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

Cities and Clusters: Economy-Wide and Sector-Specific Effects in Corporate Location

Giulio Bottazzi [a] & Ugo Gragnolati [a]
[a] Institute of Economics, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, Pisa, Italy
Version of record first published: 30 Nov 2012.

To cite this article: Giulio Bottazzi & Ugo Gragnolati (2012): Cities and Clusters: Economy-Wide and Sector-Specific Effects in Corporate Location, Regional Studies, DOI:10.1080/00343404.2012.739281

To link to this article: http://dx.doi.org/10.1080/00343404.2012.739281

Regional Studies, pp. 1–17, iFirst article

Cities and Clusters: Economy-Wide and Sector-Specific Effects in Corporate Location

GIULIO BOTTAZZI and UGO GRAGNOLATI
Institute of Economics, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, Pisa, Italy.
Emails: g.bottazzi@sssup.it and u.gragnolati@sssup.it

(Received February 2011: in revised form September 2012)

BOTTAZZI G. and GRAGNOLATI U. Cities and clusters: economy-wide and sector-specific effects in corporate location, *Regional Studies*. Are the observed spatial distributions of firms decided mostly by economy-wide urbanization economies or rather by sector-specific localization economies? This paper finds that the latter kind of forces weight systematically more than the former in deciding firm location. The analysis uses Italian data on a variety of manufacturing and service sectors spatially disaggregated at the level of local labour market areas.

Firm location   Sector-specific localization economies   Urbanization economies   Maximum likelihood estimation   Model selection

BOTTAZZI G. and GRAGNOLATI U. 城市与集群：企业区位中的整体经济和特定部门效应，*区域研究*。我们所观察到的企业空间分部，究竟绝大部分是受到整体经济层面的城市化经济所决定，抑或是受到特定部门的在地化经济所影响？本文发现，决定企业区位时，后者的影响力较前者具有更系统性的重要性。本分析运用在空间上分散在意大利在地劳动市场区域层级的各种制造业与服务业数据。

企业区位   特定部门的在地化经济   城市化经济   最大可能性评估   模型选择

BOTTAZZI G. et GRAGNOLATI U. Les grandes villes et les clusters: les effets à la fois à l'échelle économique et spécifiques à chaque secteur dans le choix d'emplacement des entreprises, *Regional Studies*. Les distributions géographiques observées des entreprises, sont-elles déterminées dans une large mesure par des économies d'urbanisation à l'échelle économique ou plutôt par des économies de localisation spécifiques à chaque secteur? Cet article trouve que cette dernière catégorie de forces motrices pèse systématiquement plus lourd que ne le fait la première dans la prise de décision quant au choix d'emplacement des entreprises. L'analyse emploie des données italiennes sur un éventail de secteurs manufacturier et des services géographiquement désagrégées au niveau des bassins d'emploi locaux.

Choix d'emplacement des entreprises   Économies de localisation spécifiques à chaque secteur   Économies d'urbanisation   Estimation du maximum de vraisemblance   Choix d'un modèle

BOTTAZZI G. und GRAGNOLATI U. Städte und Cluster: wirtschaftsweite und sektorspezifische Auswirkungen auf den Firmenstandort, *Regional Studies*. Werden die beobachteten räumlichen Verteilungen von Firmen vor allem durch wirtschaftsweite Urbanisierungsökonomien oder eher durch sektorspezifische Lokalisierungsökonomien bestimmt? In diesem Beitrag stellen wir fest, dass die zweite Art von Kräften bei der Festlegung eines Firmenstandorts systematisch stärker ins Gewicht fällt als die erste. Für unsere Analyse kommen italienische Daten über ein breites Spektrum von produzierenden und Dienstleistungsbranchen zum Einsatz, die auf der Ebene der lokalen Arbeitsmarktbereiche räumlich disaggregiert werden.

Firmenstandort   Sektorspezifische Lokalisierungsökonomien   Urbanisierungsökonomien   Maximale Wahrscheinlichkeitsschätzung   Modellauswahl

BOTTAZZI G. y GRAGNOLATI U. Ciudades y aglomeraciones: efectos en toda la economía y sectores específicos del emplazamiento corporativo, *Regional Studies*. ¿Se deciden las distribuciones espaciales observadas de las empresas sobre todo en función de toda la economía de urbanización o más bien de las economías de localización específicas a cada sector? En este artículo comprobamos que el segundo tipo de fuerzas tiene más importancia sistemática que el primero a la hora de decidir la ubicación de una empresa. En el análisis utilizamos datos italianos sobre una serie de sectores manufactureros y de servicios desglosados espacialmente en las áreas del mercado laboral de ámbito local.

Ubicación de empresas   Economías de localización por sectores   Economías de urbanización   Estimación de la probabilidad máxima   Selección de modelo

JEL classifications: C31, L1, R3

0034-3404 print/1360-0591 online/12/000001-17 © 2012 Regional Studies Association   http://dx.doi.org/10.1080/00343404.2012.739281
http://www.regionalstudies.org

INTRODUCTION

The factors leading to the formation of economic agglomerates have been identified since MARSHALL (1890) with the pecuniary advantages that clustering provides, thanks to deeper intermediary markets and larger pools of labour suppliers. The advantages stemming from the increased availability of cheaper production factors are plausibly reinforced by the demand-side effect that the simple presence of more consumers is likely to create. Together, cost-saving and revenue-enhancing pecuniary factors generate a competitive advantage for local firms via market-mediated interactions. Along these lines, in the context of the relatively recent 'New Economic Geography' models, the agglomerative pull has been typically identified with the feedback mechanism between local demand and local labour supply (KRUGMAN, 1991). In that setting, when households move to a new location reacting to differentials in labour demand, they increase the consumption level of the location they move into and, consequently, the local demand faced by *all* final good sectors. Therefore, the resulting locational effect is economy wide rather than sector specific. In this sense, the accumulation of population and demand is by all means an urbanization effect: it explains the growth of industrially diversified cities rather than the formation of specialized clusters.

Yet, several sectors of the economy show a considerable agglomeration of firms in places that are neither densely populated nor particularly well connected to metropolitan areas. These specialized clusters seem to escape the explanation provided by the pecuniary effects of demand and supply: in fact, the specific sectoral production within an industrial cluster exceeds local demand by far. In these cases, agglomeration plausibly results from non-pecuniary factors, which generate a locational advantage for firms through economic relations that escape market exchange. Typically, these advantages are identified with the existence of positive externalities produced by the flow of sector-specific technological knowledge (MARSHALL, 1890). In this view, some of the private productive knowledge incorporated in firms may spill into a knowledge pool which is available only locally. The geographic scope of such a pool is limited by the fact that much of its content is scarcely codified and thus hard to transfer in space. Hence, firms with similar productive processes need to locate near one another in order to benefit from the common knowledge pool, which in turn will be greater where more firms are settled. Therefore, these technological externalities engender automatically a self-reinforcing mechanism leading to agglomeration. Notably, though, the presence of firms with a certain knowledge base may be technologically and locationally relevant only for firms acting within the same knowledge range, while it does not generate any relevant technological spillover for most of the other firms in the economy. In this sense, Marshallian localization economies are sector specific rather than economy wide, thus being able to explain the formation of specialized clusters rather than the presence of diversified cities. Partly in contrast with this tradition, JACOBS (1969) regards technological spillovers as being most relevant across sectors, so that productivity and innovation are most favoured in cities, where the variety of the economic fabric is supposed to be greater.[1]

The present work investigates the strength of urbanization economies relative to sector-specific localization economies in shaping the observed geographical distribution of manufacturing and service activities in Italy. The pull of urbanization, plausibly being the outcome of pecuniary, market-mediated interactions, is expected to act *across* the different economic sectors, albeit with possibly varying strength. Conversely, localization economies stemming from technological and organizational specificities are expected to act *within* each sector separately. On the ground of these considerations, the observed spatial distributions of firms in each sector are compared with the ones predicted by the discrete choice model of BOTTAZZI and SECCHI (2007), thus allowing one to assess the relative strength of the various locational drivers, common or sector specific, on the attractiveness of a location. As such, while being similar in scope to other approaches in the literature (DESMET and FAFCHAMPS, 2006; DEVEREUX *et al.*, 2004; DUMAIS *et al.*, 2002; ELLISON and GLAESER, 1997, 1999; MAUREL and SÉDILLOT, 1999; OVERMAN and DURANTON, 2001), this work differs from them by producing not only estimates of the relevant locational parameters, but also a prediction of the entire spatial distribution of firms in each sector. This feature constitutes a commonality with other works by BOTTAZZI *et al.* (2007, 2008). However, the present paper improves upon them in two directions. First, it uses maximum likelihood methods to obtain point estimates of the parameters of interest and bootstrap resampling to estimate the variance and thus the statistical significance of the different marginal effects: in this way, the magnitudes of the various locational drivers can be unambiguously interpreted and directly compared. Second, the analysis is extended to include a large set of regressors to control for a multiplicity of determinants, thus allowing a more thorough understanding of the locational process.

The main result presented here regards the role played by market-mediated urbanization economies relative to technology-related localization economies. Firm location is found to be affected by both, but the weight of the former is systematically lower that the weight of the latter. These findings may be taken to imply that any attempt to explain the spatial structure of economic activities cannot rescind from considering the technological and organizational dynamics internal to each sector, since these are the major factors deciding the geography of corporate location.

The rest of the paper is structured as follows. The second section presents the database for which the analysis is carried out together with the choices that were operated on the data. The third section introduces some descriptive statistics as well as some hints regarding the distinct effects of urbanization and industrial clustering on spatial distributions. The fourth section sketches the discrete choice model on which theoretical predictions are based and estimates the marginal effects of the various factors identified as possible explanators of the attractiveness of locations. The fifth section summarizes and discusses the relevant results.

DATA

The present analysis will attempt to explain the spatial distribution of economic activities in manufacturing and service sectors in Italy for the year 2001. The census data that will be used here account approximately for 3.5 million plants and 13.8 million employees, which amount to 71.2% of employment in the Italian economy.

At the sectoral level, the universe of investigation is constituted by two-digit ATECO-NACE[2] sectors in their 2002 definition, with the exception of sector '36 – Furniture and other manufacturing activities'. This sector is further disaggregated at a three-digit level in order to capture more accurately what the 'other' activities are, namely: '361 – Furniture', '362 – Jewellery', and '363 – Musical instruments', plus the residual '36R – Residual of Sector 36'. With this disaggregation it is possible to characterize more sharply some industrial districts, which constitute an object of special interest for the present analysis and fall precisely within the 'other' activities of Sector 36. Conversely, this kind of disaggregation was not applied to the residual sector in services, that is '74 – Other business activities': in fact, despite being residual, such a sector is well characterized in productive terms as it collects mostly professionals. Across all sectors, only those business units classified as firms are taken into consideration, while non-profit and governmental units are neglected. Table 1 reports the list of sectors under analysis with some summary statistics that will be discussed below.

At the geographic level, observations are distributed across 686 spatial units identified as local labour market areas (LLMAs) in their 2001 definition (Istituto Nazionale di Statistica (ISTAT), 1997, 2001). Basically, an LLMA aggregates the 'peripheral' municipalities from which workers commute with their common 'core' municipality, which receives the commuters flows; in this sense, LLMAs are not a mere administrative repartition, but rather a socio-economic one. As such, they tend to preserve much more accurately the spatial continuity of economic phenomena, such as agglomeration, which are central to the present analysis (Sforzi, 2000). Fig. 1(a) gives an idea of the spatial extension of LLMAs.

Given these choices about the sectoral and geographical disaggregation of the data, the present analysis will attempt to explain the spatial distribution of plants through several independent variables. More specifically, the econometric exercise of the fifth section will characterize the attractiveness of each LLMA according to the local value of the following variables: (1) resident population; (2) consumption per capita; (3) average labour productivity; (4) distance from transport infrastructures; (5) local productive variety; (6) a binary variable for metropolitan areas; and (7) a binary variable for industrial districts. The data on plants, population, employees, value added, and binary variables for metropolitan areas and industrial districts come from the census of manufactures and services and from the census of population and housing, as collected by the Italian National Institute of Statistics (ISTAT, 2006). In particular, value added divided by the number of employees defines average labour productivity by LLMA. Consumption data are taken from the statistical report of the chambers of commerce, but they underwent partial elaboration as they had to be aggregated by LLMA (Unioncamere, 2009) (see below). Finally, the data on productive variety and distances from transport infrastructures result from the authors' own elaborations. Some of these data deserve further description.

The consumption data used here constitute a disaggregation by LLMA of the data that were originally produced at the province level by Unioncamere (2009). This disaggregation is obtained by assigning to each municipality a weight in terms of the population share it detains with respect to the total population of its pertaining province. Then, the province-level data on consumption are allotted to each municipality proportionally to their population weight. Finally, municipality-level data on consumption are aggregated by LLMA, thus obtaining total consumption by LLMA. Dividing this value by the population of each LLMA yields the level of per capita consumption by LLMA, as mapped in Fig. 1(a).

The data regarding the distance of each LLMA from transport infrastructures are obtained through geographical information system (GIS) elaborations. For a given infrastructure, say airports, a map is laid down identifying the presence of such infrastructure over the Italian territory. The distance of each LLMA from the infrastructure is then measured as the Euclidean distance of the centroid of the LLMA from the closest airport. This operation is repeated for five different kinds of transport infrastructures: commercial airports, commercial seaports, toll highways, state routes and railways (Fig. 1(b)). As a result, five different vectors of distances are obtained. These can either be used as five separate variables, or they can be reduced to a single index expressing the average distance of each LLMA from transport infrastructures.

Also, the spatial distribution of labour productivity deserves some comment. As displayed in Fig. 1(c), when

Table 1. Summary statistics

ATECO-NACE sector	N	W	F	S_{max}	G	D
15 – Food products	73 680	443 930	0.999	0.025	0.575	7
17 – Textile	31 984	309 487	0.904	0.178	0.834	17
18 – Apparel	46 377	298 241	0.950	0.048	0.780	28
19 – Leather products	24 195	206 035	0.701	0.097	0.904	20
20 – Wood processing	50 250	179 313	0.999	0.036	0.574	5
21 – Pulp and paper	5175	84 212	0.614	0.100	0.819	4
22 – Publishing and printing	29 166	173 431	0.940	0.146	0.802	0
23 – Coke, petroleum and nuclear fuel	913	24 537	0.370	0.051	0.835	0
24 – Organic and inorganic chemicals	7721	205 153	0.720	0.163	0.814	1
25 – Rubber and plastic products	15 115	216 876	0.757	0.097	0.804	3
26 – Non-metallic mineral products	31 177	253 664	0.988	0.030	0.640	11
27 – Basic metals	3984	139 287	0.548	0.096	0.841	0
28 – Fabricated metal products	102 295	700 984	0.999	0.060	0.711	13
29 – Industrial machinery	46 481	597 544	0.927	0.084	0.793	9
30 – Office machinery	1715	19 257	0.442	0.104	0.847	1
31 – Electrical machinery	20 282	211 404	0.832	0.115	0.808	1
32 – Radio, television and telecommunication devices	9677	107 578	0.821	0.099	0.767	2
33 – Precision instruments	26 244	126 004	0.907	0.090	0.780	10
34 – Motor vehicles and trailers	2229	172 932	0.402	0.177	0.868	0
35 – Other transport equipment	4951	103 096	0.555	0.045	0.831	0
361 – Furniture	35 784	209 188	0.943	0.094	0.797	18
362 – Jewellery	10 906	50 232	0.716	0.132	0.879	5
363 – Musical instruments	695	2740	0.241	0.180	0.916	1
36R – Residual of Sector 36	6728	39 233	0.690	0.093	0.814	0
40 – Electricity and gas	4159	109 047	0.885	0.053	0.619	0
41 – Water	1408	15 961	0.571	0.043	0.725	0
45 – Construction	529 757	1 528 629	1.000	0.048	0.652	0
50 – Sale and services of motor vehicles	164 079	457 527	1.000	0.058	0.645	0
51 – Wholesale and commission trade	404 278	1 021 666	0.999	0.083	0.776	0
52 – Retail trade	772 730	1 675 275	1.000	0.059	0.638	0
55 – Hotels and restaurants	261 304	853 122	1.000	0.056	0.616	0
60 – Land transport	135 135	531 539	1.000	0.077	0.695	0
61 – Water transport	1319	20 394	0.187	0.434	0.960	0
62 – Air transport	457	24 973	0.131	0.260	0.961	0
63 – Auxiliary transport activities	33 765	322 071	0.946	0.109	0.803	0
64 – Post and telecommunications	18 056	289 518	1.000	0.044	0.575	0
65 – Financial intermediation	30 587	392 870	1.000	0.071	0.679	0
66 – Private insurance and pensions	1771	40 591	0.465	0.106	0.856	0
67 – Auxiliary financial activities	84 677	154 227	0.994	0.075	0.750	0
70 – Real estate activities	149 990	226 736	0.926	0.142	0.840	0
71 – Renting of machinery and equipment	13 291	29 536	0.879	0.083	0.726	0
72 – Computer and related activities	84 100	354 847	0.987	0.138	0.808	0
74 – Other business activities	216 883	904 234	1.000	0.142	0.788	0
Mean	81 290	321 561	0.789	0.103	0.771	3.628
Standard deviation (SD)	153 224	378 379	0.254	0.071	0.101	6.608
Total	3 495 470	13 827 120	–	–	–	156

Note: Variables: N, number of plants; W, number of employees; F, fraction of local labour market areas (LLMAs) hosting at least one plant of the sector; S_{max}, maximum share of plants in a single LLMA; G, Gini coefficient corrected by sample size; and D, number of industrial districts per sector.

productivity is measured by dividing the total value added of each location by the number of employees it hosts, the map that comes out is rather surprising. Regions such as Calabria, Sicily and Sardinia have a higher productivity of labour relative to the rest of the country: this is surprising because those regions are known for being among the poorest in Italy (ISTAT, 2006). Such a strangeness can possibly be explained by the share of employment working in the public sector. Because public wages are on average higher than private wages, the fact that regions such as Calabria, Sicily and Sardinia are characterized by the presence of an above-average share of employment in the public sector may help one make sense of the map displayed in Fig. 1(c) (GIORDANO, 2009). Moreover, when dividing value added by the number of employees, the relatively higher rates of unemployment recorded in those regions may further contribute to diminishing the denominator of the fraction and produce higher values.

Finally, the measurement of local productive variety aims at quantifying how much variety exists in terms of economic sectors within each commuting zone. Such an

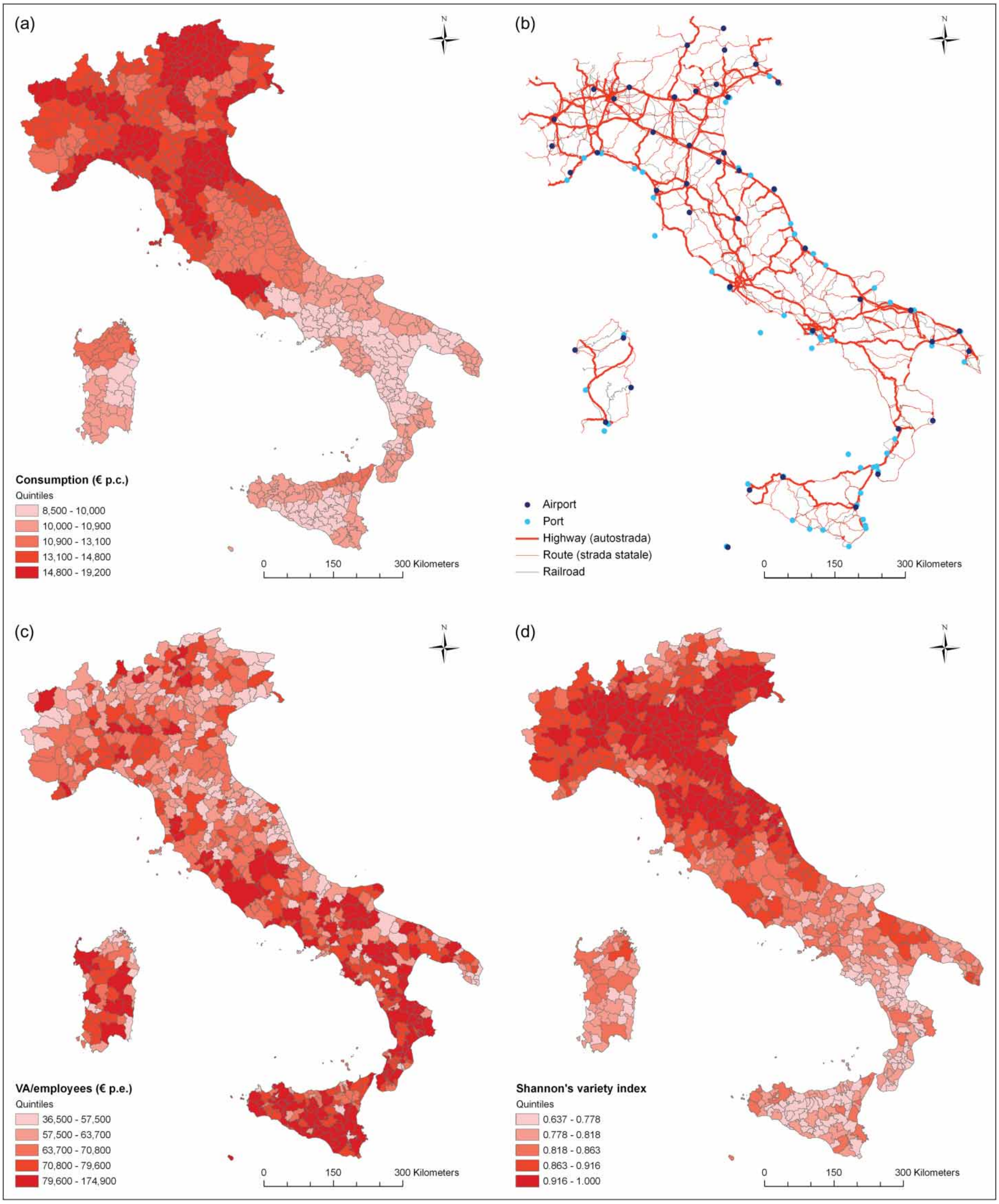

Fig. 1. *Mapping of consumption, transport infrastructures, labour productivity and industrial variety*

aim is here pursued through a variety index based on Shannon entropy, namely:

$$V_l = \sum_i s_{l,i} \log\left(1/s_{l,i}\right)$$

where $s_{l,i}$ is the share of the ith sector in location l. Indeed, V_l increases as the number of locally represented sectors increases, but it also increases with the evenness of the distribution among sectors. Given this definition of variety by LLMA, it is worth looking at the spatial distribution of V. Interestingly, bigger metropolitan areas turn out *not* to be associated with the highest values of the Shannon variety index, which instead peaks in relatively less urbanized locations (Fig. 1(d)). For example, the LLMA associated with Rome does not belong to the highest quintile of the distribution of V, while the LLMA associated with Bassano del Grappa (a town with about 40 000 residents) represents the maximum of the distribution. This tendency appears to be robust with respect to the way in which variety is measured. On the one hand, when employees rather than plants are used to compute the Shannon index in Italy, the measurement displays an 80% correlation with the values of the index obtained by using plants. On the other hand, other variety indexes end up by not identifying the biggest metropolitan areas as being among the most varied; for instance, the Simpson index of variety shows a 67% correlation with the Shannon index and it still leaves out Rome from the highest quintile. This fact is unexpected because bigger cities are often thought to host 'more of everything'. Evidently, that is the case as far as the sheer number of sectors is considered; yet, as soon as the evenness of sectoral shares is taken into account (such as in the Shannon or Simpson indexes), bigger metropolitan areas turn out to be characterized by more uneven distributions as compared with less urbanized locations.

EMPIRICAL ANALYSIS

The present analysis spans economic sectors which are strongly heterogeneous over several dimensions, as summarized in Table 1. The productive scale of firms constitutes a primary dimension of heterogeneity, as shown by the widely varying values of the total number of plants, N, and workers, W, in each sector. For example, the dataset ranges from a sector such as '30 – Office machine', where about 20 000 employees are spread among 1700 plants, to a sector such as '34 – Motor vehicles and trailers', in which about 2200 plants employ almost 173 000 workers. Moreover, other than in their productive scale, sectors differ also with respect to their spatial distributions in at least two ways. Firstly, the fraction F of locations hosting at least one firm in the sector reveals that some sectors are spread literally everywhere (for example, $F \approx 1$ in '15 – Food products'), while others are present only in some locations (for example, $F << 1$ in '27 – Basic metals'). Secondly, the maximum share of business units found in a location, S_{max}, reveals that the scale of locational effects is greater in some sectors than in others. For example, the biggest agglomeration in '17 – Textile' captures a share of the plants that is almost thrice the share captured by the biggest agglomeration in '18 – Apparel'. These various dimensions of heterogeneity in firm location end up producing geographical structures that vary across sectors as much as exemplified in Fig. 2.

As mentioned in the Introduction, the observed heterogeneity in the spatial distributions of firms possibly relates to the varying weights held by urbanization and localization economies across the various sectors. Both effects contribute to increase the overall degree of spatial concentration, yet each of them is likely to play a different role. This intuition is confirmed by verifying the correlation between different measures of

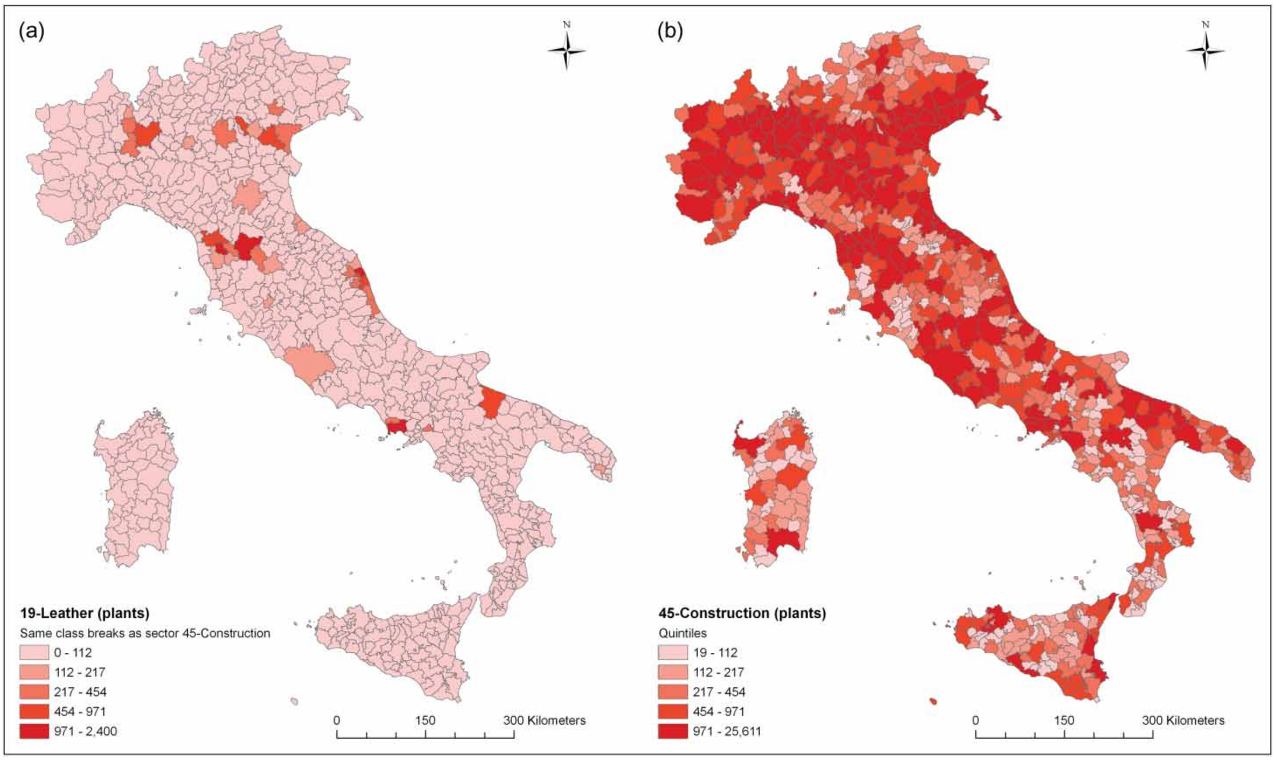

Fig. 2. Spatial distributions of plants across local labour market areas (LLMAs) in two exemplifying sectors

agglomeration, each related to distinct drivers of firm location: a weak correlation between the different measures would signal that intrinsically different phenomena are being measured. For example, it is interesting to analyse the correlation between the Gini coefficient, G, and the number of industrial districts, D. Higher values of both measures are associated with a stronger tendency to cluster; however, while G captures the overall degree of spatial concentration, D identifies only those clusters characterized by a majority of small-to-medium-sized manufacturing enterprises typically established in non-metropolitan locations (SFORZI, 2001). For this reason, G measures the joint agglomerative effect of various determinants of firm location, while, by definition, D identifies mainly those agglomerations driven by localization economies. Interestingly, running a rank correlation between G and D yields a low and weakly significant Spearman index, ρ. For instance, when only manufacturing sectors are considered, one obtains $\rho = 0.387$, with $p = 0.063$, which signals the absence of statistical significance at a 95% confidence level.[3] Although this simple finding does not bear any strong implication per se, it suggests that the spatial distributions of firms are driven by distinct determinants, whose individual effect is sufficiently big as to be detected in the data. Then, such determinants need to be qualified precisely and their effects disentangled.

A MODEL OF FIRM LOCATION

The present analysis is based on the discrete choice model described in BOTTAZZI and SECCHI (2007). The model posits a fixed number of firms in the sector and derives their equilibrium distribution in space assuming a sequence of separated profit-maximizing locational choices. The assumption of a fixed number of firms is consistent with the observation that the annual variation in the number of firms in any sector is, in general, one order of magnitude smaller than their gross entry/exit rate. Therefore, the sectoral dynamics in space, if one is merely interested in the aggregate behaviour, is mainly a 'reallocation' of activities. Given these considerations, the basic mechanism of the model is very simple.

Consider a single sector j composed of N firms. At each time step one firm is randomly chosen to die and make room for a new entrant, which will have to choose one location l among the L available. Each location is characterized in terms of its individual attractiveness, which is proportional to the expected profitability that the firm will face by locating there. The attractiveness of the location is composed of two non-negative terms. The first, a_l, is constant with respect to the number of firms but variable across location; hence, it captures the 'fixed' advantages offered by the location (higher demand, lower marginal costs, better infrastructures, etc.). Conversely, the second term, bn_l, is proportional to the number of firms of the sector, n_l, that are already established in the location. The coefficient b is constant in space. This term captures the presence of positive externalities of localization which generate and act within the sector, rather than in single locations. Hence, due to the heterogeneous preferences of firms (for more details, see BOTTAZZI and SECCHI, 2007), the entrant's probability of locating in l can be expressed as:

$$p_l \sim a_l + bn_l \tag{1}$$

Bottazzi and Secchi found that the equilibrium distribution of firms across locations:

$$n = (n_1, \ldots, n_L)$$

implied by the previous model was generally characterized by the Polya form:

$$\pi(n; a, b) = \frac{N!\Gamma(A/b)}{\Gamma(A/b + N)} \prod_{l=1}^{L} \frac{1}{n_l!} \frac{\Gamma(a_l/b + n_l)}{\Gamma(a_l/b)} \tag{2}$$

where:

$$a = (a_l, \ldots, a_L)$$

is the geographic attractiveness of the L locations; and:

$$A = \sum_l a_l$$

In the specific case of null externalities (that is $b = 0$), the equilibrium distribution of firms across locations, $n = (n_1, \ldots, n_L)$, takes the multinomial form:

$$\pi(n; a, b = 0) = N! \prod_{l=1}^{L} \frac{1}{n_l!} \left(\frac{a_l}{A}\right)^{n_l} \tag{3}$$

Equations (2) and (3) represent short-run equilibrium distributions that depend on the geographic attractiveness of the different locations. If the model were correct, the median of firms observed in each location l would fluctuate around a_l/A, while the amplitude of the fluctuations would decrease with the strength of the externality parameter b.

Maximum likelihood estimation and marginal effects

Assume that for each location a set of H variables:

$$x_l = (x_l^1, \ldots, x_l^H)$$

is known. These variables represent location-specific characteristics, such as the geographic position of the location, the local population or the average local wealth. The idea is to use these variables to model the geographic attractiveness of the different locations.

In the simpler case of null externalities, the equilibrium distribution (3) depends on a_l but not on b. Hence, a_l can be written directly as a generic function of the H regressors:

$$a_l = c(\theta, x_l)$$

depending on a set of parameters θ to be estimated. With simple substitution, the log-likelihood of the observed distribution of firms across locations as a function of the parameters reads:

$$\log \pi = \log N! - \sum_{l=1}^{L} \log(n_l!) + \sum_{l=1}^{L} n_l\big(\log c(\theta, x_l) - \log A\big) \tag{4}$$

where:

$$A = \sum_{l=1}^{L} c(\theta, x_l)$$

One can maximize the previous expression and obtain maximum likelihood point estimates for the parameters $\hat{\theta}$. In turn, these estimates define a geographic attractiveness coefficient:

$$\hat{a}_l = c(\hat{\theta}, x_l)$$

for each location; and:

$$\hat{A} = \sum_{l=1}^{L} \hat{a}_l$$

When positive externalities are allowed for, the equilibrium distribution (2) is a function of both a_l and b. However, they appear only as the fraction a_l/b. Therefore, a generic functional specification of the model can be obtained by setting:

$$a_l/b = c(\theta, x_l)$$

The log-likelihood of the observed distribution of firms then reads:

$$\log \pi = \log N! - \sum_{l} \log n_l! + \sum_{l=1}^{L} \sum_{k=0}^{n_l-1} \log\left(c(\theta, x_l) + k\right) - \sum_{k=0}^{N-1} \log(C + k) \tag{5}$$

where:

$$C = \sum_{l=1}^{L} c(\theta, x_l)$$

Again, one can maximize the previous expression and obtain maximum likelihood point estimates for the parameters $\hat{\theta}$ and the coefficients:

$$\hat{c}_l = c(\hat{\theta}, x_l) \text{ and}$$

$$\hat{C} = \sum_{l=1}^{L} \hat{c}_l$$

Assessing the effect of the H regressors and of the externality coefficient b through the estimation of the parameters $\hat{\theta}$ in the expression for c can be problematic. This is particularly true in the Polya case, where the dependence of the likelihood on the ratio a_l/b automatically induces parametric redundancy in the model. Therefore, the present analysis will characterize the impact of the different covariates using the notion of marginal effect, as often done in discrete models.[4]

To begin with, notice that once the estimates $\hat{\theta}$ are obtained, one has an estimate of $\hat{a}_l$ (in the multinomial case) or of the fraction $\hat{c}_l$ (in the Polya case). According to the original definition of the model, the probability of choosing a location l is:

$$p_l = \frac{\hat{a}_l}{\hat{A}} \tag{6}$$

in the multinomial case; and:

$$p_l = \frac{\hat{b}n_l + \hat{a}_l}{\hat{b}N + \hat{A}} = \frac{n_l + \hat{c}_l}{N + \hat{C}} \tag{7}$$

in the Polya case.

Since p_l is bounded in [0, 1] when the probability is closer to the upper bound there is less room for a further increase. An unbounded measure, which captures the total attractiveness of the location l, can be obtained by considering:

$$q_l = -\log(1 - p_l)$$

The attractiveness q_l is unbound from above and it increases with the probability p_l. In terms of this variable, the marginal effects become:

$$\frac{\partial q}{\partial \log x^h} = \sum_{l=1}^{L} \frac{\partial q_l}{\partial \log x_l^h} = \sum_{l=1}^{L} \frac{x_l^h \partial_h a_l}{A} \tag{8}$$

for the multinomial model and:

$$\frac{\partial q}{\partial \log x^h} = \sum_{l=1}^{L} \frac{\partial q_l}{\partial \log x_l^h} = \sum_{l=1}^{L} \frac{x_l^h \partial_h c_l}{N + C} \tag{9}$$

for the Polya model. For the latter case, the elasticity of q_l to the location of an additional firm reads:

$$\frac{\partial q}{\partial \log n} = \sum_{l=1}^{L} \frac{\partial q_l}{\partial \log n_l} = \frac{L}{N + C} \tag{10}$$

Notice that this expression does not depend explicitly on the functional form of c. The strength of sector-specific positive externalities depends on the total number of firms in the sector, N, and on the number of locations, L, as well as on the sum of their total attractivenesses, C. Being $C = A/b$, greater values of C are associated with a predominance of the 'fixed' attractivenesses of locations (that is, A) over the effect of sector-specific positive externalities (that is, b). Hence, the marginal effect of an additional firm decreases as C grows. In other terms, the larger the parameter C, the smaller the marginal effect of additional firms and the lesser the locational choice of a firm is actually affected by the choices of others.

A final practical warning is mandatory. In general the specification should be such that $c(\beta, x_l)$ is positive for any l. Notice that when $c \rightarrow 0^+$, the log-likelihood becomes minus infinity, so that the maximization algorithm itself should be avoided to compute the likelihood in this point. Since the analysis relies on numerical methods, however, the fulfilment of this condition is not automatically assured. Indeed, when the actual maximum is near the boundary of the positivity domain, the adopted algorithm could probe the function outside this domain, thus generating infinities. In what follows special attention will be devoted to avoid this occurrence.

Model specification

The multinomial and Polya models introduced in the previous section require the specification of a functional form which relates the values of the variables x_l in one location to its attractiveness. In this paper the following log-linear specification will be used:

$$c(\beta, x_l) = \exp\left(\sum_{h=1}^{H} \beta_h \log(x_{h,l}) + \beta_0\right) \qquad (11)$$

In the multinomial case, it is assumed that $\beta_0 = 0$. This is because in this case the log-likelihood (4) is invariant for a rescaling factor, that is, the transformation $c_l \rightarrow \lambda c_l$ applied to each c_l leaves the likelihood level invariant. Consequently, leaving β_0 to be estimated would result in an over-specified model.

The specification in (11) is equivalent to the Cobb–Douglas functional form used in consumer and producer theory. Indeed, from the previous expression one has:

$$c(\beta, x_l) = \prod_{h=1}^{H} x_{h,l}^{\beta_h} \exp(\beta_0)$$

This expression describes the attractiveness c_l as the accumulated multiplicative effect of the different variables and possesses a straightforward probabilistic interpretation. Assume that the different variables represent different aspects of the economic activity that the firm evaluates in order to choose where to locate. If, on average, the probability to choose location l according to factor h is proportional to:

$$x_{h,l}^{\beta_h}$$

and if the effects of the different factors can be assumed to be roughly independent, the combined probability of the firm to choose this location is given by expression (11). Moreover, the log-linear specification of c allows for a simple computation of the marginal effects defined in (8) and (9). Indeed, one has:

$$\frac{\partial q}{\partial \log x_h} = \hat{\beta}_h$$

for the multinomial model; and:

$$\frac{\partial q}{\partial \log x_h} = \hat{\beta}_h \frac{C}{N + C}$$

for the Polya model. Obviously, in this latter case it is also possible to compute the marginal elasticity with respect to the number of plants in the sector, n; however, as already mentioned above, such value does not depend on the functional specification, thus being given directly by equation (10).

RESULTS

It is now possible to use the methodology described so far to reconnect with the economic issues discussed above in the Introduction. What is the weight of urbanization economies relative to localization economies? How do other factors, such as infrastructural endowments, affect the locational choices of firms? These questions can be assessed by using the proper explanatory variables within the functional specification (11). Such variables will generally serve to characterize the 'fixed' advantages of each LLMA, and in particular the effect of urbanization economies. Moreover, in the context of the Polya model, their marginal effect on firm localization will be compared with the effect of the positive externalities that stand at the root of localization economies. Assuming the functional specification (11), the following seven different regressors will be taken into account to describe the 'fixed' attractiveness of each LLMA.

To begin with, the attractive pull exerted by more populated locations, that is the urbanization effect, is captured by using as a regressor the resident population (*POPULATION*). Naturally, since the demographic size of locations is likely to be positively associated with their spatial extension, it becomes necessary to introduce as a first control the area of LLMAs (*AREA*). An alternative approach would consist of using as a regressor the population density per squared kilometre. However, that method would not allow one to capture two essential size effects related to population: the revenue advantages determined by a higher local aggregate demand and the cost advantages granted by a vaster local supply of labour. Clearly, more populated areas offer advantages on both sides. Yet, in order to capture this two-fold effect just with the regressor *POPULATION*, two further controls need to be introduced: the average local level of consumption per capita (*CONSUMPTIONPC*) and the average local productivity of labour (*PRODUCTIVITY*).

On the one hand, aggregate demand depends both on the number of consumers and on their expenditure possibilities. Hence, locations with similar demographic size may express different levels of local aggregate demand due, for instance, to varying income conditions. To control for this effect, the average level of consumption per capita by LLMA is introduced as a control. Among other things, the insertion of this regressor also captures the North–South disparities shown in Fig. 1(a).

On the other hand, the cost advantages granted by a vaster local supply of labour have to be considered together with the local productivity of labour. In fact, a more productive location would need fewer employees to produce as much output as a less productive location; consequently, keeping total output fixed, the latter location would likely tend to have more residents than the former. In this sense, the cost advantage associated with a greater local availability of labour supply may be compensated by the cost disadvantage determined by a lower local productivity of labour. To control for this effect, the average productivity of labour in the entire local economy (and not in the single sector) is used as a regressor.

A further element to take in consideration with respect to the 'fixed' attractiveness of each LLMA is its infrastructural endowment, particularly for what concerns transport infrastructures. The accessibility of transportation means presumably affecting the locational choices of firms by influencing both the ease of input provisions and the facility to deliver finished products outside the local market. In this sense, transport infrastructures may determine the geographic range on which firms operate. Moreover, infrastructures can be regarded as one of the channels through which urbanization economies unfold. In fact, the cost of infrastructural services declines with the number of people and firms that use them, due to the relevant installation costs that are necessary to provide these services. As a consequence, cities can possibly provide infrastructural services for a lower cost relative to non-urban areas. The endowment in transport infrastructures is measured here with the average distance (*INFDISTANCE*) of each LLMA from five different kinds of facilities: commercial airports, commercial seaports, toll highways, state routes and railways. Being a metrical distance, a statistically significant and negative value of the marginal elasticity of this regressor indicates that an increment of the proximity to transport infrastructures increases the probability of an LLMA to attract firm location, and vice versa.

As mentioned in the Introduction, Jacobs sees the attractive pull of cities as also being determined by their supposed ability to host a more diversified economic fabric. According to this view, while industrial clusters favour localization economies among firms operating in a strictly similar productive range, cities would allow positive externalities to flow across different industries and enhance their productivity in the long run. Hence, a more varied economic fabric is expected to generate more inter-sectoral externalities, which would then actively contribute to define urbanization economies. As a matter of fact, it is hard for this whole argument to be unambiguously rescued in the data. For instance, as long as variety is measured giving some weight to the equi-distribution of sectoral shares at the local level, bigger metropolitan areas turn out to display less variety as compared with many smaller cities (see Fig. 1(d) and the discussion in the second section). Therefore, productive variety might not be necessarily associated with urbanization economies, but rather with the 'balanced' co-localization of sectors as a result, for instance, of mere input–output relations. Under this caveat, the definition of local attractiveness will also include Shannon's variety index (*VARIETY*) among the regressors.

Are the variables discussed so far sufficient to capture the effect of urbanization economies? And, by the same token, is the number of firms of the sector a sufficient indicator to capture localization economies? To make sure that these questions can be assessed, two further regressors will be introduced in order to characterize the 'fixed' attractiveness of a location: a binary variable for metropolitan LLMAs (*METRO*); and a binary variable indicating whether an LLMA is classified as an industrial district (*DISTRICT*). These two variables will serve to identify those 'residual' effects, stemming, respectively, from urbanization and localization economies, that are not detected by the other regressors. The magnitude of the elasticities connected to *METRO* and *DISTRICT* will reveal if too much has been left out from the present account of urbanization and localization economies.

Given these premises, the specification that will be used in the following econometric analysis is:

$$
\begin{aligned}
\log_c(\beta, x_1) = {} & \beta_1 \log POPULATION \\
& + \beta_2 \log AREA \\
& + \beta_3 \log CONSUMPTIONPC \\
& + \beta_4 \log PRODUCTIVITY \\
& + \beta_5\ INFDISTANCE \\
& + \beta_6\ VARIETY + \beta_7\ METRO \\
& + \beta_8\ DISTRICT + \beta_0 \qquad (12)
\end{aligned}
$$

where, as discussed above, the constant β_0 is set to zero for the multinomial model while it is regularly estimated in the Polya model, in which it represents the average effect. Notice that besides the two dummy variables, the regressors *INFDISTANCE* and *VARIETY* are also estimated without applying the logarithm; this choice is justified by the intention to signal even minimal effects associated with these regressors, which would have a risk of being statistically non-significant if the logarithm were applied. Notice also that the point estimates of the marginal elasticities are obtained here by standardizing all the regressor in equation (12) through the removal of the average and the division by the standard deviation: this operation allows one to compare directly the elasticities of the various regressors.

Multinomial versus Polya: relative performance

The methodology described so far allows one to evaluate the relative performance of two different models. The multinomial model differs from the Polya by the sole parameter b: in the former case, b is null so that no sector-specific externalities are allowed, while in the Polya case b can be non-negative. Although the analysis of the performance of models is typically relegated to the end of econometric investigations, in this case it is particularly useful to invert this order so as to streamline the exposition. In fact, the Polya model turns out to have better performance than the multinomial model across all sectors. Therefore, after having analysed the relative performances of the two models, it will be possible to focus exclusively on the Polya case without risking to lose any information.

The relative performances of the two models can be assessed with the Akaike information criterion corrected by finite sample size (AICc) (AKAIKE, 1974; HURVICH and TSAI, 1989):

$$AICc = 2k - 2\ln(L) + \frac{2k(k+1)}{n-k-1} \qquad (13)$$

where n is the sample size; k is the number of parameters in each model; and L is the maximized value of the likelihood function. Between two alternative models, the one with a lower AICc value is to be preferred as it 'dissipates' less information. In this sense, the definition of the AICc prizes the goodness of fit of the model via L, while it penalizes its parametric numerosity through k. Table 3 shows the AICc values for the multinomial and Polya models: the latter outperforms systematically the former. This suggests that no economic sector can do without an account of positive sector-specific externalities in order to explain completely the spatial distribution of firms. Yet, the weight of localization economies relative to other determinants of firm location remains to be assessed. As a matter of fact, though, it is possible to focus the rest of the exposition directly on the results of the Polya model without losing any relevant information.

Marginal elasticities

The strength of the various determinants of firm location can be evaluated through the marginal elasticities of the regressors shown in Table 2. A series of conclusions emerge from this exercise.

First, urbanization economies turn out to be diffused and relevant. Despite having introduced all the control variables discussed previously, the marginal elasticity connected to the regressor *POPULATION* is always positive and significant, and its magnitude is sizeable relative to those of the other regressors (see column (1) in Table 2). On average, a marginal increment of the population that resides in an LLMA by 0.36% attracts one extra plant. Remarkable deviations from this average appear across sectors: for instance, the value is above 1% in ATECO sectors 23, 30 and 66, while it is an order of magnitude below the average in ATECO sectors 17–20, 26, 361, 45, 51–55 and 61. In general, though, additional portions of population favour the localization of plants across all economic sectors. In this sense, the urbanization effect is economy-wide and it ends up attracting 'more of everything'.

Second, the level of consumption per capita has a positive effect especially on service sectors, yet the role of local aggregate demand unfolds mostly through the demographic scale of commuting zones. The marginal elasticity of *CONSUMPTIONPC* is statistically significant at a 99% confidence level and positive in seventeen sectors out of forty-three, while for eight of them it is negative and for the remaining eighteen it is not statistically significant (see column (3) in Table 2). In any case, even where the effect of consumption is positive, its estimated strength is rather limited. In particular, whenever the marginal elasticity of *CONSUMPTIONPC* is positive and significant, its value is typically some order of magnitude smaller than the one associated with *POPULATION*. Therefore, the effect of local aggregate demand unfolds mostly through the mere number of consumers rather than via their possibility or propensity to consume. This fact signals the predominant role played by demographic scale in deciding the demand-side effect of urbanization economies. Instead, the other control variables *AREA* and *PRODUCTIVITY*, although serving their empirical purpose, tend to be statistically non-significant or of minor importance (see columns (2) and (4) in Table 2).

Third, sector-specific localization economies are diffused and typically much stronger than urbanization economies. The marginal elasticity connected to n is always positive and significant, and its average magnitude amounts to 0.79%, which is more than twice the amount recorded for *POPULATION* (see column (9) in Table 2). Notably, this average datum does not

Table 2. Marginal elasticities with the Polya model

ATECQ – Sector	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
15 – Food products	1.26e–01**	4.66e–03**	−1.18e–03	−3.97e–04	4.59e–03**	2.69e–03*	1.62e–03	7.48e–04**	8.75e–01**
17 – Textiles	2.50e–02**	−1.03e–03*	−3.69e–04	8.93e–04**	−1.76e–04	8.32e–03**	4.02e–04	1.01e–02**	9.74e–01**
18 – Apparel	4.35e–02**	−1.95e–03**	−1.75e–03**	1.36e–04	−6.14e–04	1.02e–02**	2.12e–03**	2.31e–03**	9.68e–01**
19 – Leather products	8.96e–03**	−5.78e–04*	4.80e–04*	5.83e–04**	2.67e–04**	2.46e–03**	1.33e–03**	9.25e–03**	9.86e–01**
20 – Wood processing	8.72e–02**	4.87e–03**	1.55e–03	−8.42e–04	−1.68e–03**	5.98e–03**	1.74e–03**	5.69e–04**	9.04e–01**
21 – Pulp and paper	6.39e–01**	−3.67e–02**	−1.27e–02**	−2.39e–03	2.21e–03	1.90e–01**	2.36e–02**	1.84e–02**	7.03e–01**
22 – Publishing and printing	8.50e–01**	−4.45e–03	2.76e–02**	1.14e–02**	−6.10e–03	8.59e–02**	9.16e–03	–	6.92e–01**
23 – Coke, petroleum and nuclear fuel	1.17e+00**	6.66e–02**	−9.92e–03	6.42e–02**	1.06e–01**	4.84e–02	−1.01e–02	–	3.70e–01**
24 – Organic and inorganic chemicals	8.99e–01**	−5.61e–02**	1.93e–02	4.44e–04	−1.51e–02**	1.42e–01**	−6.69e–03	3.27e–04**	6.66e–01**
25 – Rubber and plastic products	3.54e–01**	−1.63e–02**	−1.32e–02**	2.12e–03	−7.09e–03**	1.40e–01**	−7.23e–04	7.33e–03**	8.29e–01**
26 – Non-metallic mineral products	7.41e–02**	4.60e–03**	−1.69e–03**	−2.66e–04	1.49e–03*	8.08e–03**	2.82e–03**	1.23e–03**	9.18e–01**
27 – Basic metals	4.28e–01**	−1.58e–02	−2.35e–02**	−6.53e–04	−1.30e–02**	1.75e–01**	1.04e–02	–	7.85e–01**
28 – Fabricated metal products	1.25e–01**	−1.24e–03	−2.12e–03**	−3.20e–04	−3.12e–03**	2.93e–02**	1.31e–03	2.74e–03**	9.23e–01**
29 – Industrial machinery	2.20e–01**	−8.09e–04	3.08e–03	4.04e–04	−3.50e–03**	9.07e–02**	−3.91e–03*	3.46e–03**	8.83e–01**
30 – Office machinery	1.12e+00**	7.14e–02**	−4.20e–02**	−4.16e–03	7.31e–03	2.33e–01**	4.97e–02**	7.56e–04**	5.11e–01**
31 – Electrical machinery	4.40e–01**	−1.11e–02	−6.25e–03	−1.68e–03*	−6.63e–03**	1.44e–01**	7.72e–03	1.04e–03**	8.03e–01**
32 – Radio, television and telecommunication devices	9.99e–01**	−4.81e–03	2.08e–02**	3.57e–03	−2.54e–03	1.29e–01**	−2.54e–03	2.86e–04**	5.47e–01**
33 – Precision instruments	3.89e–01**	4.93e–03*	4.73e–03*	6.10e–04	7.11e–04	7.08e–02**	9.79e–03**	2.14e–03**	8.13e–01**
34 – Motor vehicles and trailers	2.89e–01**	2.35e–02**	−1.95e–02**	−1.79e–03	−8.98e–03**	1.06e–01**	1.68e–02	–	8.34e–01**
35 – Other transport equipment	1.35e–01**	−1.40e–02**	1.03e–02**	−5.40e–04	1.83e–02**	9.59e–03**	8.98e–05	–	9.11e–01**
361 – Furniture	8.42e–02**	−2.85e–03	3.75e–03*	7.51e–04	−1.43e–03	1.03e–02**	6.74e–04	2.38e–02**	9.44e–01**
362 – Jewellery	1.04e–01**	9.60e–03*	4.79e–03	−5.27e–04	3.31e–03	−7.47e–04	7.47e–03**	8.97e–02**	9.22e–01**
363 – Musical instruments	3.23e–01**	−1.49e–02	2.13e–02*	−5.73e–03	4.34e–03	2.06e–01**	2.63e–03	2.10e–01**	7.29e–01**
36R – Residual of Sector 36	4.69e–01**	−1.21e–02*	9.61e–03*	−4.03e–03*	1.36e–03	1.19e–01**	3.19e–02**	–	7.58e–01**
40 – Electricity and gas	5.23e–01**	1.01e–01**	3.58e–02**	4.15e–03*	−1.83e–02**	7.67e–03	−8.29e–04	–	4.41e–01**
41 – Water	2.21e–01**	6.60e–02**	−2.04e–04	1.61e–03	2.04e–02**	−1.37e–02	4.08e–02*	–	6.37e–01**
45 – Construction	4.77e–02**	2.09e–03**	1.99e–03**	−4.12e–04	−5.60e–04**	3.74e–03**	−7.49e–04*	–	9.68e–01**
50 – Sale and services of motor vehicles	3.04e–01**	1.02e–02**	−7.47e–04*	−6.08e–04	1.73e–03*	2.89e–03**	−2.22e–03	–	8.12e–01**
51 – Wholesale and commission trade	6.96e–02**	−1.93e–03**	4.33e–04	2.04e–05	1.42e–03**	1.10e–02**	3.14e–04	–	9.69e–01**
52 – Retail trade	8.31e–02**	1.27e–03**	6.00e–04*	−4.87e–04**	1.43e–03**	−3.33e–03**	5.33e–04*	–	9.50e–01**
55 – Hotels and restaurants	2.61e–02**	9.32e–04**	3.66e–03**	−2.53e–04**	−2.70e–04*	−2.68e–03**	7.48e–04**	–	9.79e–01**
60 – Land transport	1.03e–01**	4.31e–03**	5.37e–03**	−9.97e–04*	1.04e–03	1.14e–02**	3.41e–03*	–	9.34e–01**
61 – Water transport	4.51e–02**	−5.55e–03*	5.30e–03**	−3.69e–04	6.12e–03*	−4.89e–03*	6.19e–03**	–	9.61e–01**
62 – Air transport	8.00e–01**	1.16e–01	1.57e–01**	4.90e–02	−2.30e–02	−1.23e–02	3.26e–02	–	7.26e–01**
63 – Auxiliary transport activities	3.50e–01**	−1.11e–02	2.42e–02**	3.42e–03*	9.26e–03**	1.64e–03	1.45e–02**	–	8.63e–01**
64 – Post and telecommunications	2.88e–01**	3.83e–02**	1.03e–02**	5.10e–03**	−5.78e–03**	2.09e–02**	1.34e–02**	–	6.76e–01**
65 – Financial intermediation	9.36e–01**	3.44e–02**	4.96e–02**	1.97e–03	−6.55e–03*	1.37e–01**	1.03e–02	–	4.02e–01**
66 – Private insurance and pensions	1.56e+00**	6.89e–03	−1.87e–03	1.04e–01**	3.07e–02	2.68e–01**	−1.83e–02	–	3.63e–01**
67 – Auxiliary financial activities	4.63e–01**	5.15e–03	1.16e–02**	2.22e–03**	2.85e–03*	5.08e–02**	−8.29e–03**	–	7.87e–01**
70 – Real estate activities	1.04e–01**	−2.77e–03**	1.32e–02**	−3.58e–04	−4.95e–04*	2.70e–02**	−3.36e–04	–	9.59e–01**
71 – Renting of machinery and equipment	3.75e–01**	1.55e–02	2.88e–02**	−2.00e–03	9.68e–03**	−6.72e–04	2.77e–03	–	7.88e–01**
72 – Computer and related activities	4.86e–01**	9.30e–04	2.16e–02**	3.82e–03*	−3.51e–03	5.74e–02**	−4.95e–03	–	8.27e–01**
74 – Business services	1.81e–01**	3.53e–04	8.02e–03**	7.18e–04*	−8.06e–04	1.58e–02**	9.35e–04	–	9.31e–01**

Notes: Variables by commuting zone: (1) Population; (2) Area; (3) Consumption per capita; (4) Labour productivity (VA/employees); (5) Average distance from transport infrastructures; (6) Shannon variety index; (7) Dummy variable for metropolitan areas; (8) Dummy variable for industrial districts; and (9) Number of plants belonging to the sector (n_i).

**Significance above 99%; *significance above 95%.

The point estimates of the marginal elasticities are obtained by maximum likelihood applying equations (9) and (10).

Table 3. *Relative and absolute performances of the Polya model*

ATECO-NACE sector	Relative performance		Absolute performance
	AICc – multinomial	AICc – Polya	Efron-R^2 Polya
15 – Food products	9.80e+03	6.13e+03	9.33e–01
17 – Textiles	2.62e+04	5.34e+03	4.00e–01
18 – Apparel	2.52e+04	5.90e+03	7.16e–01
19 – Leather products	2.99e+04	4.27e+03	4.09e–01
20 – Wood processing	1.23e+04	6.05e+03	8.75e–01
21 – Pulp and paper	3.67e+03	2.73e+03	9.09e–01
22 – Publishing and printing	5.56e+03	4.36e+03	9.63e–01
23 – Coke, petroleum and nuclear fuel	1.44e+03	1.39e+03	8.12e–01
24 – Organic and inorganic chemicals	3.78e+03	3.11e+03	8.66e–01
25 – Rubber and plastic products	7.06e+03	3.97e+03	8.83e–01
26 – Non-metallic mineral products	1.43e+04	5.64e+03	7.37e–01
27 – Basic metals	3.94e+03	2.57e+03	7.87e–01
28 – Fabricated metal products	1.77e+04	6.65e+03	9.29e–01
29 – Industrial machinery	1.31e+04	5.32e+03	8.92e–01
30 – Office machinery	1.93e+03	1.71e+03	8.87e–01
31 – Electrical machinery	7.46e+03	4.22e+03	9.20e–01
32 – Radio, television and telecommunication devices	3.73e+03	3.33e+03	9.53e–01
33 – Precision instruments	7.65e+03	4.55e+03	9.54e–01
34 – Motor vehicles and trailers	2.94e+03	1.96e+03	4.99e–01
35 – Other transport equipment	6.41e+03	2.85e+03	5.50e–01
361 – Furniture	1.65e+04	5.30e+03	8.63e–01
362 – Jewellery	8.64e+03	3.54e+03	7.46e–01
363 – Musical instruments	1.81e+03	1.13e+03	5.94e–01
36R – Residual of Sector 36	4.44e+03	3.06e+03	8.37e–01
40 – Electricity and gas	3.17e+03	3.03e+03	8.34e–01
41 – Water	2.63e+03	2.24e+03	5.98e–01
45 – Construction	2.93e+04	8.37e+03	9.77e–01
50 – Sale and services of motor vehicles	8.08e+03	6.28e+03	9.89e–01
51 – Wholesale and commission trade	3.97e+04	7.92e+03	9.56e–01
52 – Retail trade	2.15e+04	8.38e+03	9.91e–01
55 – Hotels and restaurants	4.31e+04	8.17e+03	9.50e–01
60 – Land transport	1.61e+04	6.84e+03	9.47e–01
61 – Water transport	3.99e+03	1.22e+03	9.05e–02
62 – Air transport	8.87e+02	6.97e+02	8.74e–01
63 – Auxiliary transport activities	7.86e+03	4.79e+03	9.77e–01
64 – Post and telecommunications	5.26e+03	4.59e+03	9.44e–01
65 – Financial intermediation	4.85e+03	4.52e+03	9.75e–01
66 – Private insurance and pensions	1.76e+03	1.68e+03	8.03e–01
67 – Auxiliary financial activities	7.94e+03	5.57e+03	9.87e–01
70 – Real estate activities	2.74e+04	6.35e+03	9.65e–01
71 – Renting of machinery and equipment	5.83e+03	4.03e+03	9.71e–01
72 – Computer and related activities	9.13e+03	5.59e+03	9.80e–01
74 – Other business activities	1.71e+04	6.95e+03	9.54e–01

Note: AICc, Akaike information criterion corrected by finite sample size.

derive from a concentration of extremely high magnitudes in a few sectors; on the contrary, the marginal elasticity of n exceeds that connected to *POPULATION* in thirty-five sectors out of forty-three.[5] Together with the superiority of the Polya model relative to the multinomial one, this regularity shows that localization economies are a diffused and predominant determinant of the spatial distribution of economic activities. Consequently, those technological and organizational factors on which localization economies are grounded cannot be regarded as a 'special case' concerning only some peculiar sectors or locations. In this sense, it is the rule rather than the exception that firms co-localize with other firms to which they are similar.

Fourth, the locational pull exerted by transport infrastructures is limited and specific to some sectors. The marginal elasticity connected to *INFDISTANCE* is significantly different from zero with a 99% confidence level and negative in twelve sectors out of forty-three, while in nine sectors it is significant but positive (see column (5) in Table 2). Where the sign is negative, that is when proximity to transport infrastructures attracts localization, the magnitude of the effect is quite heterogeneous across sectors: this derives presumably from the inter-sectoral differences in transportation costs. A sector like '27 – Basic metals' is characterized by fairly high transportation costs as compared, for instance, with service activities such as those of sector '55 – Hotels and restaurants'; consequently, being marginally closer to transport infrastructures increases the probability of an LLMA to attract a new plant about fifty times more in the former sector than in the latter. In a minority of sectors, transport infrastructures are instead 'repulsive' of firm location: that is, the marginal elasticity of *INFDISTANCE* is significant and positive.[6] This fact has to do with two separate features that characterize some of the sectors in question: the intensive use of some natural resources as input factors and the special attention devoted by urban development plans to polluting activities. In fact, both elements may drive firms to localize relatively far from urban or specific areas, thus tending to determine also a greater distance from transport infrastructures. For instance, when agricultural land is used intensively such as in sector '15 – Food products', plants are automatically driven toward the countryside; by the same token, the intensive use of soft water in sector '41 – Water' drives plants to follow the spatial dislocation of water sources and rivers rather than the spread of people, which is instead how infrastructures are located. On the other hand, urban plans tend explicitly to ban the location of polluting activities in residential areas and relegate them inside the perimeter of rather separated industrial areas; for this reason, a polluting sector such as '19 – Leather' may be located relatively far from highly populated areas, which is where most of the transport infrastructures are placed. However, even accounting for these various sector-specific effects, it is most striking to observe that very often the marginal proximity of an LLMA to transport infrastructures does not affect its probability to attract plants. More precisely, the marginal elasticity connected to *INFDISTANCE* is not significantly different from zero with a 99% confidence in twenty-two sectors out of forty-three. This may well derive from a sufficiently widespread presence of transport infrastructures over the Italian territory. Indeed, if, on the one hand, some areas are relatively more endowed than others, Fig. 1(b) shows that there are no 'infrastructural deserts'. Therefore, the diffusion of transport infrastructures may be sufficiently capillary as not to affect the locational choices of firms in a statistically significant way.

Fifth, local productive variety has at most a secondary effect on the attractiveness of commuting zones. The marginal elasticity connected to *VARIETY* is positive and significantly different from zero with a 99% confidence in thirty-two sectors out of forty-three, while being negative in two sectors and non-significantly different from zero in the remaining nine sectors (see column (6) in Table 2). In terms of its relative weight, the pull of variety turns out to be always much lower, for instance, than the pull of demography. As an example, the marginal elasticities of *VARIETY* and *POPULATION* are most similar for sector '19 – Leather products', where in any case the magnitude of the former regressor is about thrice as small as the magnitude of the latter. Since the values at stake are also small in absolute terms, the effect of variety is definitely of secondary importance to explain the spatial structure of economic activities. As a remark, it is worth noting that the few sectors in which variety is associated with a negative locational effect are possibly affected by the role of tourism. In fact, sectors such as '52 – Retail trade' and '55 – Hotels and restaurants' are often concentrated in commuting zones that are oriented toward tourist, rather than manufacturing activities.

Sixth, and finally, the effect of urbanization and localization economies appear to be vastly captured by the set of regressors discussed so far. In fact, the marginal elasticities connected to the binary variable for metropolitan areas, *METRO*, are always much lower than the marginal effects connected to *POPULATION*, thus signalling that any residual driver of urbanization economies is at most of secondary importance. Similarly, the marginal elasticities connected to the binary variable for industrial districts, *DISTRICT*, are always much lower than the marginal effects connected to n, which signals that localization economies are described quite accurately by the sheer number of firms in the sector.

Absolute performance of the Polya model

The absolute performance of the Polya model now needs to be evaluated. Essentially, this means one has to verify how closely the spatial distributions predicted by the model match the empirically observed ones: that is, the goodness of fit of the model.

A rather coarse assessment of the goodness of fit of the Polya model can be accomplished through histograms. These are built on the basis of occupancies, $f(n)$, which count the number of locations hosting exactly n plants. For example, $f(0)$ is the number of locations containing zero plants belonging to the sector under scrutiny; $f(1)$ is the number of locations hosting exactly one plant, and so on. Then, the general definition of occupancy is:

$$f(n) = \sum_{l=1}^{L} \delta_{n_l,n} \qquad (14)$$

where $\delta_{n_l,n}$ is the Kronecker delta. It follows that:

$$\sum_{n=0}^{+\infty} f(n) = L$$

where, although infinity appears as upper bound, the summation effectively stops with the number of plants in the most populated location. To synthesize information further, observations can be grouped into classes whose ranges are defined by the geometric

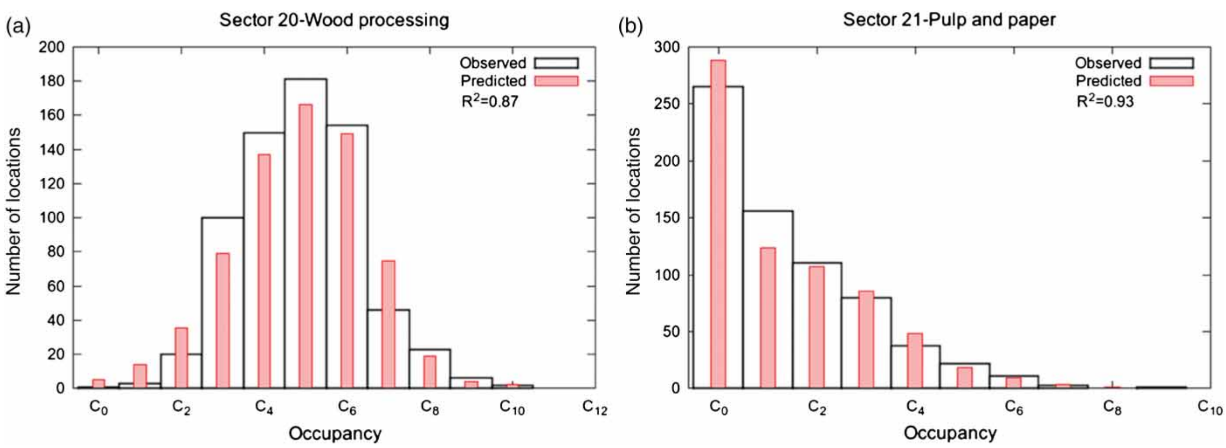

Fig. 3. Polya model: predicted and observed spatial distributions in two exemplifying sectors

progression:

$$C_k = [2^k - 1, 2^{k+1} - 2), \quad k = 0, 1, 2, \ldots \qquad (15)$$

So that occupancy classes are:

$$f(C_k) = \sum_{n \in C_k} f(n) \qquad (16)$$

With this definition, the ranges of the occupancy classes will then be $C_0 = 0$; $C_1 = 1$–2; $C_2 = 3$–6, and so forth, up to a class that includes the maximum number of plants observed in a location.

Through occupancy classes it is then possible to build histograms showing the observed spatial distributions of plants (Fig. 3), which give a visual hint about the degree of spatial concentration that characterizes each sector. An asymmetric distribution means that there is an high occupancy of the first classes compared with the latter ones: most locations host null or minimal shares of the plants in the sector, while a few other locations host big shares. Hence, the sectors tends to be spatially concentrated in a few locations. To the contrary, a more bell-shaped histogram indicates that locations tend to host similar 'typical' shares of firms, which are then more uniformly distributed in space. In this sense, histograms can immediately show that a sector like pulp and paper in Fig. 3(b) is more spatially concentrated than a sector such as wood processing in Fig. 3(a). As can be seen, the distribution of occupancy classes predicted by the model well approximates the observed one. The agreement attained in the two displayed sectors is not peculiar, but typical of the large majority of sectors.

A finer numerical measure of the goodness of fit of the Polya model can be obtained with Efron-R^2. This measure assesses the agreement between predictions and observations at the level of single locations, rather than at the level of occupancy classes. Formally:

$$\text{Efron-}R^2 = 1 - \frac{\sum_l (n_l - n_l^*)^2}{\sum_i (n_l - \bar{n})^2} \qquad (17)$$

where n_l is the observed value of the lth location; n_l^* is the value predicted by the Polya model for the lth location; and $\bar{n}$ is the average number of firm per location. Efron-R^2 is bounded between 0 and 1 and yields higher values for higher agreements, thus providing an absolute and comparable measure of the goodness of fit of the Polya model in the different sectors.

The values of Efron-R^2 in Table 3 signal that the Polya model has a rather high absolute performance. On average, the value of Efron-R^2 across sectors is 0.83, although some relevant deviations from the average do exist. In fact, the model seems able to capture the spatial distributions of sectors that are technologically, organizationally and spatially very different. For instance, the agreement between observations and predictions is high for sectors that are characterized by the presence of big firms operating with scarcely modularizable production processes, such as for the basic metals sector, as well as for sectors that are largely dominated by small-to-medium firms operating on modularizable processes, such as for the leather sector. By the same token, symmetric spatial distributions, such as the one of the wood sector, are captured with the same accuracy as more asymmetric distributions, such as the one of the office machinery sector.

Despite the generally satisfactory performance of the Polya model, there still is much room for improvement. In particular, the agreement between observations and predictions turns out to be lower in some sectors. For instance, the values for Efron-R^2 reach their minimum for sectors '17 – Textile' and '19 – Leather products', which are characterized by an especially high presence of industrial districts. This fact may entail that some relevant determinants of firm location have been overlooked for what concerns these two sectors. Similarly, the absolute performance of the Polya model is below the average for sectors '34 – Motor vehicles and trailers' and '35 – Other transport equipment', in which the spatial distribution of plants is possibly affected by the co-location of leading firms with their spin-offs, as suggested by KLEPPER (2002). Thus, also in these sectors an important driver of firm location may have been neglected.

CONCLUSION

The present work has proposed an empirical framework to explain what makes a location attractive to firms of a particular sector. A series of conclusions have been drawn. First, between two alternative but comparable models, the one not allowing for sector-specific positive externalities matched the data systematically worse than the model allowing for them. Second, both urbanization effects and sector-specific localization economies were found to affect the location of firms significantly. However, and third, localization economies were often found to be more relevant than urbanization motives. According to these findings, an explanation of the spatial structure of economic activities must rely primarily on the technological dynamics that are the likely cause of sector-specific positive externalities, and only secondarily on urbanization effects.

The analysis presented here is open to a number of further developments. To begin with, the authors' conclusion about the relevance of sector-specific positive externalities is in sharp contrast with that by KIM (1995) concerning manufacturing in the United States. Therefore, it would be interesting to apply the present framework on US data, as well as on other countries, to detect how the results would possibly change. Further developments should possibly regard the use of alternative specifications in order to verify how they affect the discrepancy between the predicted spatial distributions and the observed ones. Similarly, it would also be useful to try out other explanatory variables either to substitute or to be added to the ones used here. Finally, it is necessary to evaluate systematically how both spatial and sectoral disaggregations influence the outcome of the analysis; this operation would help one not only to test the 'robustness' of the model, but also to infer the geographic range of the various effects under scrutiny.

Acknowledgements – The research leading to this work received funding from the Italian Ministry of University and Research as a part of the project 'The Growth of Firms and Countries: Distributional Properties and Economic Determinants' (PRIN 2009 programme, Protocol Number 2009H8WPX5).

NOTES

1. A vast literature exists which tries to assess the relative importance of Marshallian and Jacobs externalities (for a remarkable comparative survey, see BEAUDRY and SCHIFFAUEROVA, 2009). The present work shares some of the interests of that literature and, in fact, it will explicitly consider the role of productive variety among the determinants of firm location. However, as discussed in the second section, variety does not turn out to be a prerogative of bigger cities as compared with smaller centres. In this sense, the present work takes a distance from the original view put forward by Jacobs.
2. ATECO is the Italian version of NACE, the statistical classification of economic activities developed by the European Statistical Office (Eurostat).
3. It makes sense to run this correlation only for the manufacturing sectors since the statistical definition of industrial district allows one to have only manufacturing districts, as described by SFORZI (2001). Therefore, running the correlation with all the sectors would be somewhat biased, since by definition $D = 0$ in service sectors. In fact, if one attempts to run the rank correlation including all sectors, both the value of the Spearman index and its statistical significance decrease by an order of magnitude, being $\rho = 0.030$, with $p = 0.847$. More generally, the Spearman index remains small and with a low significance level (that is, with $p > 0.5$) even with some minor modifications of the two variables G and D. For instance, the result holds even if D is substituted with a binary variable identifying the presence/absence of industrial districts in the sector, or when metropolitan LLMAs are excluded from the computation of G. In this sense, the weak correlation between overall spatial concentration and industrial districts seems to be a rather robust regularity.
4. For an alternative solution based on the choice of a suitable functional specification for the parameter c, see BOTTAZZI *et al.* (2007).
5. The exceptions are represented by ATECO-NACE sectors 22–24, 30, 40, 62 and 65–66.
6. These cases are represented by ATECO-NACE sectors 15, 19, 23, 35, 41, 51–52, 63 and 71.

REFERENCES

AKAIKE H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19(6)**, 716–723.

BEAUDRY C. and SCHIFFAUEROVA A. (2009) Who's right, Marshall or Jacobs? The localization versus urbanization debate, *Research Policy* **38(2)**, 318–337.

BOTTAZZI G., DOSI G., FAGIOLO G. and SECCHI A. (2007) Modeling industrial evolution in geographical space, *Journal of Economic Geography* **7(5)**, 651–672.

BOTTAZZI G., DOSI G., FAGIOLO G. and SECCHI A. (2008) Sectoral and geographical specificities in the spatial structure of economic activities, *Structural Change and Economic Dynamics* **19(3)**, 189–202.

Bottazzi G. and Secchi A. (2007) *Repeated Choices Under Dynamic Externalities*. LEM Working Paper Series, September 2007. Laboratory of Economics and Management (LEM)-Scuola Superiore Sant'Anna, Pisa (available at: http://www.lem.sssup.it/WPLem/files/2007-08.pdf).

Desmet K. and Fafchamps M. (2006) Employment concentration across US counties, *Regional Science and Urban Economics* **36(4)**, 482–509.

Devereux M. P., Griffith R. and Simpson H. (2004) The geographic distribution of production activity in the UK, *Regional Science and Urban Economics* **34(5)**, 533–564.

Dumais G., Ellison G. and Glaeser E. L. (2002) Geographic concentration as a dynamic process, *Review of Economics and Statistics* **84(2)**, 193–204.

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

Ellison G. and Glaeser E. L. (1999) The geographic concentration of industry: does natural advantage explain agglomeration?, *American Economic Review* **89(2)**, 311–316.

Giordano R. (2009) I differenziali salariali tra i settori pubblico e privato in italia, in *Mezzogiorno e politiche regionali (Banca d'Italia)* **no. 2(November)**, 137–158.

Hurvich C. M. and Tsai C. L. (1989) Regression and time series model selection in small samples, *Biometrika* **76(2)**, 297–307.

Istituto Nazionale di Statistica (ISTAT) (1997) *I sistemi locali del lavoro 1991*. ISTAT, Rome.

Istituto Nazionale di Statistica (ISTAT) (2001) *8vo Censimento generale dell'industria e dei servizi*. ISTAT, Rome.

Istituto Nazionale di Statistica (ISTAT) (2006) *Atlante statistico dei comuni italiani*. ISTAT, Rome.

Jacobs J. (1969) *The Economy of Cities*. Vintage, New York, NY.

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(4)**, 881–908.

Klepper S. (2002) The evolution of the US automobile industry and Detroit as its capital. Paper presented at the 9th Congress of the International Joseph A. Schumpeter Society, Gainesville, FL, USA, 2002.

Krugman P. R. (1991) Increasing returns and economic geography, *Journal of Political Economy* **99(3)**, 483–499.

Marshall A. (1890) *Principles of Economics*. Macmillan, London.

Maurel F. and Sédillot B. (1999) A measure of the geographic concentration in French manufacturing industries, *Regional Science and Urban Economics* **29(5)**, 575–604.

Overman H. G. and Duranton G. (2001) *Testing for Localization Using Micro-Geographic Data*. CEPR Discussion Paper Series Number 3379. Centre for Economic Policy Research (CEPR), London.

Sforzi F. (2000) Il sistema locale come unità di analisi integrata del territorio, in Gori E., Giovannini E. and Batic N. (Eds) *Verso i censimenti del 2000*, pp. 185–192.

Sforzi F. (2001) Distretti industriali e sistemi locali del lavoro 2001, in *8° Censimento generale dell'industria e dei servizi*. Istituto Nazionale di Statistica (ISTAT), Rome.

Unioncamere (2009) *Appendice statistica al Rapporto Unioncamere – Consumi finali interni per tipologia e provincia, Tav. 5.6*. Unioncamere, Rome.