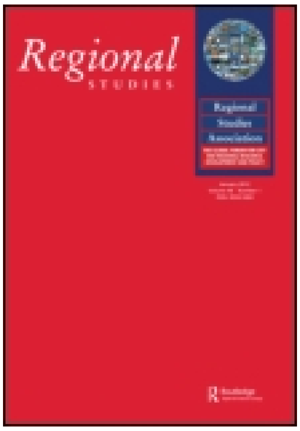

Regional Studies
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cres20

Spatial Localization in Manufacturing: A Cross-Country Analysis

Stefania Vitali[ab], Mauro Napoletano[cd] & Giorgio Fagiolo[d]
[a] ETH Zurich, Systemgestaltung KPL F 28, Kreuzplatz 5, CH-8032 Zurich, Switzerland
[b] Dipartimento di Economia, Università Politecnica delle Marche, Piazzale Martelli, 8, I-60121 Ancona, Italy. Email:
[c] OFCE and SKEMA Business School, (Batiment 2) 250 rue Albert Einstein, F-06560 Valbonne, France. Email:
[d] Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, 33, I-56127 Pisa, Italy.
Published online: 08 Dec 2011.

To cite this article: Stefania Vitali, Mauro Napoletano & Giorgio Fagiolo (2013) Spatial Localization in Manufacturing: A Cross-Country Analysis, Regional Studies, 47:9, 1534-1554, DOI: 10.1080/00343404.2011.625006

To link to this article: http://dx.doi.org/10.1080/00343404.2011.625006

Regional Studies, 2013
Vol. 47, No. 9, 1534–1554, http://dx.doi.org/10.1080/00343404.2011.625006

Routledge Taylor & Francis Group
Regional Studies Association

Spatial Localization in Manufacturing: A Cross-Country Analysis

STEFANIA VITALI*†, MAURO NAPOLETANO‡§ and GIORGIO FAGIOLO§
**ETH Zurich, Systemgestaltung KPL F 28, Kreuzplatz 5, CH-8032 Zurich, Switzerland*
†Dipartimento di Economia, Università Politecnica delle Marche, Piazzale Martelli, 8, I-60121 Ancona, Italy. Email: svitali@univpm.it
‡OFCE and SKEMA Business School, (Batiment 2) 250 rue Albert Einstein, F-06560 Valbonne, France. Email: mauro.napoletano@ofce.sciences-po.fr
§Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, 33, I-56127 Pisa, Italy. Email: giorgio.fagiolo@sssup.it

(Received September 2009: in revised form September 2011)

VITALI S., NAPOLETANO M. and FAGIOLO G. Spatial localization in manufacturing: a cross-country analysis, *Regional Studies*. This paper employs a homogeneous-firm database to investigate industry localization in European countries. More specifically, it compares, across industries and countries, the predictions of two of the most popular localization indexes, that is, the Ellison and Glaeser index of 1997 and the Duranton and Overman index of 2005. Independently from the index used, it is found that localization is a pervasive phenomenon in all countries studied; and the degree of localization is very unevenly distributed across industries in each country. Furthermore, it is shown that in all countries localized sectors are mainly 'traditional' sectors or, if one controls for country industrial structures, science-based sectors. Moreover, it is found that the two indexes significantly diverge in predicting the intensity of localization of the same industry both across and within countries. In turn, these differences point to the different role played by pecuniary versus non-pecuniary externalities in driving firms' location decisions.

Industry localization Manufacturing industries Localization indexes Spatial concentration Spatial correlation Pecuniary externalities Non-pecuniary externalities Cross-country studies

VITALI S., NAPOLETANO M. and FAGIOLO G. 制造业的空间定位：跨国分析，区域研究。本文采用同质的公司数据库考察了欧洲国家的产业本地化。更确切地说，本文比较了两大流行的本地化指标对各行业和国家的预测，即 1997 年的 Ellison－Glaeser 指数与 2005 年的 Duranton－Overman 指数。在独立于上述指标的基础上研究发现，在所有被研究的国家中本 地化都是一个普遍存在的现象，且各产业的本地化程度不一。研究进一步指出，各国中本地化的多为传统部门，或者是控制国家产业机构以及科技基础的行业。此外，研究表明，上述两种指标在预测同一产业在国家间以及国内本地化程度中存在显著分歧。这些分歧反映了资金与非资金外部性在企业选址决策中所发挥的不同作用。

工业本地化 制造业 本地化指标 空间集聚 空间相关性 资金外部性 非资金的外部性 跨国研究

VITALI S., NAPOLETANO M. et FAGIOLO G. La localisation géographique de l'industrie: une analyse transnationale, *Regional Studies*. A partir d'une base de données auprès des entreprises homogènes, cet article cherche à examiner la localisation de l'industrie dans les pays européens. Plus précisément, il fait une comparaison, à travers les industries et les pays, des prédictions de deux des indices de localisation les plus répandues, à savoir, l'indice Ellison–Glaeser qui date de 1997, et l'indice Duranton–Overton de 2005. Quelle que soit l'indice employée, il s'avère que la localisation est un phénomène omniprésent dans les pays éudiés; et que l'importance de la localisation varie de façon irrégulière à travers les industries de chaque pays. Qui plus est, on démontre que dans tous les pays les secteurs localisés sont principalement des secteurs 'traditionnels', ou bien, si l'on tient compte des structures industrielles d'un pays, les secteurs basés sur les sciences. De plus, il s'avère que les deux indices divergent sensiblement pour ce qui est de la prédiction de l'intensité de la localisation de la même industrie, à la fois au-delà et au sein des pays. A leur tour, ces différences laissent indiquer le rôle joué par les effets pécuniers et non-pécuniers externes en tant que forces motrices de la prise de décision des entreprises quant à la localisation.

Localisation de l'industrie Industries Indices de localisation Concentration géographique Corrélation géographique Effets externes pécuniers Effets externes non-pécuniers Études transnationales

VITALI S., NAPOLETANO M. und FAGIOLO G. Räumliche Lokalisation in der produzierenden Industrie: eine länderübergreifende Analyse, *Regional Studies*. In diesem Beitrag wird die Branchenlokalisation in europäischen Ländern mit Hilfe einer homogenen Firmendatenbank untersucht. Insbesondere werden die Prognosen von zwei der beliebtesten Lokalisationsindizes für verschiedene Branchen und Länder miteinander verglichen: dem Ellison–Glaeser-Index von 1997 und dem Duranton–Overman-Index von 2005. Unabhängig vom verwendeten Index stellt sich heraus, dass es sich bei der Lokalisation in sämtlichen untersuchten

Ländern um ein weit verbreitetes Phänomen handelt; das Ausmaß der Lokalisation ist in jedem Land unter den Branchen äußerst ungleichmäßig verteilt. Darüber hinaus zeigt sich, dass es sich in allen Ländern bei den lokalisierten Sektoren in erster Linie um 'traditionelle' Sektoren bzw. bei Berücksichtigung der jeweiligen Branchenstruktur des Landes um wissenschaftsbasierte Sektoren handelt. Ebenso stellen wir fest, dass die beiden Indizes hinsichtlich der Prognose der Intensität der Lokalisation in derselben Branche zwischen verschiedenen Ländern und innerhalb eines Landes signifikant divergieren. Diese Unterschiede weisen wiederum auf die unterschiedliche Rolle der pekuniären und nicht-pekuniären Externalitäten als Faktoren für die Standortentscheidungen von Firmen hin.

Branchenlokalisation Produzierende Industrie Lokalisationsindizes Räumliche Konzentration Räumliche Korrelation Pekuniäre Externalitäten Nicht-pekuniäre Externalitäten Länderübergreifende Studien

VITALI S., NAPOLETANO M. y FAGIOLO G. Localización espacial en el sector manufacturero: un análisis entre países, *Regional Studies*. En este artículo empleamos una base de datos homogénea para empresas con el objetivo de investigar la localización industrial en los países europeos. En concreto, comparamos las predicciones de dos de los índices de localización más conocidos – el índice Ellison y Glaeser de 1997 y el índice Duranton y Overman de 2005 – para diferentes industrias y países. Independientemente del índice que se utilice, observamos que la localización es un fenómeno extendido en todos los países estudiados, y el grado de localización está distribuido de modo muy desigual entre las industrias de cada país. Además, demostramos que en todos los países los sectores localizados son principalmente los sectores 'tradicionales', o, si se tienen en cuenta las estructuras industriales del país, los sectores basados en la ciencia. Asimismo, observamos que los dos índices differen en gran medida a la hora de predecir la intensidad de la localización de la misma industria tanto entre países como dentro de cada uno de ellos. A su vez, estas diferencias señalan el diferente papel desempeñado por los factores externos pecuniarios a diferencia de los no pecuniarios en cuanto a decidir la ubicación de las empresas.

Localización industrial Industrias manufactureras Índices de localización Concentración espacial Correlación espacial Factores externos pecuniarios Factores externos no pecuniarios Estudios entre países

JEL classifications: R3, R12

INTRODUCTION

This paper investigates the empirical location patterns of manufacturing industries in six European countries: Belgium, France, Germany, Italy, Spain and the United Kingdom. Drawing on a comprehensive source covering data on European manufacturing firms, it simultaneously performs both a cross-country and a cross-sector analysis of industry spatial localization patterns by employing two of the most popular localization measures: the Ellison and Glaeser (E&G) index (ELLISON and GLAESER, 1997) and the Duranton and Overman (D&O) index (DURANTON and OVERMAN, 2005). The 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 indexes across different European countries in a homogeneous way. Indeed, as argued 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 mainly to a single index. Results are therefore hardly comparable, due to the inherent heterogeneity in data collection and definitions of variables (for example, firm size).

The analysis of firms' location has attracted the attention of economists for a very long time (for example, MARSHALL, 1920). More recently, a relevant body of theoretical research in the 'New Economic Geography' literature (for example, KRUGMAN, 1991; FUJITA *et al.*, 1999) has aimed at explaining what might be considered the basic stylized fact of economic geography, that is, that firms look more clustered in space than any theory of comparative advantage would predict. New Economic Geography emphasizes the role of increasing returns, arising from various types of externalities related to the number of firms into an area, as a fundamental determinant of industry agglomeration.

Beside these theoretical contributions, a good deal of empirical research has investigated localization in manufacturing industries (among others, see ELLISON and GLAESER, 1997; MAUREL and SÉDILLOT, 1999; BARRIOS *et al.*, 2005; LAFOURCADE and MION, 2006; and 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 measures of the degree of firms' clustering in space ('localization indexes'). Despite very similar methodological approaches, however, the literature has so far been quite heterogeneous in terms of the variety of the measures employed (for a survey, see, for example, COMBES and OVERMAN, 2004).

To this end, the contributions in ELLISON and GLAESER (1997), DUMAIS *et al.* (2002), and MAUREL and SÉDILLOT (1999) provide localization indexes that simultaneously: (1) control for the overall concentration of manufacturing; (2) control for industry concentration, and (3) provide a null hypothesis against which to verify 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 indexes represent key advances in the measuring of localization. Nevertheless, by construction, they require an *ex-ante* partitioning of the geographical space (for example, a country) into smaller units (for example, 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, 2004). The division of the space into subunits has the advantage of making the computational problems involved in the measurement of localization easier. However, it also introduces possible biases in the analysis (the modifiable area unit problem [MAUP]; Arbia, 2001; and more recently, see Lafourcade and Mion, 2007; and Briant *et al.*, 2010).

First, comparisons among countries are difficult, as the areas of spatial subunits may significantly vary across different countries. Second, and relatedly, comparisons become difficult also within countries across different spatial scales (for example, departments versus regions). Finally, clusters of firms located at the borders of neighbouring 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 by Arbia (2001) and Lafourcade and Mion (2006), indexes requiring the division of the space into smaller units can only capture the 'spatial concentration' of industrial activity into some areas. They cannot measure instead 'true agglomeration', that is, the degree of spatial correlation in firms' location choices. Localization indexes that tackle the foregoing problems are those proposed in Moran (1950) and Duranton and Overman (2005). In particular, the D&O index relies on the empirical distribution of distances across firms, computed by locating firms on the basis of their postal codes. Moreover, postal codes are more detailed than any alternative spatial breakdown, and are very comparable across countries (for more on that, see the second and third sections).

This paper is an attempt to improve upon the foregoing literature along two dimensions. First, it simultaneously performs an investigation of industry localization in several European Union countries by exploiting a database of firms that is homogeneous across European countries. In this respect, countries in the European Union provide a natural arena to test the implications of many New Economic Geography theories. This is because decades of integration across European countries have dramatically lowered trade and transportation costs inside European countries. This should have spurred industry agglomeration processes largely independent from the physical attributes of locations and driven instead by increasing returns and externalities at the industry level (Ottaviano and Thisse, 2001; Ottaviano, 1999). On these grounds, one should expect a high number of localized industries in European countries, despite the different sizes of countries. Moreover, if localization processes in European Union countries are mainly driven by industry-specific forces, one should not expect significant differences in the intensity of localization of the same industry across countries. So far the answers to the above questions have been limited by the availability of harmonized data at the firm level. This work can be therefore be considered a first attempt to provide a broad analysis of agglomeration in the European Union based on firm-level data.

Second, this paper departs from the standard practice of analysing localization employing a single index by performing a comparative study of different localization indexes (E&G and D&O) to estimate localization patterns. In particular, it compares the two indexes with respect to their predictions about: (1) the number of localized sectors; (2) the intensity of industry localization forces; and (3) the types of localized sectors (identified by using Pavitt taxonomy; Pavitt, 1984). Making such a comparison is important because, as discussed above, indexes typically differ in their treatment of the space. The choice of the index could thus induce serious distortions in the results due to the MAUP. Whether or not this problem significantly affects the measurement of localization remains, however, an empirical question. One goal of this paper is precisely to evaluate the biases introduced by the MAUP along three dimensions (pervasiveness, intensity of localization forces and type of localized industries), and across European countries having different sizes and spatial units (for similar remarks, see Briant *et al.*, 2010). Finding that some results about empirical localization patterns do not differ between the two indexes is good news in this respect because it suggests that differences in localization across sectors and/or countries are not spuriously due to biases related to the MAUP, but they rather reflect differences in localization forces. Given this objective, the choice about which indexes should be compared is rather natural. Indeed, the E&G and D&O indexes lie at opposite extremes in the treatment of geographical space. It turns out that comparison across the results they return is a very good test of the distortions that the MAUP can introduce in the analysis of localization.[1]

Furthermore, there is another important dimension along which differences between the E&G and D&O indexes can provide a contribution to empirical agglomeration patterns. The E&G index is an unbiased estimate of localization resulting of natural advantages and sector-specific spillovers dependent on the number of firms in a particular area. However, as argued at more length by Ottaviano and Thisse (2001) and Ottaviano and Lamorgese (2003), the latter is an aggregate of different externalities, whose strength is highly dependent on spatial distance. 'Pecuniary' externalities, that is, those that are the by-product of market relationships (for example, related to the existence of a large market demand, or the availability of the large labour supply), typically span their effects over large distances with respect to 'non-pecuniary' externalities (for example, related to the existence of infrastructures, or to knowledge exchanges). It follows that non-pecuniary externalities are likely to introduce much stronger correlation in spatial location choices of firms. Since the latter is captured by the D&O index, one can interpret similarities and differences in results across the two indexes for the same industry as reflecting higher strengths of, respectively, pecuniary and non-pecuniary externalities in generating localization patterns.

Previous attempts in the same direction can be found in the work of BARRIOS *et al.* (2005); LAFOURCADE and MION (2006), and FORNAHL and BRENNER (2009). However, differently from these, the present paper considers a larger number of (size-heterogeneous) countries. Furthermore, to account for spatial features of the data, it employs the D&O index rather than the Moran index (MORAN, 1950). This choice has been made because, as discussed at more length by ARBIA (2001), the Moran index cannot entirely capture the observed variability in spatial permutations. Indeed, it requires an *ex-ante* partition of space (these points will be discussed in the third section).

The results do not reject the hypothesis that European Union countries are characterized by significant levels of spatial localization in their manufacturing industries. Indeed shares of localization sectors are very high in most countries analysed. Moreover, the characteristics of industry agglomeration (intensity, type of industry) are invariant across countries when one considers the overall manufacturing industry and broad categories of industries. In contrast, national specificities emerge at more disaggregated levels of analysis (that is, at four-digit industries). Furthermore, the analysis suggests that in all six countries localized industries are mainly 'traditional' ones (jewellery, wine, textiles, etc.), as well as those where scale economies are important. This outcome mainly reflects the composition of countries' industrial structures, generally skewed toward traditional and scale-intensive sectors. Once one controls for such a factor, it is found that science-based industries become those where localization is more pervasive.

How much are the above results dependent on the index used, and therefore on the treatment of geographical space? First, the analysis shows that spatial disaggregation does not matter in making predictions about the pervasiveness of industry localization, both considering aggregate manufacturing as well as broad categories of sectors. In contrast, the treatment of space becomes very important when analysing the intensity of agglomeration. In particular, employing *ex-ante* space partitions (NUTS-3 in the work) introduces spurious cross-country differences in average localization intensity. Indeed, country differences are significant when the E&G index is applied but not when the D&O index is used.

The paper is organized as follows. The second section presents the database used in the analysis. The third section describes the localization indexes employed. It begins with the E&G index and then moves to the D&O index. The fourth section is devoted to the presentation and discussion of the results on the empirical analysis of localization in European Union countries. The fifth section concludes. Finally, the Appendix contains a robustness analysis of the results discussed in the paper, with particular emphasis on the possible biases due to using firms rather than plants as the object of the analysis.

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. http://www.bvdep.com/en/ORBIS). From this extensive dataset information about location (that is, postal codes), employment and industrial classification of firms in six European countries for the period 2004–2006 have been extracted. All information is available at the firm level and is derived from companies' annual reports.

The countries chosen for the analysis are Belgium, France, Germany, Italy, Spain and the UK. 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. In addition to US-focused research (ELLISON and GLAESER, 1997; ROSENTHAL and STRANGE, 2001; HOLMES and STEVENS, 2002; KIM, 1995), existing contributions have been studying industry localization patterns in the UK (DEVEREUX *et al.*, 2004; DURANTON and OVERMAN, 2005), Belgium (BERTINELLI and DECROP, 2005), France (MAUREL and SÉDILLOT, 1999), Italy (LAFOURCADE and MION, 2006), Germany (BRENNER, 2006), and Ireland and Portugal (BARRIOS *et al.*, 2005). For a review, see COMBES and OVERMAN (2004). On the other hand, we were constrained by data availability. Indeed, both the ORBIS and the databases used for geographical coordinates (see below) cover more countries than those employed in this analysis. However, for some countries data on firm size and postal codes were available only for a very small number of sectors. In other countries, data on firm size and postal codes were available, but data on geographical coordinates (needed to compute the D&O index) were of poor quality. Only for the countries in the sample enough data were collected on geographical coordinates, firm size, and postal codes needed to compute both the E&G and D&O indexes for a sufficiently high number of sectors.

Firm-level data other than localization are available for more years. Unfortunately, the last available year differs among countries. More precisely, there are data until 2005 for Belgium, France, Italy and the UK; until 2004 for Spain; and until 2006 for Germany. Notice that information on firm localization only refers to the last available year. In order to keep as many countries as possible in the analysis, and to be sure that localization data are synchronized with other firm-specific variables, data were employed only for the last available year in the database. This, of course, prevents a proper cross-section analysis from being performed, but this is not expected to be a source of important bias to the analysis. Indeed, given the relatively short time span covered by the database, only a small fraction of all firms is going to change their locations. Similarly, sectors are not very likely to change their industrial structure dramatically.

The analysis is limited to manufacturing industries as defined by the NACE (Nomenclature générale des activités économiques dans les Communautés Européennes) classification (NACE Rev. 1 section D). More specifically, following DURANTON and OVERMAN (2005), the analysis is restricted to sectors with more than ten firms. This allows sectors to be excluded where localization is the result of location choices by a few firms and, therefore, to focus on clustering phenomena where localization forces attract a significant bunch of firms.

To identify spatial subunits, the NUTS classification is applied (cf. http://www.ec.europa.eu/eurostat/ramon/nuts). NUTS (Nomenclature des Unités Territoriales Statistiques) is a hierarchical classification at five levels (three regional and two local), extensively used for comparative statistics among European countries. In line with previous studies (for example, MAUREL and SÉDILLOT, 1999; BERTINELLI and DECROP, 2005) NUTS-3 regions are used. Firms are then assigned to each subunit 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' database).

Since one of the two indexes employed in the analysis (the D&O index, cf. the third section) requires the identification of the longitude–latitude coordinates of firms in space, data from the 'TeleAtlas Multinet Europe' database (cf. http://www.teleatlas.com) are also employed. More precisely, this database provides the spatial coordinates of the contour of the areas corresponding to postal codes in the sample. Each firm is then assigned coordinates coinciding with the centroid of each postal code area.

Table 1 shows some descriptive statistics for the sample of firms studied. The number of firms under analysis is highly variable among the countries considered, and in some cases (the 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. Furthermore, in the 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.[2] Table 2 shows instead some descriptive statistics for the sample of geographical locations considered, namely NUTS-3, and postal code areas (PCAs). Considering the available number of locations, Table 2 suggests that spatial disaggregation varies significantly across the countries considered. As might be expected, the number of NUTS-3 regions and postal codes areas is the lowest in Belgium (number of NUTS-3 = 44, number of PCAs = 1157). Moreover, Germany is the country with more NUTS-3 areas (423), whereas Spain has more postal codes areas (10 494). Looking at the location average size, however, reveals that one gains much in cross-country comparability by moving from NUTS-3 to PCAs. Indeed, the ratio between the largest NUTS-3 average size (Spain) and the smallest one (Belgium) is 14.53. Considering PCAs, this ratio (largest = France, smallest = Belgium) boils down to 3.03.

The significant cross-country variability in the number of firms and average firm size might suggest the presence of possible biases in the analysis because the characteristics of the total firm population in each country are very different from the ones in the sample. In particular, the bias is likely to be stronger the larger is firm size in the sample with respect to the size in the population. Indeed, the analysis presented in the previous sections relies on a database where

Table 1. Descriptive statistics for the sample of firms studied

Country	Number of firms	Last year available	Number of NACE sectors[a]	Average number of employees[b]	Number of firms[c]	Average firm size[b]
Belgium	13 032	2005	160 (227)	1603	65	20
France	50 396	2005	201 (237)	4051	227	16
Germany	62 588	2006	231 (239)	7445	264	27
Italy	26 940	2005	206 (234)	5472	120	42
Spain	67 809	2004	282 (309)	3711	228	15
UK	6056	2005	115 (204)	7675	157	33

Notes: [a] Number of NACE sectors with at least ten active firms. Total number of sectors in each country is given in parentheses.
[b] In NACE sectors. Size: firm employees.
[c] Average number of firms per sector.

Table 2. Descriptive statistics for the sample of geographical locations studied

Country	Number of NUTS-3	Average size of NUTS-3	SD of NUTS-3 size	Number of PCA	Average size of PCA	SD of PCA size
Belgium	44	693.84	356.02	1157	41.29	41.98
France	71	5729.56	2122.27	5622	125.29	123.15
Germany	423	823.53	609.35	8273	67.08	86.26
Italy	107	2816.17	1591.72	4415	93.24	179.07
Spain	49	10 072.61	5055.63	10494	64.53	110.75
UK	105	1438.22	1684.31	8922	45.86	158.57

Note: Both NUTS-3 and postal code areas (PCAs) sizes are measured in km^2.

observational units are firms rather than plants. 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 headquarters (for a more detailed discussion of this point, see also the Appendix). Such an upward bias is likely to be affected by the share of medium and large firms in the data, due to the positive correlation between firm size and the number of firms' plants typically observed in empirical studies (cf., for example, COAD, 2008). However a quick comparison of the sample with one of the other cross-country empirical studies having large industry coverage indicates that the aforementioned bias is not likely to be significant in the context. For instance, if one considers the comprehensive cross-country study of BARTELSMAN *et al.* (2005) (not covering all countries in the work, though), then average manufacturing firm size in the sample is 1:8 times higher in Italy.[3] However, average size is 50% lower in France, 31% lower in Germany, 19% lower in the UK (cf. the average size statistics in Table 1 with average manufacturing firm size in BARTELSMAN *et al.*, 2005, table 3). Second, the data contained in the ORBIS dataset (so far the only publicly available dataset at the firm level) are harmonized across countries. This minimizes other possible biases that could arise in a comparative cross-country analysis of localization performed on non-harmonized datasets, even having larger industry coverage.[4]

Table 3. Share of sectors localized (LOC) and dispersed (DISP) in each country

Country	E&G index			D&O index	
	LOC[a]	LOC[b]	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

Notes: [a] Share of sectors with a strictly positive E&G index.
[b] Share of sectors significantly localized according to the 2-sigma rule.

Furthermore, there are other two arguments that indicate that the biases introduced by data limitations are not strong in the context. The first is related to the fact that the indexes used in the analysis measure industry localization by controlling (partially or totally) for the characteristics of the distribution of firm size within each industry (see the third section). It follows that by employing these indexes the measurement of localization of a given industry is not affected by the particular distribution of firm size of that industry, and it is therefore independent of the possible biases that are affecting the latter. Second, it is not clear in which direction this is biasing the estimates of the extent and intensity of the overall manufacturing localization. This is because it is not clear how firm size is affecting the location choice of firms and, therefore, industry agglomeration. Empirical studies available so far (cf. HOLMES and STEVENS, 2002; LAFOURCADE and MION, 2006) have indeed provided mixed answers on the issue.

Finally, a second source of bias in the analysis might be represented by the fact that, differently from the literature, firms rather than plants data are used in the 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 concentrated in the same area (the headquarters). To check for this possible bias, the Appendix provides a more detailed analysis of localization by using different approaches to detect and locate a firm's plants in space. The results of these robustness analyses show that localization is not overestimated by using firms' rather than plants' data.

LOCALIZATION INDEXES

This section describes the properties of the localization indexes employed in the investigation. As argued in the Introduction, the literature has so far proposed several measures to capture firms' spatial clustering. A full account of the properties of the different localization indexes is beyond the scope of this paper. COMBES and OVERMAN (2004) provide a detailed description of some of the most popular indexes, together with a discussion of their properties. Here the focus is on two indexes that have gained a lot of attention in recent years: the index proposed by ELLISON and GLAESER (1997)

(E&G index) and that introduced by DURANTON and OVERMAN (2005) (D&O index). Both indexes 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 can capture.

The Ellison and Glaeser index

The E&G index proposed by 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 in such a way to maximize its profits π_{ij} from locating in area i. In turn, profits π_{ij} are determined by cross-sectoral homogeneous 'natural advantages' attached to area *i*, by sector-specific spillovers created by business units that have previously chosen that location and, finally, by a random factor idiosyncratic to the *j*th business unit (ELLISON and GLAESER, 1997; MAUREL and SÈDILLOT, 1999).

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:

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

where G_n is the 'raw-concentration index':

$$G_n = \sum_i (s_i - x_i)^2 \qquad 2)$$

In equations (1) and (2) s_i is the share of industry's employment in area *i*; and x_i is the share of aggregate manufacturing employment in area *i*. The term H_n is the Herfindahl index of industry concentration:

$$H_n = \sum_j z_j^2$$

where *z* is the share of employment of the *j*th firm in the industry.

The E&G index has many interesting properties, as compared with other indexes proposed in the literature. First, similarly to the measure proposed by KRUGMAN (1991), it controls for the overall tendency of manufacturing to localize in space (for example, 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 of what is predicted by industry concentration. Indeed, the Herfindahl index directly enters in equation (1) to re-scale the raw index G_n. Finally, ELLLISON and GLAESER (1997) show that if agents take their location decisions according to the model outlined at the beginning of the section, then the index in equation (1) is an unbiased estimate of the following relation:

$$\gamma_n = \gamma_{na} + \gamma_s - \gamma_{na}\,\gamma_s \tag{3}$$

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 to evaluate 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 factors. This in turn implies that industries characterized by a positive E&G value display 'excess' localization, as compared with 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. The E&G index has received a lot of attention in the literature, and it is still considered as a sort of benchmark against which other indexes should be compared (for a very recent attempt in the same direction in the present paper, see also BRIANT *et al.*, 2010). At the same time, this index has also some important weaknesses, some of which may introduce serious biases in the measurement of empirical localization patterns.

One of the major drawbacks of the E&G index is the lack of a statistical procedure to measure significantly the degree of excess localization (or dispersion) of an industry. To solve such a problem partially 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 $\gamma_n < 0.02$. Moreover, it is very localized if $\gamma_n > 0.5$. These thresholds were chosen by the authors via a heuristic procedure based on their application on US data and are somewhat arbitrary.

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 (for example, ROSENTHAL and STRANGE, 2001;

DEVEREUX *et al.*, 2004; BARRIOS *et al.*, 2005). Since under the null hypothesis of random location $\gamma_n = 0$ and:

$$E(G_n) = \left(1 - \sum_i x_i^2\right) H_n$$

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

$$1 - \left(\sum_i x_i^2\right) H_n$$

is twice larger (smaller) than the standard deviation σ_G of the raw concentration index (cf. ELLISON and GLAESER, 1997):

$$\sigma_G = \sqrt{\begin{array}{l} 2\{H^2[\sum_i x_i^2 - 2\sum_i x_i^3 + (\sum_i x_i^2)^2] \\ - \sum_j z_j^4 [\sum_i x_i^2 - 4\sum_i x_i^3 + 3(\sum_i x_i^2)^2]\} \end{array}} \quad (4)$$

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, such a criterion will be used to evaluate the statistical significance of localization (or dispersion) of an industry.

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 (for example, regions, departments). Unfortunately, the division of the space into subunits has also several disadvantages (grouped under the MAUP label). First, it is not a-priori clear what is the optimal spatial breakdown at which firm clustering should be measured. For example, one could decide to compute the index considering counties, regions or different NUTS layers. In addition, the number of locations and their average size can display significant cross-country heterogeneity even at same level of spatial breakdown (see Table 2 and the discussion at the end of the second section). All this undermines comparison, both cross-country and across different disaggregation levels (for a discussion of this point see also ROSENTHAL and STRANGE, 2001; and DEVEREUX *et al.*, 2004). Second, as argued at more length by ARBIA (2001), the very computation of cumulative shares of economic activity concentrated in spatial subunits implies disregarding the spatial nature of the data. Indeed, indexes based on cumulative shares (such as 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 neighbouring areas. Moreover, by focusing on the 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 neighbouring areas). This means disregarding 'true agglomeration', that is, the degree of spatial correlation in firms' location decisions (ARBIA, 2001; LAFOURCADE and MION, 2006; DURANTON and OVERMAN, 2005; BRIANT *et al.*, 2010). Accounting for the degree of spatial correlation might also be helpful when disentangling the importance of different types of externalities in the generation of localization patterns. Indeed, the E&G index is an unbiased estimate of agglomeration, resulting of natural advantages and sector-specific spillovers dependent on the number of firms in a particular area. However, as argued at more length by OTTAVIANO and THISSE (2001) and OTTAVIANO and LAMORGESE (2003), the latter is an aggregate of different externalities, whose strength is highly dependent on spatial distance. 'Pecuniary' externalities, that is, those that are the by-product of market relationships (for example, related to the existence of a large market demand, or the availability of a large labour supply), typically diffuse over large distances with respect to 'non-pecuniary' externalities (for example, related to the existence of infrastructures or to knowledge exchanges). It follows that technological externalities are likely to introduce much stronger correlation in the spatial location choices of firms than pecuniary externalities.

Two indexes that account for spatial correlation in firms' location decisions are those proposed by MORAN (1950) and DURANTON and OVERMAN (2005). The Moran index still requires the computation of the level of economic activity into predetermined areas and does not solve all the space-related problems described above (ARBIA, 2001).[5] In light of these considerations, this paper has preferred to focus on the D&O index as the only alternative to the E&G index that accounts for the spatial features of the data. Indeed, the D&O index involves the computation of distances across PCAs, therefore it seems better equipped to deal with the characteristics of the spatial distribution of firms into industries. Moreover, PCAs are much more detailed than any other spatial breakdown. In addition, they are much more comparable across countries (cf. Table 2 and the end of the second section).

To compute the D&O index, Euclidean distances must be built between pairs of economic units (plants or firms) in each industry by employing 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 $M(M-1)/2$. The density of distances can then be estimated by using the (Gaussian) kernel function:

$$K(d) = \frac{1}{M(M-1)b} \sum_{h=1}^{M-1} \sum_{j=h+1}^{M} f\left(\frac{d-d_{hj}}{b}\right) \qquad (5)$$

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 kilometers.[6]

Obviously, studying the distribution of kernel densities alone does not give information about whether or not a sector is localized. To solve this problem, the D&O index allows for a rigorous statistical test of industry localization to be made. The test involves the comparison of the empirical density to artificially generated distributions based on the random location of firms in space. Note that this procedure also controls 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 the random re-location of firms in space.

In what follows, the section then bootstraps 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. It then builds *global* confidence bands. Each confidence band 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 at several target distances.[7] For an industry n, the index of (global) localization (A) and (global) dispersion (Δ) are given by:

$$A_n(d) = K_n(d) - \overline{\overline{K}}_n \qquad (6)$$

$$\Delta_n(d) = \begin{cases} \underset{\approx}{K}_n - K_n(d) & \text{if} \sum dA_n(d) = 0 \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

where $\overline{\overline{K}}_n$ is the upper confidence band in industry I; while $\underset{\approx}{K}_n$ is the lower confidence band.

The measurement of localization without an *ex-ante* partition of geographical space is a key strength of 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 a very detailed level.

RESULTS

This section presents the results of the analysis on the spatial distribution of firms in manufacturing industries. It begins by investigating the extent of localization in the countries analysed. In other words, it studies whether the number of industries where firms co-locate is significant in the country under analysis. In addition, it studies whether the fraction of localized industries displays cross-country variation. Furthermore, it investigates how strong localization is. Indeed, the value returned by each localization index captures the strength of localization forces into an industry (see the third section). It is then worthwhile analysing whether the intensity of those forces is heterogeneous across countries and sectors. Finally, it carries 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, the results produced by the two localization indexes employed (E&G and D&O) are compared. As discussed in the third section, these indexes markedly differ in their approach to the measurement of localization. Thus, as suggested by ARBIA (2001), the comparison between the results of the two indexes may help to detect the biases introduced by the MAUP in the analysis. Moreover, as indicated by OTTAVIANO and THISSE (2001) and OTTAVIANO and LAMORGESE (2003), different pecuniary and technological externalities induce different degrees of spatial correlation in firms' location decision. Since the spatial correlation is captured by the D&O index but not by the E&G index, the comparison between the two indexes helps in identifying the different role of the above-mentioned externalities in generating localization patterns.

How many industries are localized?

This subsection starts by assessing how many industries are localized in each country considered. Table 3 shows the fraction of sectors localized and dispersed using, respectively, the E&G and D&O indexes. 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 out to be very high in all countries considered (Table 3, column 1). This result is consistent with previous studies in the literature (for example, DEVEREUX *et al.*, 2004; MAUREL and SÉDILLOT, 1999; LAFOURCADE and MION, 2006). 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 50% in Belgium (from 0.70 to 0.32), whereas in Spain and Italy 60% of all sectors still display localization after the application of the stricter rule. The fraction of sectors displaying excess dispersion is very low in all countries considered. Finally, none of

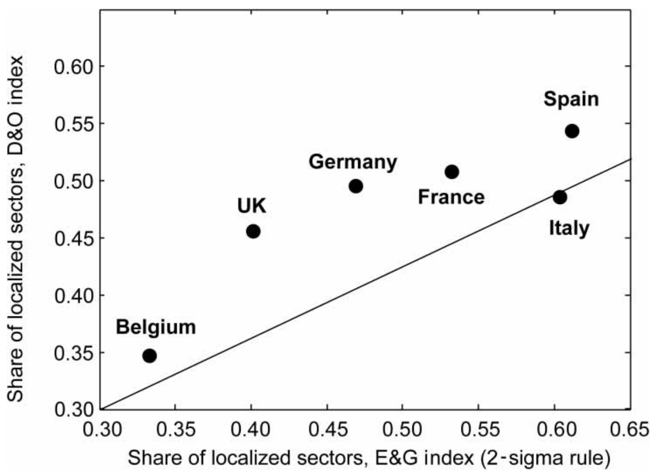

Fig. 1. Share of the Ellison and Glaeser (E&G) index localized sectors (x-axis) versus the share of the Duranton and Overman (D&O) index localized sectors (y-axis)

the sectors studied was significantly dispersed according to the 2-sigma rule criterion. Turning to the D&O index, it is found that in all countries except 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. A similar result for the UK has already been emphasized by DURANTON and OVERMAN (2005). However, the share of localized sectors is very similar with that obtained by applying the 2-sigma rule to the E&G index (cf. Table 3, columns 2 and 4). Fig. 1 summarizes the patterns of localization emerging in each country from the application of the E&G and D&O indexes. Fig. 1 starkly reveals that both indexes produce the same ranking across countries with respect to the pervasiveness of localization, although (as noted above) the share of localized sectors is, in all cases but one (Italy), higher according to the E&G index.

To sum up, the above confirms the conjecture supported by the theory (see the Introduction) that European countries are all characterized by a high number of localized industries. In addition, the absence of marked differences between the two indexes within each country (at least considering sectors for which localization is statistically significant) indicates that: (1) the treatment of geographical space does not matter too much when making predictions about the pervasiveness of localization; and (2) observed localization in European Union countries is equally driven by pecuniary and non-pecuniary externalities. The section now turns to check whether the above findings are confirmed when investigating other important aspects of localization (intensity, type of industry).

How much are industries localized?

As discussed in the third section, the magnitude of the E&G and D&O indexes can provide some information not only on whether a sector is localized or dispersed, but also on the intensity of localization, and therefore on the forces underlying its emergence. *Strictu sensu*, this is true for statistically significant values of the E&G index, whereas it applies by construction to the D&O index. For instance, the E&G index is the result of the combination of forces arising from cross-firm spillovers and geographical advantages of specific areas (see equation 3). Likewise, the value of the D&O index captures the extent to which the spatial distribution of firms in a sector deviates from the one generated under the hypothesis of random firm-location choices. In addition, it can be read as indicator of the strength of non-pecuniary externalities in driving localization patterns. In light of these remarks, this section studies in detail the cross-sector distributional properties of the two indexes in each country.

For each index, Table 4 reports the first four sample moments of within-country distributions of localized industries in the countries considered. Results clearly

Table 4. Moments of empirical distributions of localized industries

	Mean	SD	Skewness	Kurtosis
E&G Index				
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				
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

indicate that, in all countries, cross-sector 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. Figs 2 and 3).

The foregoing results are in line with previous findings in the literature (for example, ELLISON and GLAESER, 1997; MAUREL and SÉDILLOT, 1999; DURANTON and OVERMAN, 2005), which, however, make use of heterogeneous databases and statistical procedures. A perfect comparison of mean values with those of other studies in the literature is of course impossible because of differences across the samples studied. Nevertheless, considering the E&G index, DURANTON and OVERMAN (2005) found that localization is slightly more intense in the UK (0.034) than what is found in Italy (0.033) and France (0.032) by other studies (respectively, LAFOURCADE and MION, 2006; and MAUREL and SÉDILLOT, 1999). This ranking is confirmed in the empirical investigation (Tables 4 and 5). What is more, the results seem to be robust to the index employed. Indeed, both indexes deliver distributions of localized sectors having similar statistical properties (cf. Table 4 and Figs 2 and 3). Nonetheless, the two indexes produce different cross-country rankings with respect to average localization intensity. For instance, both indexes predict that average intensity is the lowest in Germany. However, the E&G index indicates that localization forces are on average higher in the 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 investigate further the cross-country differences in average localization intensity detected above, a Wilcoxon rank-sum (one-sided, non-paired) test is performed for each pair of countries (c_1; c_2).[8] The null hypothesis is that average localization intensity is the same across the selected pair of countries, that is, 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 of country c_2. Wilcoxon test statistics, together with their corresponding (exact) p-values, are reported in Table 5.

Most of the cross-country differences detected by sample moments are confirmed. Consider first the E&G index (in what follows, a 5% threshold is used for convenience). Average intensity is significantly larger in the UK with respect to all other countries

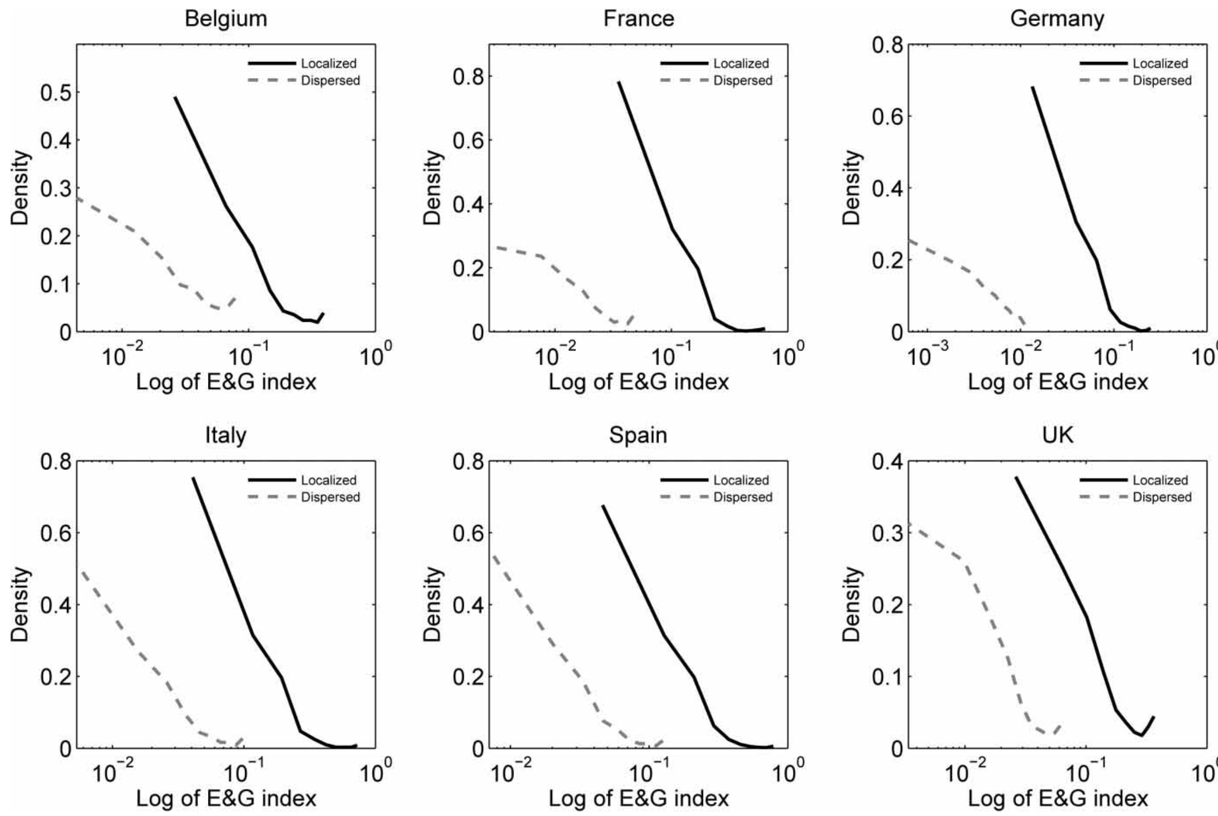

Fig. 2. Kernel density estimates of the distribution of the Ellison and Glaeser (E&G) index: localized versus dispersed sectors. A sector is localized (respectively, dispersed) if the value of the E&G index is greater (respectively, smaller) than zero

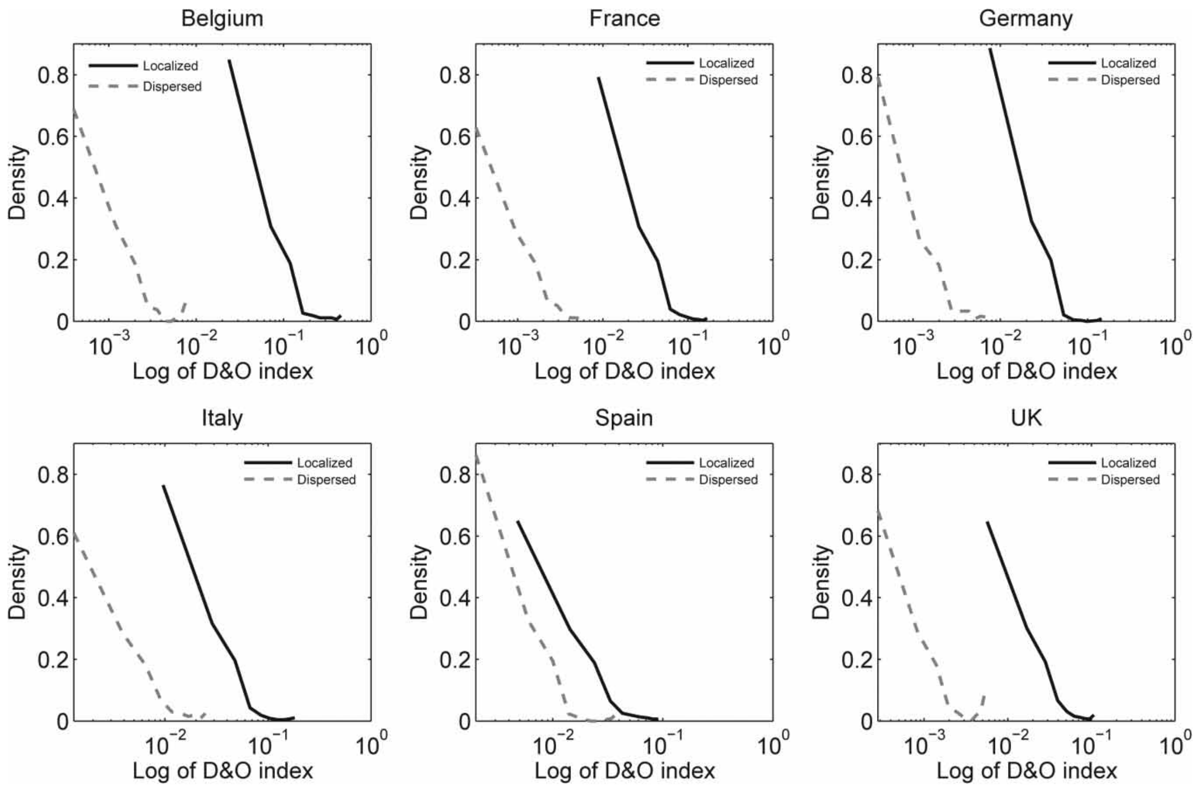

Fig. 3. Kernel density estimates of the distribution of the Duranton and Overman (D&O) index: localized versus dispersed sectors

considered. Both Belgium and Spain present an average intensity larger than that for France and Germany. Moreover, average intensity is significantly lower in Germany with respect to all countries. Notice that if Belgium is compared with Italy and Belgium is compared with Spain, the null hypothesis cannot be rejected, irrespective of the two one-sided alternatives. However, these results are by far altered when the D&O index is applied to the data. Indeed, the null of equal average is not rejected for all pairs of countries except for the whole of Germany's profile and for the Belgium–France comparison.

Additional evidence on intensity patterns comes from the analysis of cross-country Spearman correlation matrices in industry rankings produced by the two indexes employed (cf. Table 6). Predictions about rank correlations differ markedly across the two indexes employed. On the one hand, the E&G index assigns considerable positive correspondence in ranks to all countries considered, with an average level of correlation of 0.59.

Table 5. One-sided Wilcoxon rank-sum test statistics (W) and exact p-values

	Belgium		France		Germany		Italy		Spain		UK	
	W	*p*-value	*W*	*p*-value	*W*	*p*-value	*W*	*p*-value	*W*	*p*-value	*W*	*p*-value
E&G index												
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	12 106	0.0074	–	–	3240	0.9626
UK	1417	0.0390	3664	0.0000	4288	0.0000	3764	0.0003	4580	0.0374	–	–
D&O index												
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	–	–

Note: Due to the properties of the test statistics *W*, the sum of *p*-values of two entries symmetrical to the main diagonal is 1.

Table 6. Cross-country Spearman rank-correlation matrices for localized sectors

	France	Germany	Italy	Spain	UK
E&G index					
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 index					
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)

Note: Localized sectors in E&G results are computed according to the 2-sigma rule. *P*-values are given in parentheses.

Table 7. Spearman rank correlation between E&G and D&O indexes in localized sectors

	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)

Note: Localized sectors in E&G results are computed according to the 2-sigma rule. *P*-values are given in parentheses.

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 the highest rank correlation are Belgium and France, France and Spain, and the UK and Germany, with coefficients equal to 0.73, 0.71 and 0.74, respectively. In contrast, the D&O index suggests that such pairs are Belgium and Germany, and Italy and Spain.

The above results indicate that localization indexes significantly diverge in predicting the intensity of localization forces for the same industry across countries. Interestingly, the same type of divergence is observed also within countries. Indeed, as shown in Table 7, 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 are in general quite small. In particular, the correlation between the two indexes appears much weaker, vis-à-vis their predictions on within-country shares of localized sectors (cf. Fig. 1 and Table 3).

This section now provides some interpretations of the above findings. On the one hand, the fact that the shape of the distributions of index values is invariant across the European countries considered provides further support for the idea that in an economically integrated area with low trade and transportation costs the characteristics of localization are not dependent on country-specific factors. Some national specificities seem to emerge when considering the average intensity of the overall manufacturing sector (cf. the Wilcox test results above). These specificities are, however, highly dependent on the index used. On the other hand, marked cross-country differences emerge at more disaggregated levels of analysis (that is, at four-digit industries). In particular, this is evidenced by the cross-country Spearman rank correlations, obtained using the D&O index, which are in general very low and not statically significant. This means that – for the same industry – firm location choices are much more spatially autocorrelated in one country but not in another country.

Furthermore, the above findings provide empirical support in favour of the claim that the analysis of industry localization is sensitive to the type of index used (Arbia, 2001). In more 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, similar mismatches are found across countries, as indicated by cross-index differences found in average intensity and in ranking correlations. As mentioned above, the weak cross-country correlation found by applying the D&O index shows that location choices of firms in the same industry may display high levels of spatial correlation in some countries, but much less in others. In turn this might be read as an indication of the fact that non-pecuniary externalities play a very important role in generating national specificities in industry localization patterns.

Which industries are localized?

The results presented in the previous sections show that, independently of the index used and of the country analysed: (1) localization is a pervasive phenomenon; and (2) localization forces are very uneven across manufacturing sectors. Moreover, they indicate that different indexes make quite different predictions about the intensity of the forces underlying the emergence of industrial localization within each country and that the same industry can be characterized by different levels of spatial correlation in firms' location decision across different countries. This section completes the analysis of the characteristics of industry localization by investigating the composition of the type of localized industries in the countries considered.

It begins by looking at how much groups of localized industries are similar across countries. Despite similar shares (cf. Table 3), countries could indeed be very different in terms of the composition of the group of localized sectors (for example, due to different industrial structures). It therefore begins by computing the share of sectors in common between pairs of countries (Table 8). The fraction of localized sectors across countries is in general large for each index, especially for the E&G index. On average, countries share, respectively, 68% and 64% of the localized sectors according to E&G and D&O indexes. 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 shown in Tables 9–14, in each country the most localized sectors include, to a great extent, traditional industries such as jewellery, wine and textiles. Moreover, Tables 9–14 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, thirteen industries appear in the list of localized sectors in all countries considered, while this number reduces to eight if the D&O index is considered. Among these sectors, four are in common between the two indexes and belong to the publishing and printing sector group (NACE 2211, 2213, 2215 and 2222).[9]

In order to interpret these results better, a cross-sectoral investigation of all localized industries was performed by employing a taxonomy that classifies industries in macro-groups composed of sectors with relatively homogeneous characteristics. More specifically, Pavitt's taxonomy (PAVITT, 1984) is employed here. It is one of the first (and most widely used) classification frameworks proposed in the industrial organization literature. Pavitt's taxonomy classifies industries in categories by considering different indicators, which

Table 8. Share of localized sectors in common between pairs of countries

	France	Germany	Italy	Spain	UK
E&G index					
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 index					
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

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 number of localized sectors between the two countries.

Table 9. Ten most localized NACE four-digit manufacturing sectors: Belgium

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

Table 10. Ten most localized NACE four-digit manufacturing sectors: France

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

Table 11. Ten most localized NACE four-digit manufacturing sectors: Germany

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

Table 12. Ten most localized NACE four-digit manufacturing sectors: Italy

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

account for their technological characteristics (for example, internal versus external sources of the innovation process; product/process innovation; degree of appropriability of innovations, etc.), but also for other

Table 13. Ten most localized NACE four-digit manufacturing sectors: Spain

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

Table 14. Ten most localized NACE four-digit manufacturing sectors: UK

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

industry dimensions such as type-entry barriers, average size of firms in the sectors, etc. Using such classification criteria, at least four macro-categories of industries can be identified: science based, specialized suppliers, scale intensive, and traditional. The most salient features of each Pavitt category are summarized in Table 15.

If the share of localized sectors in each Pavitt category is computed for the countries considered and for each localization index (data not shown), then both the E&G and D&O indexes suggest that localized industries mainly belong to the groups of traditional and scale-intensive sectors, while the science-based sectors (typically characterized by intense internal and external research and development activity) feature the smallest fraction of localized industries.

However, this pattern could simply reflect the numerical prevalence of traditional and scale-intensive industries in the countries under examination. Thus, the apparent weakness of localization within science-based industries could be the sheer outcome of the historical evolution of different industrial structures (OTTAVIANO, 1999). To control for such a factor, the share of localized sectors in each Pavitt category was rescaled by the share of sectors in each category.[10] A value of this new index > 1 for a specific Pavitt category will indicate the presence of localization in excess of what is predicted by the share of sectors in that category. At the other extreme, a value of the index < 1 will indicate that localization in that category is less likely than what is predicted by the share of sectors in the category.

The results of this exercise are reported in Table 16. Controlling for industrial structures implies an increase in the share of science-based localized industries. Indeed, according to both localization indexes, the share is > 1 in almost all countries (except Belgium and the UK) and the largest one, as compared with that of other Pavitt categories, in half of the cases. This indicates that in science-based sectors there seems to be a more pervasive localization effect than what is predicted by the share of sectors in total manufacturing. This is in line with previous results in the literature (in particular, see AUDRETSCH and FELDMAN, 1996), that find a positive correlation between clustering and degree of innovativeness in science-based industries.

The above results provide additional clues about the characteristics of industry localization. In particular, they indicate that the type of localized industries is invariant across countries. This supports the theoretical idea that into an integrated area such as the European Union localization forces are mainly sector specific and not influenced by country-specific factors (which, however, affect the intensity of localization forces; cf. the previous section). Moreover, this outcome is independent of the index used and therefore it is independent of the treatment of geographical space in the empirical analysis.

CONCLUSIONS

This paper empirically investigated industry-localization patterns in European Union countries. Unlike the majority of existing works in this field, it employed a firm-based dataset that is homogeneous across countries, and it computed two different localization indexes, that is, the Ellison and Glaeser (E&G) index (ELLISON and GLAESER, 1997) and the Duranton and Overman (D&O) index (DURANTON and OVERMAN, 2005). This has allowed statistically sound comparisons to be made across both countries and localization indexes, at aggregate and sector-disaggregated levels. In line with previous studies, it was found that, independently from the index used, localization is a pervasive phenomenon in all countries studied. In addition, in all countries the values of localization indexes display a relevant sectoral variability. Furthermore, a cross-sectoral analysis of localized industries has shown that, in all countries and

Table 15. *Pavitt taxonomy of manufacturing industries*

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 manufacturing (leather, jewellery)	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	Research and development (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 centres	High	Very high

Note: Adapted from PAVITT (1984).

Table 16. Share of localized sectors in each Pavitt category divided by the share of sectors in each Pavitt category

	Science based	Specialized suppliers	Scale intensive	Traditional
E&G index				
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				
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

for both indexes, '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 ones. These results confirm the theoretical prediction (for example, OTTAVIANO and THISSE, 2001) that into an economically integrated area like the European Union localization is an important industry phenomenon, and that sector-specific drivers are more important than country-specific factors in determining localization patterns. In addition, the robustness to the index employed indicates that the treatment of geographical space does not introduce biases in predictions about the number and the type of localized sectors.

On the other hand, significant cross-index differences emerge with respect to predictions on the intensity of the forces underlying localization within countries. This may reflect some heterogeneity in the way geographical firm clustering looks like in space. Some clusters might indeed map 'true agglomeration', that is, 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. In turn, this difference might reflect the role played by different types of externalities in generating observed localization patterns. Indeed, the presence of differences in the prediction of intensity across indexes for a given industry within the same country can be read as an indication of the fact the pecuniary and non-pecuniary externalities operate very differently across sectors in determining the location choices of firms. This is because non-pecuniary externalities are likely to introduce stronger spatial correlation in firms location decisions than pecuniary externalities (OTTAVIANO and THISSE, 2001; OTTAVIANO and LAMORGESE, 2003). Moreover, even more intriguing is the evidence that, at the four-digit level, agglomeration intensity varies a lot across countries for the same industry using the D&O index, but not when using the E&G index. Indeed, were these cross-country differences entirely due to low factor mobility between European countries (for example, compared with the United States), then different predictions among the two indexes in this respect would not be expected. The fact that cross-country rank-correlation values are low considering the D&O index suggests that non-pecuniary externalities play a very important role in generating national specificities in industry localization patterns. This result is in line with some recent dynamic models in the economic geography literature (for example, BOTTAZZI and DINDO, 2009), that show that path-dependent dynamics are very hard to revert when non-pecuniary externalities are important, even in the presence of low trade barriers and high factor mobility.

The present work could be extended in several ways. First, a more detailed investigation of the influence of firm size and its effect on firm location choice could be performed. In addition, the time evolution of the values of the localization indexes employed in this work could be studied (for a similar attempt on more aggregate data, see BRULHART and TRAEGER, 2005). Such an attempt could also involve a full-edged analysis of the links between industry dynamics (for example, firm entry, exit and growth), and the generation and evolution of localization phenomena (for important contributions in this direction, cf. OTTAVIANO, 1999; HOLMES and STEVENS, 2002; KLEPPER, 2002; and LAFOURCADE and MION, 2006).

Second, it would be interesting to analyse in more depth the determinants of the observed cross-country and cross-index differences in localization intensity. Indeed, the present investigation followed a standard practice in the empirical localization literature, namely, comparing index results using simple statistics such as correlations (for example, ELLISON and GLAESER, 1997; DURANTON and OVERMAN, 2005; BARRIOS *et al.*, 2003; and BRIANT *et al.*, 2010, to mention only a few). However, correlations across indexes and countries provide only the first evidence about the different role played by pecuniary versus non-pecuniary externalities, as their levels can be influenced by several industry and firms characteristics. A more thorough investigation of these aspects could involve, on one hand, the use a different index (for example, that proposed by BOTTAZZI *et al.*, 2007), which explicitly disentangles sector-specific factors from location-specific ones. On the other hand, panel-data regressions could be conducted involving the values of the indexes and proxies of the strength of pecuniary and non-pecuniary externalities.

Third, the foregoing exercises have disregarded, on a first approximation, the characteristics of the areas where firms tend to localize. In fact, the dataset could be used to study in more detail the tendency of European Union firms to locate in very urbanized areas, for example, due to the presence of services (such as financial, consulting, auditing) supporting firm activity (HENDERSON and ONO, 2007; DAVIS and HENDERSON, 2008; STRAUSS-KAHN and VIVES, 2009).

APPENDIX

The analysis presented in this paper relies on a country-homogeneous database where observational units are firms rather than plants. As discussed in the second section, the use of firm-level data rather than plant-based data may induce an upward bias in the measurement of localization. In addition, the bias is likely to be increasing with the average size of firms in the sample, due to the positive correlation between a firm's size and the number of a firm's plants.

Nevertheless, two types of arguments seem to downplay the importance of the bias induced by multi-plant firms. First, recent empirical studies (for example, DURANTON and OVERMAN, 2008) indicate that in many cases firms' establishments and headquarters are closely located in space. Second, a careful scrutiny of the indexes and their characteristics suggests that the direction of the bias is rather unclear. On the one hand, the concentration of production units in the headquarters' area could indeed induce an upward-biased estimation of localization. On the other hand, moving from plants to firms also implies an increase in industrial concentration for the sector under study. The latter point can be better grasped by noticing that the Herfindahl index entering the E&G formula (1) 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 the results to such biases, two different sets of analyses were run. First, localization levels predicted by the two indexes were studied by conditioning to firm size. The assumption is that multi-plant firms are mainly firms of medium/large size. Thus, knowing how localization indexes 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, a simulation analysis of localization patterns on samples built was performed by artificially disaggregating medium-to-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, the sample was initially partitioned into two classes of firms (small versus big) employing as a size threshold the median of the industry-pooled size distribution in each country. Then, for each country and sector, the E&G and D&O indexes were computed separately for each size class, in such a way to identify the size-conditioned shares of localized and dispersed industries. Table A1 reports the results of this exercise. Notice that such results are, to a large extent, robust to different criteria for choosing the size threshold (for example, exogenously determined using the procedure employed by LAFOURCADE and MION, 2006). A comparison of size-conditioned figures with those obtained in the unconditioned analysis (cf. Table 3) seems to indicate that the effects of the firm–plant bias are quite negligible in the sample. Indeed, splitting the sample into small and large 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 the 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 with the largest median employment in the sample (for example, the UK) is comparable with those in countries with smaller median firm sizes (for example, Spain and France).

To verify further that the firm–plant bias is not that relevant in the sample, a simulation analysis of localization patterns was also performed by artificially

Table A1. Small versus large firms: share of sectors localized and dispersed in each country

	E&G index			D&O index	
	LOC[a]	LOC[b]	DISP	LOC	DISP
Small firms country					
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 country					
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

Notes: Small firms: firms with below-median employment. Large firms: firms with above-median employment. The median is computed on the industry-pooled within-country employee distribution.

[a]Share of sectors with a strictly positive E&G index.

[b]Share of sectors significantly localized according to the 2-sigma rule.

disaggregating the empirically observed firms of the sample (in each industry and country) into fixed-size plants and reallocating such plants in the geographical space. To do so, it was first assumed that the expected number of plant for a given firm increases with its size. Following LAFOURCADE and MION (2006), it was 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 simply equal to $p = s/20$ (rounded to the closest integer). Second, alternative theoretical hypotheses about the way these artificially generated plants can be spatially reallocated in the space were set up. 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 L existing locations so that the probability p_l that a plant is in a location l is independent and identically distributed and equal to

$$p_l = \frac{s_{il}}{\sum_h s_{ih}},$$

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

$$i\,\Big(\sum_h s_{ih}\Big).$$

In this way, this paper tries to reproduce the geographical distribution of small firms that, under the assumption, is in the sample the best proxy to the actual (but unobservable) spatial distribution of plants.

Under either scenarios, several independent replications ($R = 1000$) of the above procedure were performed, where in each replication all firms in the database were disaggregated and reallocated. The share of sectors that in each country turned to be localized in *all R* simulations were thus computed. Given the huge computational requirements that the D&O index places on this kind of simulation analysis (especially in countries with a huge number of firm observations), results are presented for the E&G index only (Table A2).[11] The simulation analysis seems 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 the UK), and larger than the fraction of localized sectors computed on actual data (cf. Table 3). All that hints to a downward (rather than upward) sample bias associated with considering firms instead of plants as observational units.

Table A2. *Simulation analysis of localization with artificially generated plants (E&G index)*

	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

Note: Share of sectors localized in all $R = 1000$ independent runs. Localized sectors identified using a 2-sigma rule.

NOTES

1. Moreover, despite the D&O index having clear theoretical advantages with respect to the E&G index, it is nonetheless less parsimonious in terms of data requirements and computation (see also the third section). Thus, in the absence of detailed data about firm location, the E&G index can be used as a good first-order approximation of the measurement of localization, at least for the dimensions that have been found to be robust to the index used.
2. Table 1 also reveals that censoring the sample to sectors with more than ten active firms does not alter the dimension of the sample. Indeed, the number of sectors covered is always close to the total number of sectors available in each country.
3. Note, however, that the analysis presented in the following sections shows that the share of localized sectors in Italy is comparable with that obtained in other studies (for example, LAFOURCADE and MION, 2006), conducted on more representative samples. Moreover, the Appendix shows that such a result is robust to the artificial disaggregation of firms into plants (for more details, see the Appendix).
4. Moreover, confidentiality in firm-level data has so far hampered the development of harmonized cross-country datasets having a high coverage of the universe of manufacturing firms (for a discussion of these issues, see BARTELSMAN *et al.*, 2005). It follows that for many countries (included those in the present sample) it is very difficult to obtain estimates of the distribution of firm sizes which are representative of the one of the universe of firms.
5. The Moran Index is based on a binary contiguity matrix $\boldsymbol{W}$, with $w_{il} = 1$ if and only if (i; l) are contiguous regions, and zero otherwise; or on a spatially weighted matrix, where weights are given by a function of the great circle distance between regional centroids. In any of the two cases an *ex-ante* partition into regions is needed.
6. The density function shown in equation (5) has a domain on the whole real line. Therefore, it could return positive density estimates at negative distances. Following DURANTON and OVERMAN (2005), a 'reflection correction' is imposed by replacing negative densities with zeros

and rescaling all values so as to obtain a total probability mass equal to 1.

7. Following KLIER and MCMILLEN (2008), forty target distances were selected. Moreover, for each country all distances between zero and the median distance between all pairs of manufacturing firms were considered. The median distances (kilometres) were, respectively: 105 for Belgium, 465 for France, 355 for Germany, 275 for Italy, 470 for Spain and 220 for the UK.
8. 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 the present case (see Table 4).
9. The other industries which are part of the 'core' for the E&G index are: 'Manufacture of meat' (1513), 'Manufacture of prepared feeds for farm animals' (1571), 'Finishing of textiles' (1730), 'Manufacture of other outwear' (1822), 'Manufacture of metal structure' (2811), 'Forging, pressing, stamping and roll forming of metal' (2840), 'General mechanical engineering' (2852), 'Building and repairing of ships' (3511), and 'Manufacture of chairs and seats' (3611). The other 'core' sectors for D&O index are: 'Manufacture of perfumes and toilet preparations' (2452), 'Manufacture of taps and valves' (2913), 'Manufacture of machinery for textile, apparel and leather production' (2954), and 'Manufacture of jewellery' (3622).
10. This procedure is analogous to that underlying the construction of the Balassa Index (for example, COMBES and OVERMAN, 2004; LAFOURCADE and MION, 2006).
11. Preliminary results on the D&O index, however, do not seem to contradict E&G simulation findings.

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 FELDMAN M. P. (1996) R&D spillovers and the geography of innovation and production, *American Economic Review* **86**, 630–640.

BARRIOS S., BERTINELLI L., STROBL E. A. and TEIXEIRA A.-C. F. (2003) *Agglomeration Economics and the Location of Industries: A Comparison of Three Small Countries*. Technical Report. CORE Discussion Paper, 2003. Centre for Operations Research and Econometrics (CORE), Université Catholique de Louvain.

BARRIOS S., BERTINELLI L., STROBL E. and TEIXEIRA A.-C. (2005) The dynamics of agglomeration: evidence from Ireland and Portugal, *Journal of Urban Economics* **57**, 170–188.

BARTELSMAN E., SCARPETTA S. and SCHIVARDI F. (2005) Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries, *Industrial and Corporate Change* **14**, 365–391.

BERTINELLI L. and DECROP J. (2005) Geographical agglomeration: Ellison and Glaeser's index applied to the case of Belgium, *Regional Studies* **39(5)**, 567–583.

BOTTAZZI G. and DINDO P. (2009) An evolutionary model of firms location with technological externality, in BOSCHMA R. and MARTIN R. (Eds) *Handbook on Evolutionary Economic Geography*, pp. 508–528. Edward Elgar, Cheltenham.

BOTTAZZI G., DOSI G., FAGIOLO G. and SECCHI A. (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.

BRIANT A., COMBES P.-P. and LAFOURCADE M. (2010) Dots to boxes: do the size and shape of spatial units jeopardize economic geography estimations?, *Journal of Urban Economics* **67**, 287–302.

BRULHART M. and TRAEGER R. (2005) An account of geographic concentration patterns in Europe, *Regional Science and Urban Economics* **35**, 597–624.

COAD A. (2008) Firm growth and scaling of growth rate variance in multiplant firms, *Economics Bulletin* **12(9)**, 1–15.

COMBES P.-P. and OVERMAN H. G. (2004) The spatial distribution of economic activities in the European Union, in HENDERSON J. V. and THISSE J. F. (Eds) *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2845–2909. Elsevier-North Holland, Amsterdam.

DAVIS J. and HENDERSON J. (2008) The agglomeration of headquarters, *Regional Science and Urban Economics* **38**, 445–460.

DEVEREUX M. P., GRIFFITH R. and SIMPSON H. (2004) The geographic distribution of productivity activity in the UK, *Regional Science and Urban Economics* **34**, 533–564.

DUMAIS G., ELLISON G. and GLAESER E. (2002) Geographic concentration as a dynamic process, *Review of Economics and Statistics* **84**, 193–204.

DURANTON G. and OVERMAN H. G. (2005) Testing for localization using micro-geographic data, *Review of Economic Studies* **72(4)**, 1077–1106.

DURANTON G. and OVERMAN H. G. (2008) Exploring the detailed location patterns of UK manufacturing industries using micro-geographic data, *Journal of Regional Science* **48**, 213–243.

ELLISON G. and GLAESER E. L. (1997) Geographic concentration in US manufacturing industries: a dartboard approach, *Journal of Political Economy* **105(5)**, 889–927.

FORNAHL D. and BRENNER T. (2009) Geographic concentration of innovative activities in Germany, *Structural Change and Economic Dynamics* **20**, 163–182.

FUJITA M., KRUGMAN P. and VENABLES A. (1999) *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, MA.

HENDERSON J. and ONO Y. (2007) Where do manufacturing firms locate their headquarters?, *Journal of Urban Economics* **63**, 431–450.

Holmes T. and Stevens J. (2002) 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, *Quarterly Journal of Economics* **110**, 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 McMillen D. (2008) Evolving agglomeration in the US auto supplier industry, *Journal of Regional Science* **48**, 245–267.

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

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

Marshall A. (1920) *Principles of Economics*, 8th Edn. Macmillan, London.

Maurel F. and Sédillot B. (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 continuous stochastic phenomena, *Biometrika* **37**, 17–23.

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

Ottaviano G. and Lamorgese A. (2003) Space, factors and spillovers, in *Il commercio con l'estero e la collocazione internazionale dell'economia italiana: incontro di studio in memoria di Stefano Vona*, Atti dell'incontro tenuto a Roma il 14 dicembre 2001. Banca d'Italia, Rome.

Ottaviano G. I. and Thisse F. J. (2001) On economic geography in economic theory: increasing returns and pecuniary externalities, *Journal of Economic Geography* **1(2)**, 153–179.

Pavitt K. (1984) Sectoral patterns of technical change: towards a taxonomy and a theory, *Research Policy* **13**, 343–373.

Rosenthal S. S. and Strange W. C. (2001) The determinants of agglomeration, *Journal of Urban Economics* **50**, 191–229.

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