Laboratory of Economics and Management

Sant'Anna School of Advanced Studies

Piazza Martiri della Libertà, 33 - 56127 PISA (Italy)

Tel. +39-050-883-343 Fax +39-050-883-344

Email: lem@sssup.it Web Page: http://www.lem.sssup.it/

LEM

Working Paper Series

Cities and clusters: economy-wide and sector-specific effects in corporate location

Giulio Bottazzi°

Ugo Gragnolati°

° CAFED and LEM, Scuola Superiore Sant'Anna, Italy

2010/16 September 2010

ISSN (online) 2284-0400

Cities and clusters: economy-wide and sector-specific effects in corporate location

Giulio Bottazzi and Ugo Gragnolati

CAFED and LEM, Scuola Superiore Sant'Anna, Pisa, Italy

1 September 2010

Abstract

Are the observed spatial distributions of firms decided mostly by market-mediated, economy-wide locational forces, or rather by non-pecuniary, sector-specific ones? This work 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 Labor Systems.

JEL codes: L1, C31, R3

Keywords: Industrial Location, Sector-specific Agglomeration, Urbanization Economy, Maximum Likelihood Estimation.

1 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 labor suppliers. The positive externalities 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, this agglomerative pull has been typically identified with the feedback mechanism between local demand and local labor supply (Krugman, 1991). As such, pecuniary elements tend to yield a common and general clustering of firms and households, which engenders and reinforces urban agglomeration.

At the same time, however, several sectors of the economy show a gathering of common activities in places that are not densely populated nor particularly well connected to metropolitan areas. These specialized clusters escape the explanation provided by the pecuniary effects of demand and supply: in fact, the sectoral production within a cluster exceeds local demand by far. More plausibly, in these cases agglomeration results from non-pecuniary factors, which generate a locational advantage for firms through productive 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, often scarcely codified, within organizational contexts that allow and facilitate it (Becattini, 1990; Marshall, 1890). As such, non-pecuniary factors are likely to engender the agglomeration of firms sharing very similar processes and structures and operating within the same industry, thus leading to the formation of clusters.

This work investigates the strength of the urbanization effect relative to sector-specific agglomerative forces in shaping the observed geographical distribution of manufacturing and service activities in Italy. The urbanization effect, plausibly being the outcome of pecuniary, market-mediated interactions, is expected to act *across* the different economic sectors, albeit with possibly different strengths. Conversely, sector-specific agglomerative force stemming from technological and organizational specificities are expected to act *within* each sector separately. On the ground of these considerations, we compare the observed spatial distributions of firms in each sector with the ones predicted by the discrete choice model of Bottazzi and Secchi (2007), thus allowing to assess the relative strength of 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; Kim, 1995; 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 one improves upon them by using maximum likelihood methods to obtain point estimates of the parameter of interest and Monte Carlo re-sampling to estimate the variance and thus the statistical significance of the different marginal effects. As such the magnitudes of the various locational drivers will be unambiguously interpretable and directly comparable one with the other.

The main result presented here regards the effect played by market-mediated pecuniary motives and technology-related non-pecuniary motives in deciding firm location. The present analysis finds that firm location is affected by both. However, 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 prescind 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. Section 2 presents the ISTAT database on which the analysis is carried out together with the choices that were operated on the data. Section 3 introduces some descriptive statistics as well as some hints regarding the distinct effects of urbanization and industrial clustering on spatial distributions. Section 4 sketch the discrete choice model on which theoretical predictions are based. Section 5 proceeds to estimate the marginal effects of the various factors identified as possible explanators of the attractiveness of locations. Section 6 summarizes and discusses the relevant results.

2 Data

The present analysis is based on Italian data taken from the *Atlante statistico dei comuni italiani* (hereafter Atlas) published by ISTAT ISTAT (2006). From this database we take *(i)* the census of manufactures and services, and *(ii)* the census of population and housing for the year 2001. The former census provides the data concerning the number of business units and workers in each sector, while the latter census includes the data on the population living in each geographical region. We consider only the business units classified as firms, neglecting instead non-profit and governmental organizations. Under these conditions, the data for 2001 accounts approximately for 3.5 million of business units and 13.8 million workers, which amount to 71.2% of the employment in the Italian economy.

Sectors are disaggregated following the 2-digit ATECO classification (which corresponds to the NACE classification). Sector "36-Furniture and other manufacturing activities" is further disaggregated at a 3-digit level in order to capture more accurately what the "other" activities were, namely: "361-Furniture","362-Jewelry", and "363-Musical instruments", plus the residual "36R-Residual of sector 36". With this disaggregation we can 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. Instead, this kind of disaggregation was not applied to the other residual sector concerning services, that is "74-Other business activities": in fact, despite being residual, such sector is actually well characterized in productive terms as it collects essentially professionals. Table 1 reports some summary statistics of the sectors under analysis.[1]

Business units are distributed across 686 geographical locations identified as Local Labor Systems (hereafter LLS).[2] These are preferred over purely administrative regions because they preserve the spatial continuity of phenomena, such as agglomeration, that are central to the present analysis. In fact, by discretizing space according to administrative borders, some bias can be created possibly due to an excessively coarse grained grid or to an excessively fine one, as illustrated in Figure 1.

The shaded areas on the map represent two separate agglomerations: the smaller one is contained in the micro-administrative region *7* and belongs to the macro-administrative region A; while the bigger agglomeration is located across the micro-administrative regions *1–6* and belongs to the macro-administrative regions A–D. Suppose that *an* agglomeration is defined by the statistical criterion "shaded area". How many agglomerations would be counted if micro-administrative regions were used as geographical unit of analysis? There would be seven agglomerations, which clearly constitutes an overestimation. Apparently, spatial aggregation would seem to help. If macro-administrative regions were used, the number of agglomerations would reduce to four thus improving the measurement. However, if one further aggregation is made hoping to solve completely the overestimation bias, the exact opposite problem arises: joining macro-administrative regions A–D would lead to identify only one agglomeration. Moreover, this solution would decrease substantially the geographical significance of the chosen unit of analysis; in fact, the unit A–D would be regarded as hosting an agglomeration although most of its internal space is not actually characterized by such phenomenon (i.e. the blank blocks are more than the shaded ones). In order to overcome these problems ISTAT proposes a methodology that aggregates the smallest adjoining administrative units (i.e. municipalities) into bigger geographical areas through an algorithm based on the flows of commuters. The aggregations stemming from such algorithm are the LLSs ISTAT (1997); Sforzi (2000). Basically, an LLS is a set of adjoining municipalities that have in common a relevant flow of commuters toward the same municipality; as such they share their "peripheral" role with respect to an economic "center" with which they form a sole socio-economic region. Such criterion of aggregation allows to go "beyond" administrative borders: as shown in Figure 1, the borders of the LLS manage to aggregate lower administrative units (i.e. the small squares) while crossing the borders of higher order ones (i.e. regions A–D).

[1] The 8th column of Table 1 shows the number of industrial districts in each sector. Although such data is indeed present in our database, it can be recovered more easily from the two files `tav_02_distr.xls` and `tav_16_distr.xls` downloadable from the ISTAT site:
`http://dwcis.istat.it/cis/download_distretti_industriali.htm`.

[2] We use LLSs in their 2001 definition.

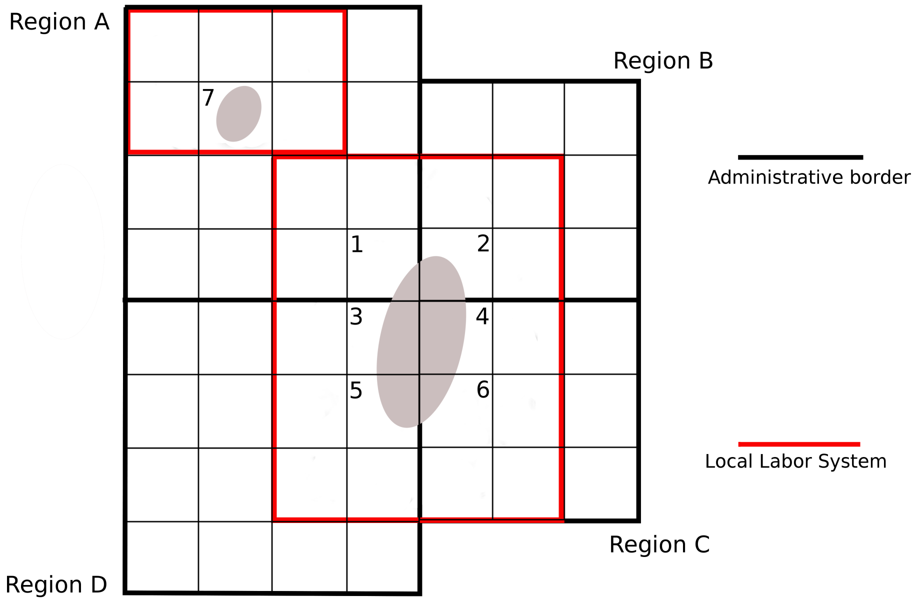

Figure 1: Administrative borders and LLS

3 Empirical analysis

The present analysis spans economic sectors which are strongly heterogeneous in several dimensions, as summarized in Table 1. First, the total number of business units N and workers W change dramatically across sectors, thus implying that structurally different economic activities are being taken into account. Second, sectors differ greatly in their spatial distributions in two main respects. On one hand, the fraction F of locations hosting at least one firm in the sector reveals that some sectors are spread literally everywhere (i.e. $F \approx 1$) while others are present only in some locations (i.e. $F << 1$). On the other hand, the maximum share of business units found in a location, $Max(n_l/N)$, reveals that the scale of locational effects is greater in some sectors than in others. As mentioned in the introduction, this heterogeneity in spatial distribution is likely related to the varying weights that urbanization and positive externalities of localization hold in each sector. Both factors contribute to increase the overall degree of spatial concentration. A measure of the latter can be obtained looking at the skewness of the distribution of business units across locations. Skewed distributions signal that many locations host null or small shares of the business units in the sector, while a few other locations capture most it; conversely, less skewed distributions are associated to a more even spread of activities in space. The Gini coefficient G corrected by sample size shows substantial differences in the degree of spatial concentration that characterizes each sector (values of G close to 1 suggest a strong clustering of activities). A second measure of the sectoral tendency to cluster is represented simply by the number D of industrial districts associated to each sector. Generally, industrial districts are located in non metropolitan areas and they are typically linked to peculiar organizational and technological elements (Becattini, 1990). So we can take their number as a proxy for the sectoral strength of positive localization externalities.

Interestingly, running a rank correlation corrected by ties between G and D reveals that the two measures are *not* positively correlated. The Spearman index is $\rho = 0.204$ and its *p-value* is 0.189: hence, ρ is not different from zero at 95% significance. Notice that in principle the

ATECO-Sector	N	W	F	$Max\frac{n_l}{N}$	G	G'	D	D'
15-Food products	73680	443930	0,999	0,025	0,575	0,516	7	1
17-Textiles	31984	309487	0,904	0,178	0,834	0,830	45	1
18-Apparel	46377	298241	0,950	0,048	0,780	0,752	0	0
19-Leather products	24195	206035	0,701	0,097	0,904	0,896	20	1
20-Wood processing	50250	179313	0,999	0,036	0,574	0,522	0	0
21-Pulp and paper	5175	84212	0,614	0,100	0,819	0,775	0	0
22-Publishing and printing	29166	173431	0,940	0,146	0,802	0,691	4	1
23-Coke,petroleum and nuclearfuel	913	24537	0,370	0,051	0,835	0,800	0	0
24-Organic and inorganic chemicals	7721	205153	0,720	0,163	0,814	0,742	0	0
25-Rubber and plastic products	15115	216876	0,757	0,097	0,804	0,770	4	1
26-Non metallic mineral products	31177	253664	0,988	0,030	0,640	0,596	0	0
27-Basic metals	3984	139287	0,548	0,096	0,841	0,815	0	0
28-Fabricated metal products	102295	700984	0,999	0,060	0,711	0,661	0	0
29-Industrial machinery	46481	597544	0,927	0,084	0,793	0,756	38	1
30-Office machinery	1715	19257	0,442	0,104	0,847	0,788	0	0
31-Electrical machinery	20282	211404	0,832	0,115	0,808	0,754	0	0
32-Radio,TV,and TLC devices	9677	107578	0,821	0,099	0,767	0,687	0	0
33-Precision instruments	26244	126004	0,907	0,090	0,780	0,704	0	0
34-Motor vehicles and trailers	2229	172932	0,402	0,177	0,868	0,828	0	0
35-Other transport equipment	4951	103096	0,555	0,045	0,831	0,818	0	0
361-Furniture	35784	209188	0,943	0,094	0,797	0,778	32	1
362-Jewlery	10906	50232	0,716	0,132	0,879	0,863	5	1
363-Musical instruments	695	2740	0,241	0,180	0,916	0,915	1	1
36R-Residual of sector 36	6728	39233	0,690	0,093	0,814	0,766	0	0
40-Electricity and gas	4159	109047	0,885	0,053	0,619	0,566	0	0
41-Water	1408	15961	0,571	0,043	0,725	0,681	0	0
45-Construction	529757	1528629	1,000	0,048	0,652	0,587	0	0
50-Sale and services of motorvehicles	164079	457527	1,000	0,058	0,645	0,563	0	0
51-Wholesale and commission trade	404278	1021666	0,999	0,083	0,776	0,700	0	0
52-Retail trade	772730	1675275	1,000	0,059	0,638	0,546	0	0
55-Hotels and restaurants	261304	853122	1,000	0,056	0,616	0,537	0	0
60-Land transport	135135	531539	1,000	0,077	0,695	0,609	0	0
61-Water transport	1319	20394	0,187	0,434	0,960	0,963	0	0
62-Air transport	457	24973	0,131	0,260	0,961	0,944	0	0
63-Auxiliary transpor tactivities	33765	322071	0,946	0,109	0,803	0,700	0	0
64-Post and telecommunications	18056	289518	1,000	0,044	0,575	0,504	0	0
65-Financial intermediation	30587	392870	1,000	0,071	0,679	0,595	0	0
66-Private insurance and pensions	1771	40591	0,465	0,106	0,856	0,800	0	0
67-Auxiliary financial activities	84677	154227	0,994	0,075	0,750	0,668	0	0
70-Real estate activities	149990	226736	0,926	0,142	0,840	0,777	0	0
71-Renting of machinery and equipment	13291	29536	0,879	0,083	0,726	0,647	0	0
72-Computer and related activities	84100	354847	0,987	0,138	0,808	0,711	0	0
74-Other business activities	216883	904234	1,000	0,142	0,788	0,680	0	0

Table 1: Summary statistics of 2/3-digit sectors in the census of manufactures and service data. N, number of business units; W, number of workers; F, fraction of locations hosting at least one business unit belonging to the sector; $Max(n_l/N)$, maximum share of business in a single location; G, Gini coefficient corrected by sample size of the distribution of business units across all locations; G', Gini coefficient corrected by sample size of the distribution of business units across all locations except metropolitan areas; D, number of industrial districts belonging to each sector; D' is such that, for sector i, $D'_i = 1$ if $D_i \geq 1$ and $D'_i = 0$ otherwise.

Metropolitan areas are defined by ISTAT as the following LLSs: 7-Torino; 57-Milano; 138-Verona; 158-Venezia; 188-Genova; 213-Bologna; 249-Firenze; 350-Roma; 409-Napoli; 457-Bari; 581-Palermo; 594-Messina; 628-Catania.

relationship between G and D can be weakened by the fact that more districts means more occupied locations, thus inducing a greater spread of firms over space and, consequently, a reduction of the Gini index. This effect could generate a measurement bias when running a correlation between G and D. However, results do not change if a binary variable detecting the presence/absence of industrial districts is considered instead of D. Call D' a variable such that $D' = 1$ if the sector contains at least one industrial district and $D' = 0$ otherwise. Then, the rank correlation corrected by ties between G and D' yields a Spearman index $\rho = 0.216$ with a *p-value* of 0.163, which makes it not different from zero at 95% significance. The simple message standing behind this statistics is that spatial concentration and industrial clustering are two separate and unrelated phenomena, thus possibly driven by different underlying forces. In particular, their discrepancy has to do with cities. To see this run the same correlation but using the Gini index G' computed across all locations *except* metropolitan areas. Now the Spearman rank correlation corrected by ties between G' and D yields a value of 0.3, having a *p-value* of 0.051. The correlation becomes statistically significant. The same is true for the rank correlation between G' and D', which yields a value of 0.3, having a *p-value* of 0.051.

To repeat, although these simple findings do not bear any strong implication *per se*, they suggest the need to distinguish those agglomeration dynamics related to urbanization from the ones linked to industrial clustering. Both phenomena generate spatial concentration, but their underlying drivers are different.

4 A model of firm location

The present analysis is based on the discrete choice model described in Bottazzi and Secchi (2007). The model assumes 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 their net growth rate in any sector is in general one order of magnitude smaller than their gross entry/exit rate. So the sectoral dynamics in space is mainly a "reallocation" of activities. Given these considerations, the basic mechanism of the model is very simple.

Consider a single sector j composed by N firms. At each time step, one of them is randomly chosen to die and make room for a new entrant, which will have to choose one location among the L that compose space. Each location is characterized by an 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 terms: the first, a_l, is constant and captures the "fixed" advantages offered by the location (higher demand, lower marginal costs, better infrastructures, etc.); the second, b, is proportional to the number of firms of the sector already located there. Due to the heterogeneous preferences of firms (see Bottazzi and Secchi (2007) for more details), the entrant's probability of locating in location l is

$$p_l \sim a_l + b\, n_l \ , \ b > 0 \tag{1}$$

where n_l is the number of firms belonging to the sector already located in l. Notice that b does not depend on the location l and is constant: this is intended to capture the presence of agglomerative forces acting on the firms of the same sector.

Bottazzi and Secchi (2007) find that the equilibrium distribution of firms across locations $\boldsymbol{n} = (n_1, \ldots, n_L)$ implied by the previous model is generally characterized by the Polya form

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

where $\boldsymbol{a} = (a_l, \ldots, a_L)$ are the geographic attractiveness of the L locations and $A = \sum_l a_l$.

In the specific case of null agglomeration economies (that is $b = 0$), the equilibrium distribution of firms across locations $\boldsymbol{n} = (n_1, \ldots, n_L)$ has the multinomial form

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

Equations (2) and (3) represent short-run equilibrium distributions that depend on the sector-specific geographic attractiveness of the different locations a_l. 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.

5 Maximum likelihood estimation and marginal effects

Assume that each location is characterized by a set of H variables $\boldsymbol{x}_l = (x_l^1, \ldots, x_l^H)$. The idea is to use these variables to model the geographic attractiveness of the different locations.

In the simpler case of null agglomeration economies, 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, $a_l = c(\boldsymbol{\theta}, \boldsymbol{x}_l)$, depending on a set of parameters $\boldsymbol{\theta}$ to be estimated. With simple substitution, the log-likelihood of the observed distribution of firms across locations as a function of the parameters reads

$$\log \pi = \log N! - \sum_{l=1}^{L} \log(n_l!) + \sum_{l=1}^{L} n_l \left(\log c_l - \log C\right) \ , \tag{4}$$

where $C = \sum_{l=1}^{L} c(\boldsymbol{\theta}, \boldsymbol{x}_l)$. One can maximize the previous expression and obtain ML point estimates for the parameters $\hat{\boldsymbol{\theta}}$. In turn, these estimates define a geographic attractiveness coefficient $\hat{c}_l = c(\hat{\boldsymbol{\theta}}, \boldsymbol{x}_l)$ for each location.

In the case of positive agglomeration economies, 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 $\frac{a_l}{b} = c_l = c(\boldsymbol{\theta}, \boldsymbol{x}_l)$. Then, the log-likelihood of the observed distribution of firms reads

$$\log \pi = \log N! - \sum_{l} \log n_l! + \sum_{l=1}^{L} \sum_{k=0}^{n_l - 1} \log(c(\boldsymbol{\theta}, \boldsymbol{x}_l) + k) - \sum_{k=0}^{N-1} \log(C + k) \ . \tag{5}$$

Again, one can maximize the previous expression and obtain ML point estimates for the parameters $\hat{\boldsymbol{\theta}}$ and the coefficients $\hat{c}_l$.

Assessing the effect of the H regressors and of the agglomerative pull b through the estimation of the parameters $\hat{\boldsymbol{\theta}}$ 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 $\frac{a_l}{b}$ automatically induces parametric redundancy in the model. In the present analysis we will characterize the impact of the different covariates using the notion of marginal effect, as often done in discrete models.[3] We proceed as follows.

To begin with, notice that once the estimates $\hat{\boldsymbol{\theta}}$ 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

[3]For an alternative solution based on the choice of a suitable functional specification for c see Bottazzi et al. (2007).

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

$$p_l = \frac{bn_l + a_l}{bN + A} = \frac{n_l + c_l}{N + C} \tag{7}$$

in the Polya case.

Then, the marginal effect of a given variable can be obtained by considering the elasticity of the probability p_l to its variation:

$$\sum_{l=1}^{L} \frac{\partial p_l}{\partial \log x_l^h} = \sum_{l=1}^{L} x_l^h \frac{\partial p_l}{\partial x_l^h} . \tag{8}$$

Notice that this represents a weighted average, across locations, of the marginal effect that the variable exerts in the "total pull" of each location. Locations with higher level of the variable weight more than locations with lower levels.[4] Equation (8) simplifies to

$$\sum_{l=1}^{L} \frac{\partial p_l}{\partial \log x_l^h} = \sum_{l=1}^{L} \frac{x_l^h \partial_h a_l}{A} (1 - p_l) \tag{9}$$

for the multinomial model and

$$\sum_{l=1}^{L} \frac{\partial p_l}{\partial \log x_l^h} = \sum_{l=1}^{L} \frac{x_l^h \partial_h c_l}{N + C} (1 - p_l) \tag{10}$$

for the Polya model. In this latter case, the agglomerative strength exerted by the externality coefficient b can be measured with the elasticity of p_l to the location of an additional firm:

$$\sum_{l=1}^{L} \frac{\partial p_l}{\partial \log n_l} = \sum_{l=1}^{L} \frac{n_l}{N + C} (1 - p_l) , \tag{11}$$

Notably, the presence of the factor $(1 - p_l)$ in (9), (10) and (11) indicates that the marginal effects decrease when $p_l \to 1$. 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 is increasing in the probability p_l. In this case, the marginal effects are

$$\frac{\partial q}{\partial \log x^h} = \sum_{l=1}^{L} \frac{x_l^h \partial_h a_l}{A} . \tag{12}$$

for the multinomial model and

$$\frac{\partial q}{\partial \log x^h} = \sum_{l=1}^{L} \frac{x_l^h \partial_h c_l}{N + C} . \tag{13}$$

[4] Alternatively, one can consider the average marginal effect $\sum_{l=1}^{L} \frac{\partial p_l}{\partial x_l^h} / L$. Since our regressors are positive (see next Section) we prefer the former specification.

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

$$\frac{\partial q}{\partial \log n} = \frac{N}{N+C} \ . \tag{14}$$

Notice that this expression does not depend explicitly on the functional form of c: the strength of sector-specific positive externalities is measured via a combination of the number N of firms in the sector and the sum of the estimated local attractiveness C. Being $C = \frac{A}{b}$, greater values of C are associated with a predominance of the total attractiveness of locations (i.e. A) over the effect of sector-specific positive externalities (i.e. b); in turn, the marginal effect of an additional firm decreases as C grows. In other terms, the larger the parameter C, and consequently the smaller the marginal effect of additional firms, the lesser the locational choice of a firm is actually affected by the choices of others.

A final warning is mandatory. In general the specification should be such that $c(\beta, x_l)$ is positive for any l. Notice that when $c \to 0^+$ the log-likelihood becomes minus infinity, so that the maximization itself should automatically avoid this point. Since we rely on numerical methods, however, the fulfillment 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 we will pay particular attention to avoid this occurrence.

6 Model specification and results

The multinomial and Polya models introduced in the previous Section require the specification of a functional form which relates the values of the variables $\boldsymbol{x}_l$ in one location to its attractiveness. In what follows we will use the following log-linear specification

$$c(\beta, x_l) = \exp\left(\sum_h \beta_h \log(x_{h,l}) + \beta_0\right) \ . \tag{15}$$

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

The specification in (15) is equivalent to the Cobb-Douglas functional form used in consumer and producer theory. Indeed from the previous expression one has

$$c(\beta, x_l) = \prod_h x_{h,l}^{\beta_h} \ \exp(\beta_0) \ .$$

This expression describes the attractiveness c_l as the accumulated multiplicative effect of the different variables. Assume that the different variables represent different economic aspects of the firm activity that the firm evaluates in order to chose were 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 as roughly independent, the combined probability of the firm to choose this location is given by the expression in (15). Moreover, the log-linear specification of c allows for a straightforward computation of the marginal effects defined in (12) and (13). Indeed one has

$$\frac{\partial q}{\partial \log x^h} = \sum_{l=1}^{L} \hat{\beta}^h \ ,$$

for the multinomial model and

$$\frac{\partial q}{\partial \log x^h} = \sum_{l=1}^{L} \frac{\hat{\beta}^h C}{N + C} ,$$

for the Polya model.

Starting from the functional specification in (15) we consider a model with three regressors. First, we capture the pull of more populated area, i.e. the urbanization effect, by taking the total population of the location (POP). In this respect, our definition of urbanization forces captures all the effect related to the sheer number of inhabitants, like the increased demand for final goods and the availability of a larger labor market.

Then we aggregate, in each location, the firms belonging to the sectors appearing in Table 1 in two groups, manufacturing firms ($MANUF$) and service firms ($SERV$): $MANUF$ includes the sectors ranging from "15-Food products" to "36R-Residual of sector 36", while $SERV$ includes those from "40-Electricity and gas" to "74-Other business activities". Notice that the latter contains services to both local population and local firms. These variables are used to construct two further regressors. First, the relative abundance of services, computed as the ratio $SERV/POP$, is taken as a generic proxy of local economic development. In this respect the inclusion of the construction sector and of the utilities (distribution of electricity, gas and water) in the definition of services accounts for the availability of infrastructural resources. Second, a measure of local specialization is obtained by considering the ratio $MANUF/SERV$. This regressor can account for the existence of vertical relations in the production chain or of cross-sectoral externalities among firms belonging to similar sectors. With this choice we end up with three variables having very low correlation across sectors: Spearman's correlation coefficients are always lower than 10% and never significant. The orthogonality of the regressors will increase the estimation efficiency and is compatible with the choice of the log-linear form (15). With these considerations we end up following the specification

$$\log c(\beta, x_l) = \beta_0 + \beta_1 \log \mathrm{POP} + \beta_2 \log \frac{\mathrm{SERV}}{\mathrm{POP}} + \beta_3 \log \frac{\mathrm{MANUF}}{\mathrm{SERV}} , \tag{16}$$

where the constant term β_0 is set to 0 in the multinomial case while being regularly estimated in the Polya case. Estimation is performed taking the z-score of the three regressors. In this way the constant term $\exp(\beta_0)$ represents the average effect. i.e. the value of the function c associated to a location with an average value of the regressors. The z-scores are obtained by removing the mean and dividing by the standard deviation each covariate. This procedure allows to compare directly the coefficients associated with the different regressors and their marginal elasticities.

We also estimated an augmented six regressors model including, in addition to the variables in (16), the surface of the location, a dummy variable taking value one for locations characterized by the presence of industrial districts (irrespectively of their sector) and a dummy variable taking value one for larger metropolitan area. The inclusion of the location's surface variable is intended to capture the possible existence of congestion effects. The "districts" dummy captures the knowledge spillover effect generated by a particularly high concentrations of certain economic activities. The "urban" dummy captures plausible effects exerted by larger metropolitan area on the local demand. With the only exception of the urban dummy for sector "61-Water transportation", none of these variables resulted significant in none of the sectors under analysis for both the multinomial and the Polya specification. Hence they were dropped from the analysis and we will discuss the reduced model.

The model (16) was estimated on each sector separately.[5] However, we wanted to avoid

[5]The values of the regressor is slightlty different when different sectors are considered because the variable $MANUF$ and $SERV$ do not contain the firms of the sector under consideration.

Sector	POP	$\frac{SERV}{POP}$	$\frac{MANUF}{SERV}$	-log-like
15-Food products	1.12e+00 (0.11)	2.09e-03 (0.67)	-5.22e-03 (0.48)	7.75e+00
17-Textiles	1.68e+00 (0.12)	3.56e-01 (0.08)	3.19e-01 (0.15)	3.75e+01
18-Apparel	1.88e+00 (0.11)	7.38e-02 (0.20)	1.96e-01 (0.24)	1.94e+01
19-Leather products	1.32e+00 (0.12)	2.75e-01 (0.10)	2.41e-02 (0.46)	5.22e+01
20-Wood processing	9.94e-01 (0.11)	8.92e-02 (0.18)	9.21e-03 (0.30)	8.99e+00
21-Pulp and paper	2.38e+00 (0.12)	1.71e-01 (0.18)	1.58e-01 (0.33)	2.68e+00
22-Publishing and printing	3.06e+00 (0.13)	1.62e-01 (0.27)	-5.90e-03 (0.96)	3.76e+00
23-Coke,petroleum and nuclearfuel	2.10e+00 (0.13)	-4.79e-03 (2.49)	7.07e-02 (0.39)	1.05e+00
24-Organic and inorganic chemicals	2.87e+00 (0.13)	1.75e-01 (0.22)	3.05e-02 (1.09)	2.71e+00
25-Rubber and plastic products	2.29e+00 (0.12)	1.83e-01 (0.16)	2.12e-01 (0.29)	5.53e+00
26-Non metallic mineral products	1.19e+00 (0.11)	1.88e-02 (0.37)	2.72e-03 (0.74)	1.08e+01
27-Basic metals	2.39e+00 (0.12)	1.31e-01 (0.18)	3.18e-01 (0.25)	2.80e+00
28-Fabricated metal products	1.74e+00 (0.12)	1.18e-01 (0.19)	6.14e-02 (0.40)	1.89e+01
29-Industrial machinery	2.16e+00 (0.12)	3.04e-01 (0.13)	1.11e-01 (0.42)	9.68e+00
30-Office machinery	2.80e+00 (0.12)	8.17e-02 (0.32)	4.29e-03 (4.40)	1.38e+00
31-Electrical machinery	2.58e+00 (0.12)	1.92e-01 (0.19)	5.92e-02 (0.66)	5.20e+00
32-Radio,TV,and TLC devices	2.38e+00 (0.13)	1.29e-01 (0.24)	8.42e-05 (101.41)	2.61e+00
33-Precision instruments	2.38e+00 (0.13)	1.40e-01 (0.23)	-1.18e-03 (8.80)	6.16e+00
34-Motor vehicles and trailers	2.56e+00 (0.12)	1.90e-01 (0.19)	6.61e-02 (0.55)	2.44e+00
35-Other transport equipment	1.55e+00 (0.14)	2.36e-01 (0.14)	7.57e-02 (0.33)	4.88e+00
361-Furniture	1.77e+00 (0.12)	1.71e-01 (0.15)	1.04e-01 (0.35)	2.61e+01
362-Jewlery	1.88e+00 (0.13)	2.20e-01 (0.16)	-4.37e-05 (54.68)	2.10e+01
363-Musical instruments	1.48e+00 (0.12)	2.75e-01 (0.10)	9.75e-02 (0.26)	1.98e+00
45-Construction	1.54e+00 (0.12)	1.30e-01 (0.17)	8.47e-03 (1.44)	2.47e+01
50-Sale and services of motorvehicles	1.66e+00 (0.12)	6.48e-03 (0.51)	4.24e-05 (2.25)	6.05e+00
51-Wholesale and commission trade	2.52e+00 (0.12)	4.37e-02 (0.60)	4.31e-02 (0.50)	2.16e+01
52-Retail trade	1.65e+00 (0.12)	-9.43e-04 (0.41)	1.71e-02 (0.42)	1.54e+01
55-Hotels and restaurants	1.21e+00 (0.12)	2.00e-01 (0.14)	3.50e-02 (0.39)	2.85e+01
60-Land transport	1.86e+00 (0.12)	1.50e-01 (0.18)	6.04e-03 (1.38)	1.10e+01
63-Auxiliary transpor tactivities	2.83e+00 (0.13)	1.18e-01 (0.31)	4.04e-02 (0.66)	5.87e+00
64-Post and telecommunications	1.12e+00 (0.11)	4.44e-02 (0.26)	1.58e-03 (2.12)	4.14e+00
65-Financial intermediation	1.78e+00 (0.12)	1.49e-01 (0.18)	3.07e-03 (1.82)	3.37e+00
66-Private insurance and pensions	2.84e+00 (0.13)	9.37e-02 (0.34)	2.31e-02 (0.75)	1.26e+00
67-Auxiliary financial activities	2.31e+00 (0.13)	1.06e-01 (0.27)	6.63e-04 (7.39)	5.18e+00
70-Real estate activities	2.96e+00 (0.14)	5.02e-01 (0.10)	6.17e-02 (0.54)	1.40e+01
71-Renting of machinery and equipment	1.86e+00 (0.14)	1.95e-01 (0.18)	3.38e-02 (0.56)	3.42e+00
72-Computer and related activities	3.07e+00 (0.13)	1.90e-01 (0.24)	6.99e-04 (14.09)	5.69e+00
74-Other business activities	2.97e+00 (0.13)	1.45e-01 (0.26)	1.84e-03 (2.68)	9.03e+00

Table 2: Maximum likelihood estimation of the multinomial log-linear model in (4) and (16). Standard errors (in parentheses) are reported as percentage of the point estimate. The last column reports the negative log likelihood: lower level are associated with better agreement.

Sector	POP	$\frac{SERV}{POP}$	$\frac{MANUF}{SERV}$	n	-log-like
15-Food products	1.20e-01 (0.12)	9.41e-05 (0.74)	-5.71e-04 (0.48)	8.89e-01 (0.00)	4.51e+00
17-Textiles	1.31e-02 (0.20)	1.29e-03 (0.22)	1.36e-03 (0.29)	9.83e-01 (0.00)	3.99e+00
18-Apparel	3.73e-02 (0.10)	1.57e-03 (0.22)	1.88e-03 (0.31)	9.74e-01 (0.00)	4.33e+00
19-Leather products	4.88e-03 (0.21)	7.94e-04 (0.24)	3.34e-04 (0.36)	9.92e-01 (0.00)	3.21e+00
20-Wood processing	8.81e-02 (0.10)	5.60e-03 (0.21)	1.12e-03 (0.28)	9.07e-01 (0.00)	4.41e+00
21-Pulp and paper	5.53e-01 (0.13)	3.38e-02 (0.19)	3.09e-02 (0.34)	7.60e-01 (0.01)	1.95e+00
22-Publishing and printing	8.70e-01 (0.12)	4.38e-02 (0.26)	-1.08e-03 (0.92)	7.08e-01 (0.01)	3.11e+00
23-Coke,petroleum and nuclearfuel	1.24e+00 (0.14)	-3.05e-03 (2.13)	3.45e-02 (0.44)	4.04e-01 (0.03)	1.01e+00
24-Organic and inorganic chemicals	7.52e-01 (0.11)	4.41e-02 (0.18)	7.19e-03 (1.09)	7.14e-01 (0.02)	2.25e+00
25-Rubber and plastic products	1.76e-01 (0.11)	1.23e-02 (0.15)	1.28e-02 (0.30)	9.09e-01 (0.01)	2.89e+00
26-Non metallic mineral products	7.87e-02 (0.11)	2.68e-04 (0.54)	1.53e-04 (0.74)	9.25e-01 (0.00)	4.12e+00
27-Basic metals	3.34e-01 (0.11)	1.64e-02 (0.20)	2.31e-02 (0.32)	8.36e-01 (0.01)	1.84e+00
28-Fabricated metal products	5.27e-02 (0.11)	3.11e-03 (0.19)	1.13e-03 (0.49)	9.65e-01 (0.00)	5.02e+00
29-Industrial machinery	1.11e-01 (0.13)	1.35e-02 (0.13)	4.80e-03 (0.44)	9.42e-01 (0.00)	3.86e+00
30-Office machinery	1.28e+00 (0.13)	3.38e-02 (0.32)	2.58e-03 (3.46)	5.35e-01 (0.02)	1.22e+00
31-Electrical machinery	2.93e-01 (0.12)	2.13e-02 (0.17)	4.99e-03 (0.76)	8.77e-01 (0.00)	3.04e+00
32-Radio,TV,and TLC devices	1.00e+00 (0.12)	5.31e-02 (0.24)	-1.84e-04 (16.24)	5.72e-01 (0.01)	2.38e+00
33-Precision instruments	3.55e-01 (0.13)	1.92e-02 (0.24)	-2.39e-04 (5.91)	8.49e-01 (0.00)	3.29e+00
34-Motor vehicles and trailers	2.64e-01 (0.18)	1.68e-02 (0.22)	1.04e-02 (0.41)	8.63e-01 (0.02)	1.42e+00
35-Other transport equipment	1.10e-01 (0.17)	1.00e-02 (0.20)	1.83e-03 (0.57)	9.23e-01 (0.01)	2.09e+00
361-Furniture	3.07e-02 (0.13)	2.71e-03 (0.19)	5.80e-04 (0.54)	9.74e-01 (0.00)	4.10e+00
362-Jewlery	3.29e-02 (0.22)	1.45e-03 (0.40)	-3.30e-07 (90.33)	9.70e-01 (0.01)	2.78e+00
363-Musical instruments	2.55e-01 (0.26)	4.65e-02 (0.29)	4.14e-03 (0.54)	8.29e-01 (0.04)	8.40e-01
45-Construction	3.59e-02 (0.11)	3.09e-03 (0.17)	1.68e-04 (1.57)	9.76e-01 (0.00)	6.17e+00
50-Sale and services of motorvehicles	2.87e-01 (0.12)	1.16e-03 (0.51)	5.12e-06 (2.33)	8.25e-01 (0.00)	4.59e+00
51-Wholesale and commission trade	6.66e-02 (0.14)	1.21e-03 (0.59)	1.11e-03 (0.48)	9.73e-01 (0.00)	5.59e+00
52-Retail trade	7.25e-02 (0.12)	-6.68e-05 (0.42)	7.58e-04 (0.41)	9.56e-01 (0.00)	6.09e+00
55-Hotels and restaurants	2.87e-02 (0.11)	4.34e-03 (0.15)	8.01e-04 (0.38)	9.76e-01 (0.00)	5.88e+00
60-Land transport	1.15e-01 (0.12)	8.94e-03 (0.17)	3.40e-04 (1.39)	9.36e-01 (0.00)	4.94e+00
63-Auxiliary transpor tactivities	3.62e-01 (0.13)	1.44e-02 (0.31)	4.54e-03 (0.69)	8.65e-01 (0.00)	3.48e+00
64-Post and telecommunications	2.82e-01 (0.10)	1.18e-02 (0.25)	3.22e-04 (2.46)	7.35e-01 (0.01)	3.43e+00
65-Financial intermediation	9.00e-01 (0.12)	7.48e-02 (0.18)	1.50e-03 (1.83)	4.91e-01 (0.01)	3.21e+00
66-Private insurance and pensions	1.52e+00 (0.14)	4.48e-02 (0.34)	9.75e-03 (0.82)	4.46e-01 (0.03)	1.20e+00
67-Auxiliary financial activities	4.23e-01 (0.14)	1.88e-02 (0.29)	1.62e-04 (5.93)	8.16e-01 (0.00)	3.95e+00
70-Real estate activities	1.04e-01 (0.13)	1.72e-02 (0.11)	1.85e-03 (0.58)	9.63e-01 (0.00)	4.47e+00
71-Renting of machinery and equipment	5.32e-01 (0.14)	5.05e-02 (0.19)	6.88e-03 (0.65)	7.15e-01 (0.01)	2.79e+00
72-Computer and related activities	4.72e-01 (0.12)	2.94e-02 (0.23)	1.17e-04 (12.21)	8.43e-01 (0.00)	3.95e+00
74-Other business activities	2.46e-01 (0.11)	1.17e-02 (0.24)	1.19e-04 (2.73)	9.15e-01 (0.00)	4.86e+00

Table 3: Maximum likelihood estimation of the Polya log-linear model in (5) and (16). Standard errors (in parentheses) are reported as percentage of the point estimate. The last column reports the negative log likelihood: lower level are associated with better agreement.

sectors characterized either by a strong presence of publicly owned enterprises or by a decisive dependence from public infrastructures. For this reason “40-Electricity and gas”, “41-Water”, “61-Water transport” and “62-Air transport” were dropped from the analysis. Moreover, we ignored the residual sector “36R-Residual of sector 36”, since it actually includes fairly inhomogeneous activities. Table 2 and Table 3 report the estimated marginal elasticities of the three regressors considered for the multinomial and Polya specification respectively; for the latter case, Table 3 reports also the marginal elasticity of additional firms on the probability for the location to attract settlements within the sector. Standard errors are obtained via Monte Carlo re-sampling of the original distribution and reported as a percentage of the point estimate. Values larger than 20% identify non-significant regressors. In any case, variables having values larger than 15% should be regarded as only mildly significant. Finally, the last columns report the model maximum log-likelihood (per observation). Since the number of observation is large (686) and the differences in the number of parameters minimal (1), these numbers can be taken as a relative measure of the goodness of fit of the two models.

Let us start by inspecting the results for the multinomial model reported in Table 2. Remember that this model does not contain any account for the existence of sector-specific positive externalities. Nonetheless, a number of conclusions can be drawn.

First, the urbanization effect is very strong. The marginal elasticities with respect to POP are always positive and significant. Addition of a small portions of population increase, on average, the attractiveness of locations. This being the rule across almost all sectors, it is possible to infer that higher degrees of urbanization end up attracting “more of everything” with respect to economic activities, although the size of such effect appears heterogeneous across sectors: it is, for instance, three times larger for sector “22-Publishing and printing” and sector “72-Computer and related activities” than for sector “20-Wood processing”.

Second, our proxy for local development plays only a mild role. In a minority of the sectors under investigation, additional small portions of services per head have a limited positive effect on the attractiveness of a location. They are however totally irrelevant for the majority of them. In fact the marginal elasticities connected to $SERV/POP$ is significantly different from zero only in few sectors (like “17-Textiles” or “70-Real estate activities”), and even in those cases they turn out being about ten times smaller than the marginal elasticities connected to POP.

Third, specialization does not matter. The marginal elasticities connected to $MANUF/SERV$ is never significantly different from zero. Yet, it must be stressed that this finding does not deny *per se* the role of vertical or horizontal linkages among firms: rather, it detects the economic irrelevance of such linkages *at this level* of sectoral disaggregation. Such conclusion may well be reverted if a finer sectoral disaggregation were used.

Moving to the Polya model, whose results are reported in Table 3, the first thing to notice is that the qualitative effect of the different regressor remains the same. So what was said for the multinomial model is still true when one accounts for sector-specific positive externalities. This extension is however not trivial nor neutral. In fact, *sector-specific* positive externalities appear as the rule in the economy. Comparing the negative log-likelihoods for the Polya model in Table 3 with those for the multinomial model in Table 2 reveals that the former model fits systematically better than the latter. Moreover, the marginal elasticity with respect to n is always positive and significantly different from zero. This suggests not only the relevance of sector-specific positive externalities, but also the irrelevance of congestion effects. This latter aspect is particularly striking given the fine level of spatial disaggregation granted by the use of LLSs. Therefore, those technological and organizational factors that are the likely cause of sector-specific positive externalities should not be regarded as a “special case” neither for what concerns the various economic activities nor for what concerns their location. To the contrary, positive externalities of location pervade the whole economy and seem to start as soon as the

Class	Range of localized business units
C_0	0
C_1	1-2
C_2	3-6
C_3	7-14
$\vdots$	$\vdots$
C_{15}	32767-65534

Table 4: Occupancy classes

borders of a market exceed those of a town, which is precisely the spatial dimension captured by LLS.

A more direct comparison between the urbanization effect and the sector-specific externalities can be obtained by comparing the marginal elasticities relative to n with those relative to POP in Table 3. The latter outperform the former in most sectors (precisely 31 out of 38). For the most disaggregated sectors, "362-Jewelry" and "363-Musical instruments", the urbanization effect turns out to be irrelevant (marginal effect non significantly different from zero) and the self-reinforcing effects generated by the co-location of firms belonging to the same sector remains as the unique explanatory variable. As expected, the relevance of the sectoral effect gets enhanced when a finer disaggregation is considered. However, notice that this cannot be a general pattern, otherwise the specialization regressor $MANUF/SERV$ would result significant.

The improved fit of the Polya model has two implications: on one hand, it confirms that positive externalities play a crucial role not only in some special sectors but rather across the entire economy; on the other hand, it implies that a general account of the spatial structure of economic activities cannot prescind from sector-specific technological and organizational considerations. In other words, non-pecuniary positive externalities may perhaps not leave a "paper trail" but they certainly leave a strong footprint on the spatial distribution of firms, possibly more than market-mediated forces do.

Finally, although the Polya model outperforms systematically the multinomial one, its absolute goodness of fit remains to be assessed. A possible method consists in comparing, for each sector, the predicted spatial distribution stemming from the Polya model with the one observed in the data. To accomplish such comparison we construct histograms that synthesize the relevant information into proper occupancy classes.

Occupancies $f(n)$ count the number of locations hosting exactly n firms. For example, $f(0)$ is the number of locations containing 0 firms belonging to the sector under scrutiny; $f(1)$ is the number of locations hosting exactly 1 firm, and so on. Then, the general definition of occupancy is

$$f(n) = \sum_{l=1}^{L} \delta_{n_l,n} \tag{17}$$

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

$$\sum_{n=0}^{+\infty} f(n) = L \quad \forall j \tag{18}$$

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

$$C_k = [2^k - 1, 2^{k+1} - 2) \quad k = 0, 1, 2, \cdots \tag{19}$$

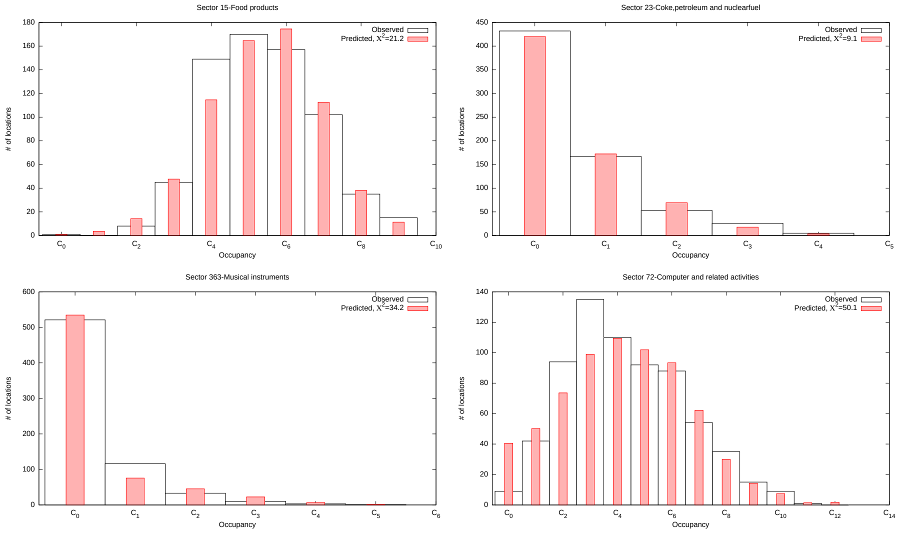

Figure 2: Spatial distributions predicted by the Polya model against observed ones.

So occupancy classes are

$$f(C_k) = \sum_{n \in C_k} f(n) \tag{20}$$

To summarize, Table 4 reports the range values for some of the occupancy classes.

From occupancy classes it is immediately possible to grasp whether the sector under scrutiny is spatially concentrated or dispersed (see Figure 2). An high occupancy of the first classes compared to the latter indicates that most locations host null or minimal portions of the sector, while a few others have big shares. Therefore, the sector is spatially concentrated. To the contrary, a more bell-shaped histogram indicates that locations tend to host similar "typical" portions of the sector, which is then more uniformly distributed in space.

As exemplified in Figure 2, the absolute goodness of fit of the Polya model varies across sectors. In some cases, the predicted spatial distributions match very closely the observed ones, while in other cases the discrepancy is greater. This heterogeneity in the goodness of fit might depend on a series of factors. First, the specifications adopted in (15) and (16) may work well in some sectors and less so in others; consequently, alternative specifications could possibly improve the performance of the Polya model. Second, the variables used to estimate the coefficients in (16) may not be the best explanators of the ratio a_l/b, which weights the "fixed" attractiveness of a location relative to the sector-specific externalities of the sector. Third, the level of sectoral disaggregation that is necessary to characterize properly a sector in geographic terms may vary across sectors, in particular when services are analyzed next to manufacturing (see Section 3). Naturally, this set of considerations about the goodness of fit of the Polya model create room for future empirical efforts addressed at the further amelioration of predictive performances.

7 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 positive externalities were found to affect significantly the location of firms. However, and third, externalities were often found to be more relevant than urbanization motives. According to these considerations, an explanation of the spatial structure of economic activities must rely primarily on the 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, our conclusion about the relevance of sector-specific positive externalities is sharply in contrast with the one of Kim (1995) concerning manufacturing in the US; therefore, it would be interesting to apply our framework on US data, as well as on other countries, to detect how the results would possibly change. Further developments should regard also 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 not only to test the "robustness" of the model but also to infer the geographic range of the various effects under scrutiny.

References

Becattini, G. (1990). The marshallian industrial district as socio-economic notion. In F. Pyke, G. Becattini, and W. Sengenberger (Eds.), *Industrial districts and inter-firm cooperation in Italy*, pp. 37–51. International Institute of Labour Studies.

Bottazzi, G., G. Dosi, G. Fagiolo, and A. Secchi (2007, September). Modeling industrial evolution in geographical space. *Journal of Economic Geography 7*(5), 651–672.

Bottazzi, G., G. Dosi, G. Fagiolo, and A. Secchi (2008). Sectoral and geographical specificities in the spatial structure of economic activities. *Structural Change and Economic Dynamics 19*(3), 189–202.

Bottazzi, G. and A. Secchi (2007). Repeated choices under dynamic externalities. LEM Working Paper 2007-08, Scuola Speriore Sant'Anna.

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

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

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

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

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

ISTAT (1997). I sistemi locali del lavoro 1991.

ISTAT (2006). Atlante statistico dei comuni italiani.

Kim, S. (1995). Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987. *The Quarterly Journal of Economics 110*(4), 881–908.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy 99*(3), 483–499.

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

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

Overman, H. and G. Duranton (2001). Testing for localization using micro-geographic data. *CEPR Discussion Paper Series* (3379).

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