated at the industry level, in such a way to efficiently compute firm geographical coordinates and distances.

All information is available at the firm level and is derived from companies' annual reports. Firm-level data other than localization is available from 2004 to the last year available in the database, which unfortunately differs among countries. More precisely, data are available until 2005 for Belgium, France, Italy, and United Kingdom; until 2004 for Spain; and until 2006 for Germany. However, information on firm localization only refers to the last year available. In principle, one would like to keep as many countries as possible in the analysis and, at the same time, be sure that localization data are synchronized with other firm-specific variables. In order to meet these two conditions, we have thus decided to employ data, in each given country, only for the last year available in the database. This of course prevents us from performing a proper cross-section analysis, but we do not expect this to be a source of important bias to the analysis. Indeed, given the relatively short time span covered by the database, we only a small fraction of all firms are going to change their locations. Similarly, sectors are not very likely to dramatically change their industrial structure.

We limit our analysis to manufacturing industries as defined by the NACE classification (NACE Rev.1 section D). More specifically, following Duranton and Overman (2005), we restricted our analysis to sectors with more than 10 firms. This allows us to exclude sectors where localization is the result of location choices by few firms and therefore to focus on clustering phenomena where localization forces attract a significant bunch of firms.

To identify spatial sub-units, we apply the NUTS (Nomenclature of Territorial Units for Statistics) classification (cf. www.ec.europa.eu/eurostat/ramon/nuts). NUTS is a hierarchical classification at five levels (three regional and two local), extensively used for comparative statistics among European countries.[3] For our purposes we use NUTS-3

[2]In addition to US-focused research (Ellison and Glaeser, 1997; Rosenthal and Strange, 2001; Holmes and Stevens, 2002a; Kim, 1995), existing contributions have been studying industry localization patterns in UK (Devereux et al., 2004; Duranton and Overman, 2005); Belgium (Bertinelli and Decrop, 2005); France (Maurel and Sédillot, 1999); Italy (Lafourcade and Mion, 2007); Germany (Brenner, 2006); Ireland and Portugal (Barrios et al., 2005). See Combes and Overman (2004a) for a review. Notice that Brenner (2006) and Holmes and Stevens (2002a) do not employ neither the E&G nor the D&O index.

[3]The NUTS partition of the EU territory has been used by EU since 1988 as a single uniform breakdown of territorial units for the production of regional statistics for the European Union. The classification does not build only on administrative boundaries, but follows maximum and minimum population thresholds for the size of the region and analytical criteria. In particular, the former criterion takes account of geographical structure and socio-economic characteristics of the territory, so that related area under different administrative layers could be embraced in the same NUTS-3 level.

regions, which are quite homogeneous and comparable between countries. We then assign firms to each sub-unit on the basis of their postal codes. The data needed to map NUTS-3 postal codes come from the European Commission Database ("Regional Indicator and Geographical Information Unit").

Since one of the two indices we employ in our analysis (D&O, cf. Section 3.2), requires the identification of the longitude-latitude coordinates of firms in space, we also employ data from the "TeleAtlas Multinet Europe" database (cf. www.teleatlas.com). More precisely, this database provides the spatial coordinates of the contour of the areas corresponding to postal codes in our sample. Each firm is then assigned coordinates coinciding with the centroid of the postal-code area.

Table 1 shows some descriptive statistics. The number of firms under analysis is highly variable among countries considered, and in some cases (UK) it is quite low, mainly because of a lack of data to match firm postal codes with Geographical Information System (GIS) coordinates and NUTS-3 regions.[4] Note that censoring the sample to sectors with more than 10 active firms does not have a significant effect on the dimension of the sample. Indeed, the fraction of sectors covered is always more than 50% of total sectors available in each country. Finally, in our sample average firm size considerably varies across countries. Average firm size is rather large in the UK and Italy, and relatively small in Spain and France.

3 Localization Indices

This section describes the properties of the localization indices that we employ in our investigation. As we argued in the introduction, the literature has so far proposed several measures to capture firms' spatial clustering.[5]. Here we shall focus on two indices that have gained a lot of attention in the recent years. These are the index proposed in Ellison and Glaeser (1997) (E&G Index) and that introduced in Duranton and Overman (2005) (D&O Index). Both indices present solutions to problems affecting older measures of localization. However, they markedly differ in their approach to geographical space and in the type of localization phenomena they are able to capture.

[4] Cf. www.gis.com/. Indeed, some difficulties typically arise in tracking the spatial evolution of postal codes. This can make geographic geo-referencing far from straightforward. For instance, boundaries of postal-code areas move continuously due to new addresses, sometimes they can change name, and/or new ones enter the stage. These events are quite common in UK, for more details, cf. www.statistics.gov.uk/geography/.

[5] A full account of the properties of the different localization indices is beyond the scope of this paper. Combes and Overman (2004b) provide a detailed description of some of the most popular indices, together with a discussion of their properties.

3.1 The Ellison and Glaeser Index

The E&G index proposed in Ellison and Glaeser (1997) is based on a probabilistic model of location choice, where each business unit (plant or firm) sequentially chooses its location. More precisely, the j-th business unit chooses its location v_j in such a way to maximize its profits π_{ji} from locating in area i:

$$\log \pi_{ji} = \log \bar{\pi}_i + g_i(v_1, ..., v_{j-1}) + \varepsilon_{ji}, \tag{1}$$

where $\bar{\pi}_i$ is a random variable reflecting cross-sectorally homogeneous profits arising from "natural advantages" attached to area i (e.g., presence of a river, or favorable weather conditions). The term $g_i(v_1, ..., v_{j-1})$ captures the effect of sector-specific spillovers created by business units that have previously chosen that location. Finally, ε_{ji} is an additional random component modeling factors that are idiosyncratic to the j-th business unit.

On the basis of this model of location choice, Ellison and Glaeser (1997) derive an index γ_n, measuring the propensity of firms in industry n to co-locate in space:

$$\gamma_n = \frac{G_n - (1 - \sum_i {x_i}^2)\, H_n}{(1 - H_n)\,(1 - \sum_i {x_i}^2)}, \tag{2}$$

where G_n is the "raw-concentration index":

$$G_n = \sum_i (s_i - x_i)^2. \tag{3}$$

In (2) and (3) s_i is the share of industry's employment in area i, x_i is the share of aggregate manufacturing employment in area i. The term H_n is the Herfindahl index of industry concentration $H_n = \sum_j {z_j}^2$, with z being the share of employment of the jth firm in the industry.

The E&G index has many interesting properties, as compared to other indices proposed in the literature. First, similarly to the measure proposed in Krugman (1991), the E&G index controls for the overall tendency of manufacturing to localize in space (e.g., spatial concentration due to difference in population across areas), as captured by the term $1 - \sum_i {x_i}^2$. However —differently from earlier statistics— the E&G index also measures localization in excess by what predicted by industry concentration. Indeed, the Herfindahl index directly enters in (2) to re-scale the raw index G_n. Finally, the value of γ_n is related to the theoretical model of location choice underlying Eq. (1) by the following relation:

$$\gamma_n = \gamma_{na} + \gamma_s - \gamma_{na}\gamma_s, \tag{4}$$

where γ_{na} and γ_s parametrize, respectively, the importance of natural advantages and spillovers in driving location choices of the business units. The above relation implies two fundamental properties of the E&G index. First, the value of the index can be directly interpreted as reflecting the (non-linear) combination of localization forces due to natural advantage and spillovers. Second, it provides a null hypothesis against which evaluating the degree of localization of an industry. Indeed, a value of the index equal to zero implies that the effect of natural advantage and spillovers on location choices is null. This corresponds to the case of "random location": observed localization is in this case entirely due to the effect of the random idiosyncratic term ε_{ji}. This in turn implies that industries characterized by a positive E&G value display "excess" localization, as compared to what would be predicted by the overall localization of manufacturing and by industry localization. The observed localization is thus driven by the combined effect of natural advantages and firm spillovers. Conversely, industries with excess spatial dispersion will exhibit a negative E&G value, whereas a value of the index equal to zero indicates no localization. The latter situation corresponds to the benchmark scenario where the observed spatial distribution is solely the result of random-location choices by firms in the industry.

One of the major drawbacks of the E&G Index is the lack of a statistical procedure to significantly measure the degree of excess localization (or dispersion) of an industry. To partially solve such a problem, Ellison and Glaeser (1997) proposed some threshold values to interpret and classify positive values of γ_n. According to their criterion, an industry is not very localized when γ_n is below 0.02. Moreover, it is very localized if $\gamma_n > 0.05$. These thresholds were chosen by the authors via an heuristic procedure based on their application on US data and are somewhat arbitrary.[6] Other contributions using the E&G index have instead relied on more rigorous criteria to evaluate the statistical significance of γ_n's values. In particular, a procedure based on a standard "2-sigma rule" has been proposed (see e.g. Rosenthal and Strange, 2001; Devereux et al., 2004; Barrios et al., 2005). Since under the null hypothesis of random location $\gamma_n = 0$ and $E(G_n) = (1 - \sum_i {x_i}^2)H_n$, an industry will be significantly localized (dispersed) whenever the difference between the empirical value of the raw concentration index G_n and its expected value $(1 - \sum_i {x_i}^2)H_n$ is twice larger (smaller) than the standard deviation σ_G of the raw concentration index (cf. Ellison and Glaeser, 1997):

$$\sigma_G = \sqrt{2\left\{H^2[\sum_i x_i^2 - 2\sum_i x_i^3 + (\sum_i x_i^2)^2] - \sum_j z_j^4[\sum_i x_i^2 - 4\sum_i x_i^3 + 3(\sum_i x_i^2)^2]\right\}}. \quad (5)$$

Note that country and industry specific terms enter the expression of both the expected value and the standard deviation of the raw index G_n. This makes the "2-sigma rule" criterion more suitable to account for country and industry specificities in the analysis. In what follows, we will use such a criterion to evaluate the statistical significance of

[6] Ellison and Glaeser define the values above by ranking sectors according to the average and median of γ_n. They find that 25% of US industries are highly localized while 50% of them shows weak localization.

localization (or dispersion) of an industry.

3.2 The Duranton and Overman Index

The E&G index is based on an exogenous division of the geographical space into subunits. Space partitions have the advantage of alleviating the computational problems involved in the measurement of industry localization. Indeed, measuring the propensity of firms to co-locate in space boils down to calculating the concentration of industrial activity into $m > 1$ areas (e.g. regions, departments). Unfortunately, the division of the space into subunits has also several disadvantages. First, it is not a-priori clear the optimal spatial breakdown at which firm clustering should be measured. One could, e.g., decide to compute the index considering counties, regions or different NUTS layers. This undermines comparison, both cross-country and across different disaggregation levels (see e.g. Rosenthal and Strange, 2001; Devereux et al., 2004, for a discussion of this point). Second, as argued at more length in Arbia (2001), the very computation of cumulative shares of economic activity concentrated in spatial subunits implies disregarding the spatial nature of the data. Indeed, indices based on cumulative shares (like the E&G index) are generally invariant to any spatial permutation of the subunits under investigation. However, having the bulk of industrial economic activity split among two distant regions is totally different from splitting it in two neighboring areas. Moreover, by focusing on total activity in one or more regions, one can only investigate *spatial concentration*, that is the uneven distribution of industry activities across regions. One cannot instead evaluate how industry activities are spatially distributed in the region (or across two neighboring areas). This means disregarding "true agglomeration", i.e. the degree of spatial correlation in firms' location decisions (cf. Arbia, 2001; Lafourcade and Mion, 2007; Duranton and Overman, 2005).

Two indices that account for the spatial features of industrial data are those proposed by Moran (1950) and by Duranton and Overman (2005). As briefly mentioned above, the Moran index still requires an ex-ante partition of the space. However, it is based on a weighting matrix W, whose generic element represents the weight of location l for location i. Weights represent contiguity relationships: $w_{il} > 0$ if and only if (i, l) are contiguous regions and zero otherwise (with $w_{ii} = 0$, all i). As discussed in Arbia (2001), this index does not solve all the space-related problems described above. For example, the Moran index is invariant to different spatial permutations involving the same number of contiguous regions. In fact, this index —by construction— can only capture the degree of firm clustering across neighboring regions and is not able to distinguish between different neighboring regions (e.g., from West to East, rather than from West to South). Similarly it cannot account for the actual distribution of firms within the contiguous regions where clusters are observed.

In light of these considerations, in this paper we have preferred to focus on the D&O index as our only alternative to the E&G index that accounts for the spatial features of the data. Indeed, D&O index does not require any ex-ante division of the space into subunits and therefore it seems to be better equipped to deal with the characteristics of the spatial distribution of firms into industries.

To compute the D&O index, one needs to build Euclidean distances between pairs economic units (plants or firms) in each industry, employing from their actual position in geographical space. Geographic positioning is identified by firms' postal codes. If the number of firms is M, the number of unique bilateral distances is $M(M-1)/2$. We can then estimate the density of distances through the (Gaussian) kernel function:

$$K\left(d\right)=\frac{1}{M\left(M-1\right)b}\sum_{h=1}^{M-1}\sum_{j=h+1}^{M}f\left(\frac{d-d_{hj}}{b}\right), \qquad (6)$$

where d_{hj} is the distance between firms h and j, b is the bandwidth and f is the (Gaussian) kernel function. All distances are computed in kilometres.[7]

Obviously, studying the distribution of kernel densities alone does not give us information whether a sector is localized or not. To solve this problem, the D&O index allows for a rigorous statistical test of industry localization. The test involves the comparison of the empirical density to artificially generated distributions based on random location of firms in space. Note that this procedure controls also for industry concentration. Indeed, if the industry were only characterized by an uneven distribution of market shares, then its spatial density would not be statistically different from the one generated by randomly re-location of firms in space.

In what follows, we shall then bootstrap 1000 samples generated by randomly allocating the position of firms in a given sector, over the whole population of locations occupied by firms in manufacturing. We then build a *local* confidence interval by ranking the samples in ascending order for each target distance (d) and taking the 5th and the 95th percentile for the lower 5% and the upper 95% confidence interval.[8] In this way, 95% of the distribution shall fall inside the confidence bands at each target distance. This means that the measures of localization (α) and dispersion (δ) are given by:

$$\alpha_I(d)=K_I(d)-\overline{K}_I \qquad (7)$$

$$\delta_I(d)=\widetilde{K}_I-K_I(d), \qquad (8)$$

[7]In the estimation of the densities, one typically restricts the support to the positive domain, replacing negative densities with zeros and re-scaling all values to get the densities estimates sum up to one. However, since the density function in Equation (6) has domain on the whole real line, it could return positive density estimates at negative distances, even if the natural domain of our study is a positive interval.

[8]Following Klier and McMillen (2008), we have selected 40 target distances.

where $\overline{K}_I$ is the upper confidence band in industry I, while $\widetilde{K}_I$ is the lower confidence band. Note that the indices α and δ allow us to draw implications only at the "local" level, i.e. at a given distance. By employing them one cannot say anything in general about the degree of localization (dispersion) of an industry as a whole. To cope with this limitation a *global* confidence interval is needed. The difference with local bands is that the kernel density is *jointly* evaluated at several target distances. The global confidence interval is built in such a way that no more than the 95% of the random distribution lies outside the interval between the upper and the lower global confidence bands. For an industry n the index of (global) localization (A) and (global) dispersion (Δ) are thus given by:

$$A_n(d) = K_n(d) - \overline{\overline{K}}_n \tag{9}$$

$$\Delta_n(d) = \begin{cases} \widetilde{\widetilde{K}}_n - K_n(d) & \text{if } \sum dA_n(d) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

This definition implies that an industry will be considered localized if its kernel density lies at least once above the interval delimited by global confidence bands. Conversely, it will be considered dispersed if its kernel density is always below the lower confidence band.

The measurement of localization without a previous partition of geographical space is a crucial value added for the D&O index, but it can also represent an important drawback. Indeed, data to compute geographical coordinates are often not easily accessible and barely available at very detailed level. In addition, compared to the E&G index, computational problems become huge. One indeed moves from calculating cumulative shares of industry (and manufacturing) activities across $L > 1$ regions to tracking the entire distribution of $M(M-1)/2$ distances computed on the basis of $M >> L$ postal codes areas.

4 Results

In this section we present the results of our analysis on the spatial distribution of firms in manufacturing industries. We begin by investigating the extent of localization in the countries analyzed. In other words, we study whether the number of industries where firms co-locate is significant in the country under analysis. In addition, we study whether the fraction of localized industries displays cross-country variation. Furthermore, we investigate how strong is localization. Indeed, the value returned by each localization index captures the strength of localization forces into an industry (see Section 3). It is then worthwhile analyzing whether the intensity of those forces is heterogeneous across countries and sectors. Finally, we carry out a detailed analysis of the sectoral composition of localized industries, to check which kinds of industries are more often localized.

In all the investigations below, we compare the results produced by the two localization indices employed (E&G and D&O). As we discussed in Section 3, these indices markedly differ in their approach to the measurement of localization. Thus, as suggested in Arbia (2001), the identification of invariances and diversities in localization patterns captured by the two indices may help unveiling different characteristics of industry localization in the countries under study.

4.1 How many industries are localized?

Let us start by assessing how many industries are localized in each country considered. Table 2 shows the fraction of sectors localized and dispersed using respectively the E&G index and D&O index. In general, localization emerges as a widespread phenomenon in all countries. The share of industries for which the value of the E&G index is strictly positive turns to be very high in all countries considered (see Table 2, column 1). This result is consistent with previous studies in the literature (see e.g. Devereux et al., 2004; Maurel and Sédillot, 1999; Lafourcade and Mion, 2007). However, the fraction of localized industries reduces considerably when one applies the 2-sigma rule to evaluate the statistical significance of localization. For instance, the fraction is reduced by more than half in Belgium (from 0.7 to 0.32), whereas in Spain and Italy 60% of total sectors still display localization after the application of the stricter rule. The fraction of sectors displaying excess dispersion (i.e., such that $\gamma_n < 0$) is very low in all countries considered. Finally none of the sectors studied was significantly dispersed according to the 2-sigma rule criterion. Turning to the D&O index, we find that in all countries but Belgium the share of localized sectors is around 50%. Overall, the figures are lower than those obtained by considering sectors with a positive value of the E&G index.[9] However, the share of localized sectors is very similar with the one obtained applying the 2-sigma rule to the E&G index (cf. Table 2, columns 2 and 4).

Figure 1 summarizes the patterns of localization emerging in each country from the application of the E&G index and the D&O index. It is tempting to roughly classify countries in three groups according to their pervasiveness of localization. More precisely, we can identify a group of "high-localization" countries, including Spain and Italy, wherein the share of localized sectors is between 50% and 60% of the total; a group of "intermediate-localization" countries, including France, Germany and (to a less extent) UK, wherein such a fraction is between 40% and 50%; and, finally, a "low-localization" country — Belgium— where only about one third of sectors is localized. Note, that the ranking of countries is invariant to the type of index used. Indeed, as the Figure 1 shows, there exists a clear monotonic non-decreasing relation between the share of sectors significantly localized according to the E&G index and to the D&O index.

[9] A similar result for UK was already emphasized by Duranton and Overman (2005).

4.2 How much are industries localized?

As discussed in Section 3, the magnitude of the E&G and D&O indices can provide some information not only on whether a sector is localized or dispersed, but also on the *intensity* of localization, and therefore of the forces underlying its emergence.[10] For instance, the E&G index is the result of the combination of forces arising from cross-firms spillovers and geographical advantages of specific areas (see Eq. 4). Likewise, the value of the D&O index captures the extent to which the spatial distribution of firms in the sector deviates from the one generated under the hypothesis of random firm location choices. In light of these remarks, this section studies in detail the cross-sector distributional properties of the two indices in each country.

For each index, Table 3 reports the first four sample moments of within-country distributions of localized industries in the countries considered. Results clearly indicate that, in all countries, cross-sectors distributions are very right-skewed and display excess kurtosis. This suggests that, within each country, localization forces operate very unevenly across manufacturing sectors. In particular, all countries are characterized by the co-existence of a vast majority of sectors displaying very low levels of localization, together with few "outliers" where forces underlying the emergence of localization are extremely strong. This is confirmed by kernel density estimates for E&G and D&O sectoral distributions, cf. Figures 2 and 3.

The foregoing results are in line with previous findings in the literature (e.g. Ellison and Glaeser, 1997; Maurel and Sédillot, 1999; Duranton and Overman, 2005), which however make use of heterogeneous databases and statistical procedures. What is more, they seem to be robust to the index employed. Indeed, both indices deliver distributions of localized sectors having similar statistical properties (cf. Table 3, Figures 2 and 3). Nonetheless, the two indices produce different cross-country rankings with respect to *average* localization intensity. For instance, both indices predict that average intensity is the lowest in Germany. However, the E&G index indicates that localization forces are on average higher in UK, Belgium and Spain, whereas France and Italy (together with Belgium) are the countries where localization is more intense according to the D&O Index.

To further investigate the cross-index differences in average localization intensity detected above, we perform a Wilcoxon rank-sum (one-sided, non-paired) test for each pair of countries (c_1, c_2).[11] The null hypothesis is that average localization intensity is the same across the selected pair of countries —i.e., that the two distributions of localization intensity are the same— whereas the alternative hypothesis is that the distribution of localization intensity for country c_1 is significantly shifted to the right of the distribution

[10] *Strictu sensu*, this is true for statistically-significant values of the E&G index, whereas it applies by construction to the D&O index.

[11] Being non-parametric, the Wilcoxon test appears a good candidate for the analysis at stake. Standard t-tests indeed rely on the assumption of normality of the distribution which does not seem appropriate in our case (see Table 3).

of country c_2. Wilcoxon test statistics, together with their corresponding (exact) p-values are reported in Table 4.

Most of the cross-index differences detected by sample moments are confirmed. Consider first the E&G index (in what follows, we use a 5% threshold for convenience). Average intensity is significantly larger in UK with respect to all other countries considered. Both Belgium and Spain present an average intensity larger than France and Germany. Moreover, average intensity is significantly lower in Germany with respect to all countries. Notice that for the cases Belgium vs. Italy and Belgium vs. Spain one cannot reject the null hypothesis irrespective of the two one-sided alternatives. These results are by far altered when one applies the D&O index to data. Indeed, the null of equal average is not rejected for all pairs of countries except for the whole Germany's profile and for the Belgium-France comparison.

The evidence just described points to the presence of important cross-index differences. Additional evidence on intensity patterns comes from the analysis of cross-country Spearman correlation matrices in industry rankings produced by the two indices employed, cf. Table 5. Predictions about rank correlations differ markedly across the two indices employed. On one hand, the E&G index assigns considerable positive correspondence in ranks to all countries considered, with an average level of correlation equal to 0.59. On the other hand, rank correlation values decrease sharply using the D&O index and in many cases they are not even statistically significant. The pairs of countries displaying the highest rank correspondence are also different. The E&G index predicts that the pairs of countries displaying highest rank correlation are Belgium and France, France and Spain, and UK and Germany, with values of the coefficients respectively equal to 0.73, 0.71 and 0.74. In contrast, the D&O index suggests that such pairs are Belgium and Germany, and Italy and Spain.

By and large, the above results indicate that localization indices significantly diverge in predicting the intensity of localization forces both *within industries* and *across countries*. Interestingly, the same type of divergence is observed also *within countries*. Indeed, as Table 6 shows, the Spearman rank correlation coefficient between the E&G index and the D&O index —among localized sectors— are always significant (with the exception of the UK). Nevertheless, correlation coefficients appear in general quite small. In particular, the correlation between the two indices appears much weaker vis-à-vis their predictions on within-country shares of localized sectors (cf. Figure 1 and Table 2).

To sum up, the foregoing findings provide some empirical support in favor of the claim that the analysis of industry localization is sensitive to the type of index used (Arbia, 2001). More in detail, the low rank correlation observed within countries indicates the presence of sectors that are spatially concentrated at the NUTS-3 level, but wherein firm location choices are not spatially correlated. Interestingly, the same mismatches are found across countries, as indicated by cross-index differences found in average intensity and in ranking correlations. The weak cross-country correlation found applying the D&O index shows that the *same* industry can be very agglomerated in one country (i.e., location choices may display high levels of spatial correlation) but much less in others.

4.3 Which industries are localized?

The results presented in the previous sections show that, independently of the index used: i) localization is a pervasive phenomenon in most EU countries; ii) in each country localization forces are very uneven across manufacturing sectors. Moreover, they indicate that different indices make quite different predictions about the intensity of the forces underlying the emergence of industrial localization. In this section we investigate the composition of the group of localized industries in the countries considered.

We begin by looking at how much groups of localized industries are similar across countries. Despite similar shares (cf. Table 2), countries could indeed be very different in terms of the composition of the group of localized sectors (e.g., due to different industrial structures). We therefore begin by computing the share of sectors in common between pairs of countries, see Table 7. The fraction of localized sectors across countries is in general large for each index, especially E&G. On average, countries share respectively 68% and 64% of the localized sectors according to E&G index and D&O index. In some cases, ratios go up to 75%. In particular, Spain shares the major number of clustered sectors with other countries.

But which are the most localized sectors in each country? As Tables 8 – 13 show, in each country the most localized sectors include to a great extent "traditional" industries, like jewelery, wine, and textiles. Moreover, the tables reveal the presence of a relevant cross-country variability in localized industries. Nonetheless, it is possible to detect the presence of a "core" of localized sectors that is invariant across countries. More precisely, considering the E&G index, 13 industries appear in the list of localized sectors in all countries considered, while this number reduces to 8 if we consider the D&O index. Among these sectors, 4 are in common between the two indices and belong to the publishing and printing sector group (NACE 2211, 2213, 2215 and 2222).[12]

In order to better interpret these results, we have performed a cross-sectoral investigation of *all* localized industries, employing a taxonomy that classifies industries in macro groups composed of sectors with relatively homogeneous characteristics. More specifically, we employ here Pavitt's taxonomy (Pavitt, 1984), one of the first (and most widely used)

[12] The other industries being part of the "core" for the E&G index are: "Manufacture (Mfg) of meat" (1513), "Mfg of prepared feeds for farm animals" (1571), "Finishing of textiles" (1730), "Mfg of other outwear" (1822), "Mfg of metal structure" (2811), "Forging, pressing, stamping and roll forming of metal" (2840), "General mechanical engineering" (2852), "Building and repairing of ships" (3511) and "Mfg of chairs and seats" (3611). The other "core" sectors for D&O index are: "Mfg of perfumes and toilet preparations" (2452), "Mfg of taps and valves" (2913), "Mfg of machinery for textile, apparel and leather production" (2954) and "Mfg of jewelery" (3622).

classification frameworks proposed in the industrial organization literature. Pavitt's taxonomy classifies industries mostly on the basis of their technological characteristics (e.g., internal vs. external sources of the innovation process; product/process innovation; the degree of appropriability of innovations, etc.), but takes also into account other industry dimensions like type entry barriers, average size of firms in the sectors, etc. Using such classification criteria, one is able to identify at least four macro-categories of industries: *science based*, *specialized suppliers*, *scale intensive* and *traditional*.[13] The most salient features of each Pavitt category are summarized in Table 14.

We begin by computing the share of localized sectors in each Pavitt category, for the countries considered and for each localization index (cf. Table 15). Note, first, that the two indices do not exhibit great differences in the share of localized sectors across Pavitt categories. Furthermore both E&G and D&O suggest that localized industries mainly belong to the groups of "traditional" and scale intensive sectors. Taken together, these two groups cover more than 70% of localized industries in all countries. Second, science based sectors (typically characterized by intense internal and external R&D activity) feature the smallest fraction of localized industries.

All this suggests that localization is more pervasive across "traditional" industries and across industries where scale economies are important. As argued at more length in Bottazzi et al. (2005), these two groups of sectors are characterized by two distinct localization phenomena. On the one hand, firms in traditional sectors are often spatially organized in *horizontally diversified clusters*, with the coexistence of many producers of similar but differentiated products. On the other hand, localization patterns in scale intensive industries takes typically the form of *hierarchical clusters*, involving an oligopolistic core together with subcontracting networks.

Furthermore, the results in Table 15 seem to indicate that localization phenomena driven by *knowledge complementarities* (e.g., in science based industries) is a quite rare phenomenon. This might contrast with previous results in the literature (see in particular Audretsch and Feldman, 1996), that find a positive correlation between clustering and degree of innovativeness in science-based industries. This seemingly puzzling pattern could however reflect the numerical prevalence of traditional and scale intensive industries in the countries under examination (see Table 16). Thus, the apparent weakness of localization within science-based industries could simply be the sheer outcome of the historical evolution of different industrial structures (Ottaviano, 1999). To control for such a factor, we have re-scaled the share of localized sectors in each Pavitt category by the share of sectors in each category. The results of this exercise are reported in Table 17. Controlling for industrial structures implies a significant increase in the share of science-based localized industries. Indeed, according to both localization indices, the share is above one in all

[13] The group of "traditional" industries corresponds to the group of "supplier dominated" industries in the original Pavitt taxonomy.

countries (but Belgium and UK) and is almost always the largest one, as compared to that of other Pavitt categories. This indicates that in science-based sectors there seems to be a more pervasive localization effect than what predicted by the share of sectors in total manufacturing.

Notice also that, unlike what happens with raw shares, cross-index differences emerge if one considers re-scaled ratios. More precisely, according to the E&G index, localization seems in general weaker in all Pavitt's categories other than science based, whereas this does not occur using the D&O index. In the latter case, scale intensive industries and specialized suppliers do exhibit a marked "excess localization" in most countries.

Such a cross-index heterogeneity provides additional clues on the characteristics of industry localization. On one hand, the excess localization predicted by both localization indices in science-based sectors hints at firms that, in those sectors, tend to co-locate at very small distances. In particular, distances look smaller than those defined by NUTS-3 level regions. On the other hand, the D&O index predicts a higher share of scale-intensive and specialized-suppliers industries vs. the E&G index. This suggests that firms in these sectors are more often localized at the borders of neighboring regions and/or at distances that go beyond those defined by the NUTS-3 breakdown level.

5 Robustness Analysis

The analysis presented in the previous sections indicate that industry localization is a pervasive phenomenon across European countries. Localized sectors are mainly "traditional" or scale-intensive industries or —taking into account country-heterogeneity in industrial structures— science based industries. Albeit such results are robust to the type of localization index employed, E&G and D&O make different predictions about the strength of the forces underlying localization. In particular, both cross-country and within-country differences emerge with respect to the ranking of localized sectors generated by each index.

As mentioned, all our results rely on a country-homogenous database where observational units are firms rather plants. At a first glance, the use of firm-level data (as opposed to plant-based data) may induce an upward bias in the measurement of localization, as different production units belonging to the same (large, multi-plant) firm would wrongly show up in the data as they were located in the area of the their headquarter. Such an upward bias is likely to be affected by the share of medium and large firm in the data, due to the positive correlation between firm size and the number of firms' plants typically observed in empirical studies (cf., e.g., Coad, 2008). A more careful scrutiny of the indices and their characteristics, however, suggests that the direction of the bias is rather unclear. On the one hand, the concentration of production units in the headquarter's area could indeed induce an upward-biased estimation of localization. On the other hand, moving from plants to firms implies also an increase in industrial concentration for the sector under study. The latter point can better grasped by noticing that the Herfindahl index entering the E&G formula (2) involves the sum of squares of employment shares of business units (firms or plants). Moving from plants to firms implies *coeteris paribus* an increase of industry concentration, just because all double products for plants belonging to the same firm are now counted. This may therefore counterbalance the upward bias discussed above.

In order to check the robustness of our results to such biases, we run two different sets of analyses. First, we study localization levels predicted by the two indices by conditioning to firm size. The assumption is that multi-plant firms are mainly firms of medium/large size. Thus, knowing how localization indices perform, within each industry, across firms belonging to different size classes may convey useful information on the direction of the purported firm-plant bias. Second, we perform a simulation analysis of localization patterns on samples built by artificially disaggregating medium-large firms in several production units, and by locating those units in space according to different theoretical scenarios.

To perform a size-conditioned localization study, we have initially partitioned our sample into two classes of firms (small vs. big) employing as a size threshold the median of the industry-pooled size distribution in each country. Then, for each country and sector, we have computed E&G and D&O indices separately for each size class, in such a way to identify the size-conditioned shares of localized and dispersed industries. Table 18 reports the results of this exercise.[14] A comparison of size-conditioned figures with those obtained in the unconditioned analysis (cf. Table 2) seems to indicate that the effects of the firm-plant bias are quite negligible in our sample. Indeed, splitting the sample into small and big firms causes, on the one hand, a modest reduction in the share of localized industries in both size classes, and in all countries considered. Such a decrease is actually weaker in the small-size class, which under our assumption, should mainly reflect location choices made by single-plant firms. On the other hand, the decrease in the fraction of localized industries in countries having the largest median employment in our sample (e.g., UK) is comparable to those in countries having smaller median firm sizes (e.g., Spain and France).

To further verify that the firm-plant bias is not that relevant in our sample, we also performed a simulation analysis of localization patterns by artificially disaggregating the empirically-observed firms of our sample (in each industry and country) into fixed-size plants and reallocating such plants in the geographical space. To do so, we first assumed that the expected number of plant for a given firm increases with its size. Following Lafourcade and Mion (2007), we assumed that, in all industries and countries, the plant size is $d = 20$ employees. This means that the number of plants p of each firm of size s is

[14] Notice that such results are, to a large extent, robust to different criteria for choosing the size threshold (e.g., exogenously determined using the threshold employed by Lafourcade and Mion, 2007).

simply equal to $p = s/20$ (rounded to the closest integer). Second, we set up alternative theoretical hypotheses about the way these artificially-generated plants can be spatially reallocated in the space. More specifically, plants in a given country and industry can be assigned to existing locations according to one of the following scenarios:

- *Uniform distribution*: Distribute plants randomly among all existing locations with probability $1/L$;
- *Small-firm based distribution*: Distribute plants randomly among all existing locations so that the probability p_l that a plant is in a location l is i.i.d. and equal to $p_l = \frac{s_{il}}{\sum_l s_{il}}$. i.e. the ratio between the number of employees of small firms in industry i in location l (s_{il}) and the number of employees of small firms in industry i ($\sum_l s_{il}$). In this way, we try to reproduce the geographical distribution of small firms that, under our assumption, is in our sample the best proxy to the actual (but to us unobservable) spatial distribution of plants.

Under either scenarios, we performed several independent replications (R=1000) of the above procedure, where in each replication we have disaggregated and reallocated all firms in the database. We thus computed the share of sectors that in each country turned to be localized in *all R* simulations. Given the tremendous computational requirements that the D&O index places on this kind of simulation analysis (especially in countries with a huge number of firm observations), we present results for the E&G index only (see Table 19).[15] Our simulation analysis seems indeed to confirm, by and large, the findings obtained using a size-conditioned analysis. Indeed, in both allocation scenarios, the share of sectors that are localized in all simulation runs according to the 2-sigma rule is quite high in almost all countries (except for Belgium and UK), and larger than the fraction of localized sectors computed on actual data (cf. Table 2). All that hints to a downward (rather than upward) sample bias associated to considering firms instead of plants as observational units.

6 Conclusions

In this paper, we have empirically investigated industry-localization patterns in EU countries. Unlike the majority of existing works in this field, we have employed a firm-based dataset that is homogeneous across countries, and computed two different localization indices —i.e., the Ellison and Glaeser index (see Ellison and Glaeser, 1997) and the Duranton and Overman index (see Duranton and Overman, 2005). This has allowed us to make statistically-sound comparisons both across countries and across localization indices, at aggregate and sector-disaggregated levels. In line with previous studies, we have found

[15]Preliminary results on the D&O index, however, do not seem to contradict E&G simulation findings.

that, independently from the index used, localization is a pervasive phenomenon in all countries studied. However, countries significantly differ in the share of localized sectors. In addition, in all countries the values of localization indices display a relevant sectoral variability. Furthermore, a cross-sectoral analysis of localized industries has shown that, in all countries and for both indices, "traditional" and scale intensive sectors are those displaying the highest tendency to localize. These results partly reflect the composition of country industrial structures: once one controls for such a factor, science-based sectors turn out to be the most localized. Finally, we have detected significant cross-index differences with respect to predictions on the intensity of the forces underlying localization. This may reflect some heterogeneity in the way geographical firm clustering looks like in space. Some clusters might indeed map "true agglomeration", i.e. strong spatial correlation among firms' location choices, while other clusters might only reflect the "spatial concentration" of firms in some ex-ante, exogenously determined, areas.

The present work could be extended in several ways. First, one could perform a more detailed investigation of the influence of firm size in affecting firm location choice. In addition, one could bridge this attempt to a full-fledged analysis of the links between industry dynamics (e.g., firm entry, exit, and growth) and the generation and evolution of localization phenomena, cf. Ottaviano (1999), Holmes and Stevens (2002b), Klepper (2002) and Lafourcade and Mion (2007) for important contributions in this direction.

Second, our results suggest that groups of sectors differ in their tendency to localize. However, the indices employed in this work do not allow one to distinguish sector-specific drivers from drivers that are independent on technological factors and are specific to geographical areas (e.g., natural advantages). One could therefore employ a different index (e.g., the one proposed in Bottazzi et al., 2007) that explicitly disentangle sector-specific factors from location-sectoral ones, and then rigorously study the influence of sectoral characteristics in firm location choices.

Third, in the foregoing exercises we have disregarded, on a first approximation, the characteristics of the areas where firms tend to localize. In fact, one could use our firm-based dataset, to study in more detail the tendency of EU firms to locate in very urbanized areas, e.g. due to the presence of services (e.g. financial, consulting, auditing) supporting firm activity (see Henderson and Ono, 2007; Davis and Henderson, 2008; Strauss-Kahn and Vives, 2009).

Finally, one could develop more rigorous testing procedures for the Ellison and Glaeser index, in order to overcome the limits of the 2-sigma rule, which is implicitly based on a non-tested assumption of normality of E&G statistic distribution. For instance, following a procedure in line with Duranton and Overman (2005), one could generate artificial samples based on a model of random location of firms, and then compute a theoretical distribution of the Ellison and Glaeser index. This way, statistical-significance tests for empirical values of the index would not depend either on the characteristics of the country studied or on the characteristics of the data source used.

Acknowledgments

We are indebted to Frank Schweitzer, Kerstin Press and Stefano Battiston for their invaluable advice and support. Thanks also to Giulio Bottazzi, Thomas Brenner, Giovanni Dosi, Davide Fiaschi, Koen Frenken, and Gianmarco Ottaviano for their comments and suggestions. Moreover, we thank participants to the DRUID Winter 2008 Conference (Aalborg, Denmark); the 12th International Schumpeter Society Conference (Rio de Janeiro, Brazil); the 2008 ERSA conference (Liverpool, UK); the GREDEG-CNRS Workshop on "Agglomeration and International Trade" (Sophia-Antipolis, France); and the DIME Economic Geography Workshop in (Utrecht, Netherlands), where earlier versions of this paper were presented. All usual disclaimers apply.

References

Arbia, G. (2001), "The role of spatial effects in the empirical analysis of regional concentration", *Journal of Geographical Systems*, 3: 271–281.

Audretsch, D. B. and M. P. Feldman (June 1996), "R&D Spillovers and the Geography of Innovation and Production", *American Economic Review*, 86: 630–40.

Barrios, S., L. Bertinelli, E. Strobl and A.-C. Teixeira (2005), "The dynamics of agglomeration: evidence from Ireland and Portugal", *Journal of Urban Economics*, 57: 170–188.

Bertinelli, L. and J. Decrop (2005), "Geographical Agglomeration: Ellison and Glaeser's Index Applied to the Case of Belgium", *Regional Studies*, 39: 567–583.

Bottazzi, G., G. Dosi and G. Fagiolo (2005), "On Sectoral Specificities in the Geography of Corporate Location", in S. Breschi and F. Malerba, (eds.), *Clusters, Networks, and Innovation*, Oxford University Press, Oxford.

Bottazzi, G., G. Dosi, G. Fagiolo and A. Secchi (2007), "Modeling industrial evolution in geographical space", *Journal of Economic Geography*, 7: 651–672.

Brenner, T. (2006), "Identification of Local Industrial Clusters in Germany", *Regional Studies*, 40: 991–1004.

Coad, A. (2008), "Firm growth and scaling of growth rate variance in multiplant firms", *Economics Bulletin*, 12: 1–15.

Combes, P. and H. Overman (2004a), "The Spatial Distribution of Economic Activities in the European Union", *Handbook of Regional and Urban Economics*, 4: 2845–2909.

Combes, P.-P. and H. G. Overman (2004b), "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*, Elsevier-North Holland, Amsterdam.

Davis, J. and J. Henderson (2008), "The agglomeration of headquarters", *Regional Science and Urban Economics*, 38: 445–460.

Devereux, M. P., R. Griffith and H. Simpson (2004), “The geographic distribution of productivity activity in the UK”, *Regional Science and Urban Economics*, 34: 533–564.

Dumais, G., G. Ellison and E. Glaeser (2002), “Geographic Concentration as a Dynamic Process”, *Review of Economics and Statistics*, 84: 193–204.

Duranton, G. and H. G. Overman (2005), “Testing for Localization Using Micro-Geographic Data”, *Review of Economic Studies*, 72: 1077–1106.

Ellison, G. and E. L. Glaeser (1997), “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach”, *Journal of Political Economy*, 105: 889–927.

Fujita, M., P. Krugman and A. Venables (1999), *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge.

Haaland, J. I., H. J. Kind and K. Ulltveit-Moe (Feb 1999), “What Determines the Economic Geography of Europe?”, CEPR Discussion Papers 2072, C.E.P.R. Discussion Papers.

Henderson, J. and Y. Ono (2007), “Where do manufacturing firms locate their headquarters?”, *Journal of Urban Economics*, 63: 431–450.

Holmes, T. and J. Stevens (2002a), “Geographic concentration and establishment scale”, *Review of Economics and Statistics*, 84: 682–690.

Holmes, T. and J. Stevens (2002b), “Geographic Concentration and Establishment Scale”, *Review of Economics and Statistics*, 84: 682–690.

Kim, S. (1995), “Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987”, *The Quarterly Journal of Economics*, : 881–908.

Klepper, S. (2002), “The capabilities of new firms and the evolution of the US automobile industry”, *Industrial and Corporate Change*, 11: 645–666.

Klier, T. and D. McMillen (2008), “Evolving Agglomeration in the US Auto Ssupplier Industry”, *Journal of Regional Science*, 48: 245–267.

Krugman, P. (1991), *Geography and Trade*, MIT Press, Cambridge.

Lafourcade, M. and G. Mion (2007), “Concentration, agglomeration and the size of plants”, *Regional Science and Urban Economics*, 37: 46–68.

Marshall, A. (1920), *Principles of Economics*, 8th edition, MacMillan, London.

Maurel, F. and B. Sédillot (1999), “A measure of the geographic concentration in french manufacturing industries”, *Regional Science and Urban Economics*, 29: 575–604.

Moran, P. A. P. (1950), “Notes on continous stochastic phenomena”, *Biometrika*, 37: 17–23.

Ottaviano, G. (1999), “Integration, geography and the burden of history”, *Regional Science and Urban Economics*, 29: 245–256.

Pavitt, K. (1984), "Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory", *Research Policy*, 13: 343–373.

Rosenthal, S. S. and W. C. Strange (2001), "The Determinants of Agglomeration", *Journal of Urban Economics*, 50: 191–229.

Strauss-Kahn, V. and X. Vives (2009), "Why and where do headquarters move?", *Regional Science and Urban Economics*, 39: 168–186.

Country	No. of Firms	Last Year Available	No. of NACE Sectors[1]	Avg. No. of Employees[2]	No. of Firms [3]	Avg. Firm Size[2]
Belgium	13032	2005	160 (227)	1603	65	20
France	50396	2005	201 (237)	4051	227	16
Germany	62588	2006	231 (239)	7445	264	27
Italy	26940	2005	206 (234)	5472	120	42
Spain	67809	2004	282 (309)	3711	228	15
UK	6056	2005	115 (204)	7675	157	33

Table 1: Descriptive statistics. Notes: [1] Number of NACE sectors with at least 10 active firms. Total number of sectors in each country in brackets. [2] In NACE sectors. Size: firm employees. [3] Average number of firms per sector.

	E&G index			D&O index	
Country	LOC[1]	LOC[2]	DISP	LOC	DISP
Belgium	0.7000	0.3187	0.3000	0.3312	0.1000
France	0.7910	0.5274	0.2090	0.5025	0.1741
Germany	0.7749	0.4675	0.2251	0.4935	0.1039
Italy	0.8107	0.5922	0.1893	0.4757	0.1990
Spain	0.8191	0.6028	0.1809	0.5355	0.1560
UK	0.7043	0.4000	0.2957	0.4435	0.1913

Table 2: Share of sectors localized (LOC) and dispersed (DISP) in each country. Notes: [1] Share of sectors with strictly positive E&G index. [2] Share of sectors significantly localized according to the 2-sigma rule.

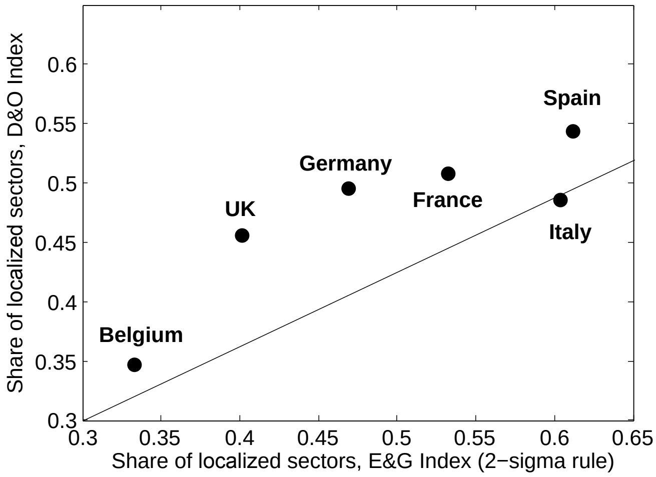

Figure 1: Share of localized sectors according to the E&G index (x-axis) vs. share of localized sectors according to the D&O index (y-axis).

E&G Index	Mean	Std Dev	Skewness	Kurtosis
Belgium	0.0853	0.0964	1.8069	5.8337
France	0.0536	0.0938	4.3353	25.7457
Germany	0.0269	0.0374	3.0638	16.1371
Italy	0.0708	0.1000	3.8002	22.7231
Spain	0.0862	0.1030	3.4528	20.5689
UK	0.1012	0.0964	1.7770	5.2609
D&O Index	**Mean**	**Std Dev**	**Skewness**	**Kurtosis**
Belgium	0.0367	0.0849	3.5573	16.3644
France	0.0140	0.0265	3.4644	17.8371
Germany	0.0065	0.0173	6.3512	52.3214
Italy	0.0150	0.0283	3.9659	22.2184
Spain	0.0114	0.0173	2.7720	11.6615
UK	0.0133	0.0200	2.8677	12.8419

Table 3: Moments of empirical distributions of localized industries.

E&G	Belgium		France		Germany		Italy		Spain		UK	
	W	P-value	W	P-value	W	P-value	W	P-value	W	P-value	W	P-value
Belgium	–	–	3614	0.0003	4231	0.0000	3409	0.1606	4063	0.7515	929	0.9610
France	1792	0.9997	–	–	7111	0.0011	4881.5	0.9993	5520	1.0000	1212	1.0000
Germany	1277	1.0000	4337	0.9989	–	–	3461	1.0000	3663	1.0000	680	1.0000
Italy	2813	0.8394	8050.5	0.0007	9715	0.0000	–	–	8634	0.9926	1848	0.9997
Spain	4607	0.2485	12500	0.0000	14697	0.0000	12106	0.0074	–	–	3240	0.9626
UK	1417	0.0390	3664	0.0000	4288	0.0000	3764	0.0003	4580	0.0374	–	–

D&O	Belgium		France		Germany		Italy		Spain		UK	
	W	P-value	W	P-value	W	P-value	W	P-value	W	P-value	W	P-value
Belgium	–	–	3195	0.0243	4224.5	0.0000	2874.5	0.1397	4364	0.1634	1497	0.1720
France	2158	0.9757	–	–	6924	0.0052	4459.5	0.8859	6781	0.9318	2349	0.8116
Germany	1817.5	1.0000	4590	0.9948	–	–	3785.5	1.0000	5509	1.0000	2008	0.9992
Italy	2319.5	0.8603	5438.5	0.1141	7386.5	0.0000	–	–	7277.5	0.5866	2505.5	0.4896
Spain	3639	0.8366	8470	0.0682	11705	0.0000	7520.5	0.4134	–	–	3876.5	0.4713
UK	1206	0.8280	2802	0.1884	3806	0.0008	2492.5	0.5104	3824.5	0.5287	–	–

Table 4: One sided Wilcoxon rank sum test statistics (W) and exact p-values in brackets. Null hypothesis: Row and column distributions are the same. Alternative hypothesis: Row distribution shifted to the right of column distribution. Note: Due to the properties of the test statistics W, the sum of p-values of two entries symmetric to the main diagonal is one.

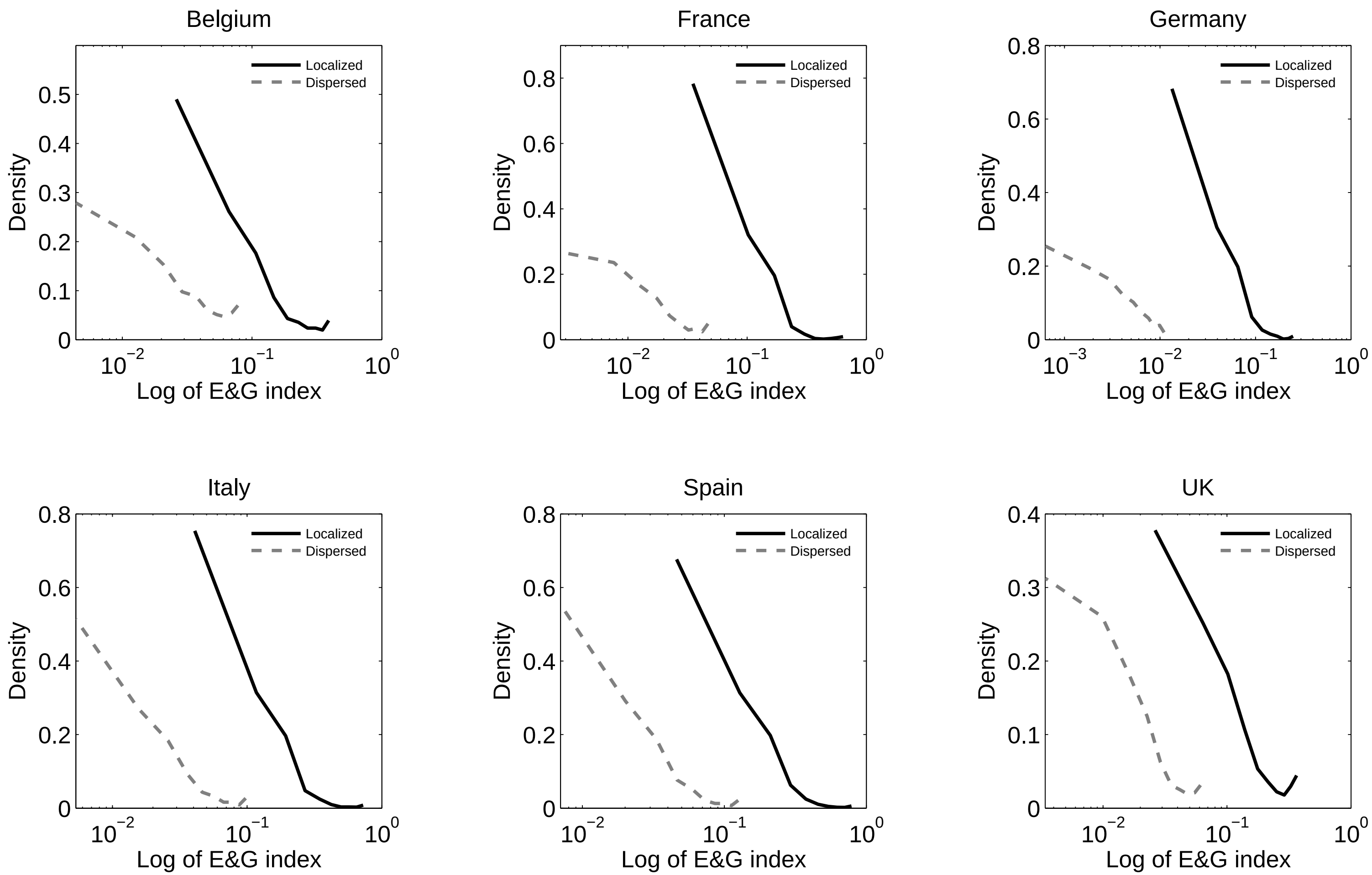

Figure 2: Kernel density estimates of the distribution of E&G index. Localized vs. dispersed sectors. Notes: Log scale on the x-axis. A sector is localized (resp., dispersed) if the value of the E&G index is greater (resp., smaller) than zero.

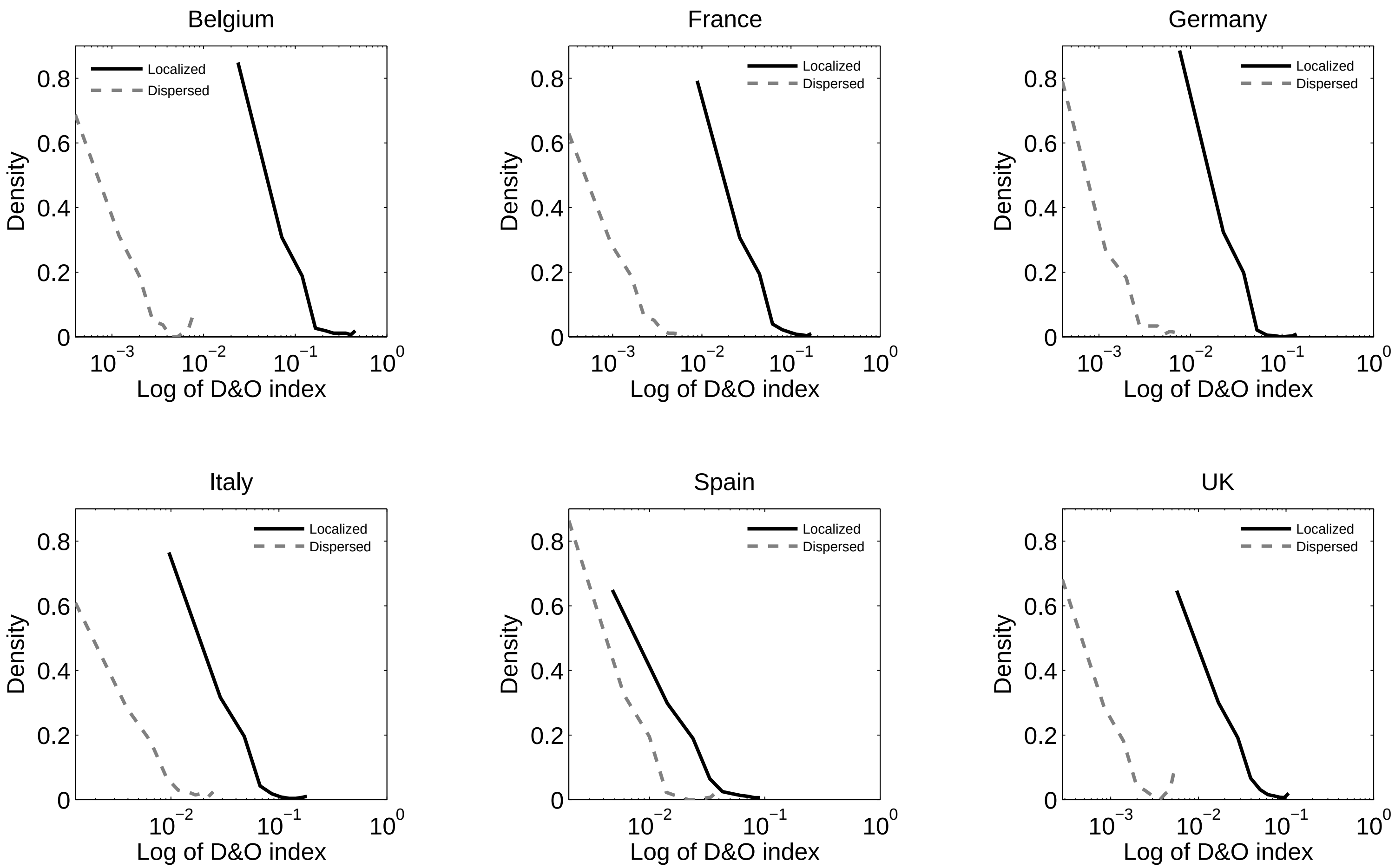

Figure 3: Kernel density estimates of the distribution of D&O index. Localized vs. dispersed sectors. Note: Log scale on the x-axis.

G&E	France	Germany	Italy	Spain	UK
Belgium	0.7347	0.5620	0.5269	0.6660	0.6772
	(0.0000)	(0.0008)	(0.0007)	(0.0000)	(0.0014)
France		0.6639	0.5771	0.7085	0.6056
		(0.0000)	(0.0000)	(0.0000)	(0.0008)
Germany			0.5939	0.4754	0.6618
			(0.0000)	(0.0000)	(0.0002)
Italy				0.5985	0.2054
				(0.0000)	(0.2851)
Spain					0.5844
					(0.0001)

D&O	France	Germany	Italy	Spain	UK
Belgium	0.3713	0.5190	0.3262	0.4352	0.3735
	(0.0364)	(0.0010)	(0.0733)	(0.0040)	(0.1541)
France		0.2703	0.4193	0.2536	0.2966
		(0.0282)	(0.0006)	(0.0215)	(0.1115)
Germany			-0.0021	0.1847	0.4149
			(0.9870)	(0.0988)	(0.0226)
Italy				0.5081	0.4646
				(0.0000)	(0.0168)
Spain					0.1810
					(0.2909)

Table 5: Cross-country Spearman rank correlation matrices for localized sectors: E&G vs. D&O indices. Note: Localized sectors in E&G results are computed according to the 2-sigma rule. P-values in brackets.

	Localized sectors*	
Country	Correlation	P-value
Belgium	0.5561	(0.0021)
France	0.5118	(0.0000)
Germany	0.3204	(0.0051)
Italy	0.5725	(0.0000)
Spain	0.4735	(0.0000)
UK	0.1560	(0.4280)

Table 6: Spearman rank correlation between E&G index and D&O index in localized sectors. Note: Localized sectors in E&G results are computed according to the 2-sigma rule. P-values in brackets.

G&E	France	Germany	Italy	Spain	UK
Belgium	0.6863	0.6275	0.7451	0.8235	0.4130
France		0.6226	0.6887	0.7547	0.5870
Germany			0.6019	0.7778	0.5870
Italy				0.7869	0.6304
Spain					0.8696

D&O	France	Germany	Italy	Spain	UK
Belgium	0.6038	0.6981	0.5849	0.7925	0.3137
France		0.6535	0.6429	0.8119	0.5882
Germany			0.6633	0.7168	0.5882
Italy				0.7959	0.5098
Spain					0.7059

Table 7: Share of localized sectors in common between pairs of countries. E&G index vs. D&O index. Note: Localized sectors in E&G results are computed according to the 2-sigma rule. Shares are computed dividing the number of localized industries in common between each pair of countries by the minimum of the number of localized sectors between the two countries.

E&G Index		D&O Index	
NACE	*Sector Name*	*NACE*	*Sector Name*
1724	Silk-type weaving	1724	Silk-type weaving
3622	Mfg of jewelery	1714	Prep. and spinning of flax-type fibres
2626	Mfg of refractory ceramic products	1721	Cotton-type weaving
2954	Mfg of machinery for textile	1751	Mfg of carpets and rugs
1772	Mfg of knitted & crocheted pullovers	1725	Other textile weaving
3511	Building and repairing of ships	3622	Mfg of jewelery
1751	Mfg of carpets and rugs	2954	Mfg of machinery for textile
2913	Mfg of taps and valves	1772	Mfg of knitted & crocheted pullovers
2462	Mfg of glues and gelatines	2213	Publishing of journals and periodicals
1725	Other textile weaving	2211	Publishing of books

Table 8: Ten most localized NACE 4-digit manufacturing sectors, Belgium.

E&G Index		D&O Index	
NACE	*Sector Name*	*NACE*	*Sector Name*
2861	Mfg of cutlery	1715	Throwing and preparation of silk
3350	Mfg of watches and clocks	1724	Silk-type weaving
2211	Publishing of books	2861	Mfg of cutlery
2214	Publishing of sound recordings	2411	Mfg of industrial gases
1715	Throwing and preparation of silk	2214	Publishing of sound recordings
2320	Mfg of petroleum products	1725	Other textile weaving
2213	Publishing of journals and periodicals	2211	Publishing of books
2461	Mfg of explosives	1713	Prep. and spinning of fibres
1593	Mfg of wines	2213	Publishing of journals and periodicals
3661	Mfg of imitation jewelery	2231	Reproduction of sound recording

Table 9: Ten most localized NACE 4-digit manufacturing sectors, France.

	E&G Index		D&O Index
NACE	*Sector Name*	*NACE*	*Sector Name*
2861	Mfg of cutlery	2861	Mfg of cutlery
3613	Mfg of other kitchen furniture	2732	Cold rolling of narrow strips
2732	Cold rolling of narrow strips	1593	Mfg of wines
3622	Mfg of jewelery	2214	Publishing of sound recordings
2411	Mfg of industrial gases	2840	Forging, pressing, stamping and roll forming of metal
1543	Manufacture of margarine & fats	3622	Mfg of jewelery
1717	Prep. and spinning of other textile fibres	3350	Mfg of watches and clocks
3661	Mfg of imitation jewelery	2215	Other publishing
2511	Mfg of rubber tyres and tubes	2225	Ancillary activities related to printing
2741	Precious metals production	1772	Mfg of knitted & crocheted pullovers

Table 10: Ten most localized NACE 4-digit manufacturing sectors, Germany.

	E&G Index		D&O Index
NACE	*Sector Name*	*NACE*	*Sector Name*
1722	Woollen-type weaving	1722	Woollen-type weaving
1724	Silk-type weaving	1724	Silk-type weaving
2213	Publishing of journals and periodicals	1713	Prep. and spinning of fibres
2622	Mfg of ceramic sanitary fixtures	2731	Cold drawing
2630	Mfg of ceramic tiles and flags	2955	Mfg of machinery for paper
3541	Mfg of motorcycles	1725	Other textile weaving
1910	Tanning and dressing of leather	2913	Mfg of taps and valves
2411	Mfg of industrial gases	1910	Tanning and dressing of leather
3661	Mfg of imitation jewelery	1721	Cotton-type weaving
1771	Mfg of hosiery	1771	Mfg of hosiery

Table 11: Ten most localized NACE 4-digit manufacturing sectors, Italy.

E&G Index		D&O Index	
NACE	*Sector Name*	*NACE*	*Sector Name*
2630	Mfg of ceramic tiles and flags	1760	Mfg of knitted and crocheted fabrics
1594	Mfg of cider and other fruit wines	1715	Throwing and preparation of silk
2624	Mfg of other technical ceramic products	2954	Mfg of machinery for textile
1930	Mfg of footwear	1713	Prep. and spinning of worsted-type fibres
3630	Mfg of musical instruments	1723	Worsted-type weaving
1713	Prep. and spinning of worsted-type fibres	1717	Prep. and spinning of other textile fibres
1723	Worsted-type weaving	1721	Cotton-type weaving
1717	Prep. and spinning of other textile fibres	1712	Prep. and spinning of woollen-type fibres
2052	Mfg of articles of cork	1722	Woollen-type weaving
3650	Mfg of games and toys	1711	Prep. and spinning of cotton-type fibres

Table 12: Ten most localized NACE 4-digit manufacturing sectors, Spain.

E&G Index		D&O Index	
NACE	*Sector Name*	*NACE*	*Sector Name*
3511	Building and repairing of ships	2221	Printing of newspapers
2625	Mfg of other ceramic products	2213	Publishing of journals and periodicals
2954	Manufacture of machinery for textile etc	2211	Publishing of books
3650	Mfg of games and toys	2840	Forging, pressing, stamping of metal
1712	Prep. and spinning of woollen-type fibres	3622	Mfg of jewelery
2221	Printing of newspapers beverages	2415	Mfg of fertilizers and nitrogen compounds
1591	Mfg of distilled potable alcoholic	2320	Mfg of refined petroleum products
1582	Mfg of rusks and biscuits	2513	Mfg of other rubber products
2751	Casting of iron compounds	2215	Other publishing
2415	Mfg of fertilizers and nitrogen	2625	Mfg of other ceramic products

Table 13: Ten most localized NACE 4-digit manufacturing sectors, UK.

Category	Typical Sectors	Average Firm Size	Type of Innovation	Main External Source of Innovation	Main Internal Source of Innovation	Appropriability Conditions	Entry Barriers
Traditional	Traditional Mfg (leather, jewelery)	Small/Medium	Process	Embodied Innovation	Learning-by-doing	Low	Low
Scale Intensive	Bulk materials (steel, glass) Assembly (durables, automobiles)	Medium/Large	Product/Process	Supply Relationships	R&D	Medium	Medium
Specialized Suppliers	Machinery (machinery/equipment) Instruments (medical, precision, optical instruments)	Small	Product	Customer Relationships	Learning-by-doing	High	Medium
Science Based	Electronics/Electrical Pharmaceutical Chemicals	Small/Large	Product/Process	Universities	R&D and R&D centers	High	Very High

Table 14: Pavitt taxonomy of manufacturing industries. Note: Adapted from Pavitt (1984).

E&G Index	Science Based	Specialized Suppliers	Scale Intensive	Traditional
Belgium	0.0196	0.1176	0.3529	0.5098
France	0.0943	0.1038	0.3962	0.4057
Germany	0.0741	0.1019	0.3889	0.4352
Italy	0.0738	0.1639	0.3525	0.4098
Spain	0.0706	0.1235	0.3059	0.5000
UK	0.0652	0.0870	0.4348	0.4130
D&O Index	Science Based	Specialized Suppliers	Scale Intensive	Traditional
Belgium	0.0566	0.1132	0.3019	0.5283
France	0.0990	0.1584	0.3960	0.3465
Germany	0.0708	0.1150	0.4336	0.3805
Italy	0.0918	0.1531	0.4592	0.2959
Spain	0.0728	0.1788	0.3576	0.3907
UK	0.0392	0.1373	0.4706	0.3529

Table 15: Share of localized sectors in each Pavitt category.

Country	Science Based	Specialized Suppliers	Scale Intensive	Traditional
Belgium	0.0688	0.1313	0.3750	0.4250
France	0.0597	0.1443	0.3532	0.4428
Germany	0.0563	0.1255	0.3723	0.4459
Italy	0.0631	0.1311	0.3689	0.4369
Spain	0.0532	0.1241	0.3652	0.4574
UK	0.0957	0.1565	0.3652	0.3826

Table 16: Share of manufacturing sectors in each Pavitt category.

E&G Index	Science Based	Specialized Suppliers	Scale Intensive	Traditional
Belgium	0.2852	0.8964	0.9412	1.1995
France	1.5802	0.7193	1.1217	0.9162
Germany	1.3162	0.8113	1.0446	0.9760
Italy	1.1690	1.2508	0.9553	0.9381
Spain	1.3271	0.9953	0.8375	1.0930
UK	0.6818	0.5556	1.1905	1.0795

D&O Index	Science Based	Specialized Suppliers	Scale Intensive	Traditional
Belgium	1.2041	1.2615	1.1774	1.8180
France	1.9967	1.3220	1.3499	0.9423
Germany	1.3179	0.9600	1.2202	0.8941
Italy	1.6813	1.3492	1.4380	0.7825
Spain	1.5055	1.5837	1.0763	0.9390
UK	0.8235	1.7614	2.5882	1.8529

Table 17: Share of localized sectors in each Pavitt category divided by the share of sectors in each Pavitt category.

Small Firms	E&G index			D&O index	
Country	LOC[1]	LOC[2]	DISP	LOC	DISP
Belgium	0.4623	0.2736	0.5377	0.2547	0.0755
France	0.6813	0.4500	0.3187	0.4000	0.1812
Germany	0.5482	0.2944	0.4518	0.3959	0.1015
Italy	0.7500	0.4940	0.2500	0.3631	0.1726
Spain	0.7860	0.5802	0.2140	0.4733	0.1523
UK	0.5294	0.3088	0.4706	0.3235	0.0735

Large Firms	E&G index			D&O index	
Country	LOC[1]	LOC[2]	DISP	LOC	DISP
Belgium	0.6829	0.2846	0.3171	0.3008	0.0813
France	0.7784	0.5455	0.2216	0.3920	0.1932
Germany	0.7033	0.4306	0.2967	0.4067	0.0718
Italy	0.8092	0.5838	0.1908	0.4971	0.1272
Spain	0.8496	0.6301	0.1504	0.5813	0.1423
UK	0.7703	0.4459	0.2297	0.2973	0.0541

Table 18: Small vs. large firms. Share of sectors localized and dispersed in each country. Small firms: firms with below-median employment. Large firms: firms with above-median employment. Median computed on the industry-pooled within-country employee distribution. Notes: [1] Share of sectors with strictly positive E&G index. [2] Share of sectors significantly localized according to the 2-sigma rule.

	Simulation Scenario	
Country	Uniform Distribution	Small-firm based distribution
Belgium	0.5726	0.5214
France	0.8058	0.7934
Germany	0.9545	0.9421
Italy	0.8445	0.8319
Spain	0.7903	0.6677
UK	0.4935	0.4935

Table 19: Simulation analysis of localization with artificially generated plants (E&G index). Share of sectors that are localized in all $R = 1000$ independent runs. Localized sectors identified using a 2-sigma rule.