Sectoral and Geographical Specificities in the Spatial Structure of Economic Activities

Giulio Bottazzi*
Giovanni Dosi*
Giorgio Fagiolo*
Angelo Secchi*

*Scuola Superiore Sant'Anna, Pisa

2004/21 March 2007

Sectoral and Geographical Specificities in the Spatial Structure of Economic Activities[*]

Giulio Bottazzi[†] Giovanni Dosi Giorgio Fagiolo Angelo Secchi

Sant'Anna School of Advanced Studies, Pisa, Italy

March 16, 2007

Abstract

This work explores the spatial distribution of productive activities in the Italian manufacturing industry. We propose an econometric model which tries to disentangle location-specific from sectoral drivers in the dynamic process of spatial agglomeration. The basic idea is that the former typically apply "horizontally" (i.e. across all industrial sectors), while the latter unfold in the form of non-decreasing dynamic returns to the current stock of installed business units. Three different specifications of the model are tested against Italian data on the location of manufacturing activities, studying the distribution of the number of firms and employees. Our results suggest that different locations exert different structural influences on the distribution of both variables. Moreover, a significant horizontal power of "urbanization", which makes some locations, especially metropolitan areas, more attractive irrespectively of the sector, does emerge. However, after controlling for the latter, one is still left with very significant sector-specific forms of dynamic increasing returns to agglomeration, which vary a lot across different manufacturing activities and which plausibly have to do with sectoral-specific and localized forms of knowledge accumulation and spin-offs.

JEL codes: C1, L6, R1

Keywords: Industrial Location, Agglomeration, Markov Chains, Dynamic Increasing Returns.

---

[*] We gratefully acknowledge the support to the research by the Italian Ministry of Education, University, and Research (MIUR, Project EFIRB03GD), the Scuola Superiore Sant'Anna (Grant ERIS03GF) and The University of Manchester. Part of the research that has lead to this work has been undertaken as part of the activities of the DIME Network of Excellence, sponsored by the European Union. Comments by Thomas Brenner, Yannis Ioannides, Karl Schlag, Alan Kirman and two anonymous referees have helped along the various revisions of the work.

[†] Corresponding author. Mail address: Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33, I-56127 Pisa, Italy. Tel: +39-050-883365 Fax: +39-050-883344. Email: `giulio.bottazzi@sssup.it`

1 Introduction

This work studies the structure of the statistical distribution of economic activities in the geographical space. In particular, we propose different econometric exercises, based on the stochastic Markov model of firm location developed in Bottazzi et al. (2007), aimed to disentangle two distinct classes of agglomeration drivers: “location-specific” drivers, which “horizontally” cut across different types of economic activities and “technology-specific” drivers, whose effect changes across different lines of production.

The ways economic activities are distributed over geographical space along relatively ordered patterns has been a concern for economic analysis at least since Alfred Marshall. Indeed, the first basic stylized fact of economic geography is that locational patterns, over the whole history for which we have some records throughout the world, tend to be much more clustered than any theory of comparative advantage might predict (cf. Krugman (1991) and Fujita et al. (1999), among many others).

At the same time, the evidence suggests a remarkable inter-sectoral variability in agglomeration structures. This applies across different countries such as the US, France, the UK, Germany and Italy: cf. Ellison and Glaeser (1997), Maurel and Sedillot (1999), Devereux et al. (2004), Overman and Duranton (2002) and Brenner (2003). That same evidence hints also at diverse degrees of “attractiveness” of different locations. So, for instance, there are several locations where business units belonging to almost all sectors are equally represented. On the contrary, in many other sites, agglomeration occurs only for business units belonging to a small number of sectors (in some cases, one or two). For example, as discussed in Bottazzi et al. (2006), in the Italian case a quite large fraction of sectors is not even represented in more than 50% of locations. Moreover, any measure of agglomeration appears to be quite stable over time, notwithstanding the great variability of agglomeration observed across locations and a turbulent underlying micro-dynamics with persistent flows of entry, exit, and variation in the relative sizes of incumbents (Dumais et al., 2002). Taken together, the foregoing pieces of evidence suggest a general picture characterized by different drivers of agglomeration, which might be economy-wide, location-specific or sector-specific.

More specifically, acknowledging the heterogeneous nature of the different agglomeration forces, in this work we investigate the relative role of location-specific mechanisms of agglomeration, independent of individual sectors and technologies *vs.* sector-specific drivers of agglomeration (or dispersion) of economic activities, applying across different locations within similar ensembles of production activities. The idea behind the present analyses is that cross-sectoral differences in agglomeration forces ought to be, at least partly, explained on the grounds of underlying differences in the relative importance of phenomena such as localized knowledge spillovers; inter- vs. intra-organizational learning; knowledge complementarities fueled by localized labor mobility; innovative explorations undertaken through spin-offs, and, more generally, the birth of new firms.

The proposed econometric exercises are different specifications of the simple stochastic model developed in Bottazzi et al. (2007). This model is built upon the idea of dynamic increasing returns and shares its general structure and several hypotheses with the models explored by Arthur (1994), Dosi et al. (1994) and Dosi and Kaniovski (1994). However, in order to obtain empirically testable predictions, instead of the irreversible pure-birth dynamics characterizing those models, we consider a Markov dynamics where the reversibility of locational choices by firms entails a notion of stochastic equilibrium (i.e., invariant limit distribution). Bottazzi and Secchi (2007) show that this equilibrium, under rather general hypothesis about the selection mechanism characterizing a heterogeneous population of agents, is equivalent to the Ehrenfest-Brillouin urn-scheme (cfr. Garibaldi and Penco (2000) and Garibaldi et al. (2002)). Building on this notion of dynamic equilibrium characterizing the spatial distribution of "productive units", which can be either plants or unit of employment, we obtain, under different assumptions, three different statistical models that we estimate using Italian data, disaggregated by "locational units" and by sector.

Let us illustrate the intuition behind our analysis borrowing from the "dartboard" metaphor in the seminal work by Ellison and Glaeser (1997), with which the following has indeed several points in common. Suppose that the economic space is a sort of dartboard where darts of different colors are thrown (that is, economic activities belonging to different sectors are located). Here, the null hypothesis (i.e. "agglomeration does not matter") is a distribution of darts on the board solely due to random factors. In departing from pure randomness, however, one might observe systematic patterns ultimately due to three different factors. The first one has to do with the generic attractiveness (or repulsiveness) of some areas on the board: hence, one will systematically find there more (or less) darts *of all* colors than what sheer randomness would predict. That is, to trivialize, one will find "more of everything" in New York as compared to Pisa, irrespectively of any finer pattern of comparative advantage. Second, on the top of these generic locational patterns, one may observe specific patterns *distinctive of any one color* (that is, sectoral specificities). Finally, the last concerns the different *size* of different darts (that is, different degrees of lumpiness of single investments).

Ellison and Glaeser (1997, 1999) and Dumais et al. (2002) control for the latter, as captured by the concentration in plant size distribution, and study the importance of sector-specific agglomeration factors as compared to inter-sectoral, location-wide, ones (which they call "natural advantage" of a location).[1]

Our exercise largely shares a similar spirit, albeit with some distinct features. Indeed, we do not "wash out" any lumpiness effect. We do it partly out of necessity and partly out of choice. The constraint is that given our small spatial units (defined in terms of local labor mobility basins, typically smaller than most US counties) and our fine-grained sectoral

---

[1] Refinements and applications of this basic methodology are in Maurel and Sedillot (1999), Devereux et al. (2004) and Overman and Duranton (2002). See also the detailed reviews in Combes and Overman (2004) and Ottaviano and Thisse (2004).

partition, it is very hard to find the relevant sectoral/spatial breakdown of the data. At the same time, at a conceptual level, it is not entirely uncontroversial that one should take out the "size effect". In order to see this, think of, say, five entities located in one particular place which at some point merge into one. This does not mean that agglomeration has fallen, but rather that whatever forces driving agglomeration have now been internalized within a single firm. Thus, complementary information may be usefully obtained by studying, side by side, the agglomeration dynamics in terms of number of firms and of employment units.

The rest of the paper is organized as follows. After a brief description of our data, in Section 3 we present the basics of the stochastic model derived from Bottazzi et al. (2007) which constitutes the conceptual framework for the econometric specifications discussed in Section 4. In Section 5 we test these specifications against data on locational patterns of different sectors of the Italian manufacturing industry, using both firms and workers as proxy for production units. Finally, Section 6 concludes.

2 Data

This research draws upon the "Census of Manufacturers and Services", a database developed by the Italian Statistical Office (ISTAT) that contains observations about five millions employees and more than half a million business units (BUs).[2] Each observation identifies the location of the employees and of the business units at a given point in time (1996), as well as the industrial sector which they belong to. We consider data disaggregated according to the Italian ATECO classification (which corresponds to the NACE classification system). Among all industries, we focus on the manufacturing segment excluding, however, the sector "16 - Tobacco products" which presents a too limited number of business units.

Business units and employees are classified with respect to 784 geographical locations. Each geographical location represents a "local system of labor mobility" (LSLM), that is a geographical area characterized by relatively high internal labor commuters' flows. LSLMs are periodically updated by multivariate cluster analyses employing census data about social, demographic, and economic variables (see Sforzi (2000) for details). Table 1 reports for each sector a brief description of the occupancy distribution of employees and business units across sites.

---

[2] Incidentally note that in the Italian case in more than 88% "business units" and "firms" coincide.

	Statistics of the Occupancy Distribution									
	Business Units					Employees				
Sector	Number	Mean	Std Dev	Min	Max	Number	Mean	Std Dev	Min	Max
15 Food products	75420	96.2	170.2	1	1854	434515	554.2	1254.2	4	20673
17 Textiles	36217	46.2	262.4	0	6675	345338	440.5	1980.5	0	38667
18 Apparel	49782	63.5	179.3	0	2297	346387	441.8	1036.4	0	9036
19 Leather products	25451	32.5	145.7	0	2311	230543	294.1	1282.1	0	17502
20 Wood processing	50662	64.6	119.0	0	1728	170294	217.2	405.6	0	3579
21 Pulp and paper	5268	6.7	26.0	0	577	85424	109.0	376.3	0	6943
22 Publishing and printing	28183	36.0	193.1	0	4162	175012	223.2	1549.3	0	35391
23 Coke, refined petroleum and nuclear fuel	825	1.1	3.1	0	45	24147	30.8	218.8	0	4496
24 Organic and Inorganic chemicals	7593	9.7	48.3	0	1197	209242	266.9	1976.7	0	51772
25 Rubber and plastic products	14626	18.7	64.7	0	1364	198401	253.1	909.3	0	17691
26 Non-metallic mineral products	30709	39.2	79.9	0	943	250824	319.9	877.7	0	17173
27 Basic metals	4034	5.1	19.5	0	353	136123	173.6	704.9	0	9843
28 Fabricated metal products	94771	120.9	323.3	2	5576	621642	792.9	2277.0	2	35873
29 Industrial machinery and equipment	42984	54.8	176.7	0	3605	554105	706.8	2447.4	0	46634
30 Office machinery	592	0.7	4.5	0	94	18609	23.7	257.4	0	6454
31 Electrical machinery	17312	22.1	91.5	0	2055	205797	262.5	1390.8	0	33261
32 Radio, TV and TLC devices	9773	12.5	48.8	0	980	103161	131.6	942.3	0	23064
33 Precision instruments	28280	36.1	142.0	0	2808	129448	165.1	834.1	0	17699
34 Motor vehicles and trailers	2261	2.9	12.8	0	297	185748	236.9	2186.6	0	57705
35 Other transport equipment	4514	5.8	17.5	0	166	100780	128.5	635.4	0	11525
36 Furniture	59627	76.1	257.8	0	4040	309911	395.3	1372.2	0	20509
37 Recycling	2061	2.6	7.5	0	105	8327	10.6	32.6	0	510

Table 1: Descriptive statistics of the firm occupancy distribution by sector in 1996.

3 A stochastic model of location with dynamic increasing returns

As discussed in the introduction, the aim of this paper is to describe the spatial distribution of economic activities among different locations in the attempt of disentangling location-specific mechanism of agglomeration, independent of individual sectors and technologies, from the sector-specific drivers of agglomeration, applying across different locations within similar ensembles of production activities. To this aim, in what follows, we propose a series of econometric exercises rooted in a stochastic model of location built upon the idea of dynamic increasing returns. To set the stage, in this section we briefly present the basic skeleton of the model.

The model considers a single-sector economy composed by a fixed number of location, L, which can be thought as production sites, and populated by a constant number, N, of heterogeneous agents representing different production units. Agents, which are assumed to be boundedly-rational profit seekers, have to choose where to locate themselves among the set of available locations. The sequence of locational choices by agents is described as a stochastic process: at each time step an agent is chosen at random to die (i.e. to leave the location where it operates) and, once the exit took place, a new agent enters the economy selecting as productive location the one which maximize his expected utility. The possibility that agents posses heterogeneous preferences and beliefs is introduced by assuming that the expected return associated to different locations posses a common component and an individual, idiosyncratic, one. In turn, the common component is characterized by a constant term which describes the intrinsic "geographic attractiveness" of each locations and by a "social term" which depends on the actual distribution of agents across different locations and captures the strength of agglomeration forces. Bottazzi and Secchi (2007) show that, under rather general assumptions about agents' preferences structure, their locational choices are, in probability, driven exclusively by the common component of the expected individual return. Assuming a linear form for the social term, the new entrant chooses location $l \in \{1, \ldots, L\}$ with probability

$$p_l \sim a_l + b_l n_{l,t-1} \tag{3.1}$$

where $n_{l,t-1}$ is the number of agents present in that location at the end of the previous time step. The coefficient a_l represents the geographical attractiveness of location l and captures the gain that an agent on average expects by choosing to locate its activity in a given site irrespectively of the choices of other agents. This coefficient might be interpreted as controlling for intrinsic exogenous geographical factors (e.g., cost of inputs, infrastructures, etc.). Conversely, the parameter b_l represents the social term and measures the strength of agglomeration economies in a given location: it is the amount by which the advantages obtained by locating in a certain site increases as a function of the number of agents already located there due for instance to technological factors and externalities of various types. A larger value of the parameter b implies that the incentive for an agent to locate in that site increases faster with the number of agents that have already settled there.

Before we illustrate how this model can be used to build empirically testable specifications, two remarks are in order. First, notice that the new "entrant" may well choose a location different from the one where "death" occurred. Thus the model is designed to capture both genuine entry of new agents and the reversibility of locational decisions of incumbents which might exit from one site just to select another one elsewhere. Second, in this model one may refer to events of birth and death as concerning both firms (more precisely business units) and employment opportunities (that is, the apparence and disappearance of employment units). In both cases the assumption that entry rates are positive, constant and equal to exit rates can be justified on an empirical ground. Indeed the share of firms (employees) belonging to a given sector which enter and leave a given location in a relatively short period of time (e.g. a year) is typically much larger than the net growth of industry size, so that the time-scale at which spatial reallocations occur is generally quite short.[3] Similar considerations apply to employment turnover whereby one observes quite high gross turnover even in presence of low net variations.[4]

Our model has many points in common with the Polya-Urn schemes popularized by Arthur (1994) and studied in Dosi et al. (1994) and Dosi and Kaniovski (1994). However, in the Polya-Urn framework the population grows through time, locational decisions are irreversible and the impact of any single locational decision becomes less and less important as time goes by. As a consequence such schemes describe a process that is non-ergodic and allows degenerate asymptotic states to emerge.

Conversely in our model, the dynamics implied by the rules we assumed for entry and exit is equivalent to a finite Markov chain whose state space is the set of all the possible distribution of the N agents across the L locations. In particular, it can be shown (cfr. Bottazzi and Secchi, 2007) that the assumptions of zero net-entry together with the reversibility of individual locational decisions and the constant impact of any single decision on the state of the system (implied by the equation 3.1) guarantee that the evolution of locational choices is an ergodic process that allows for non degenerate limit distributions. Moreover, Bottazzi et al. (2007) show (cfr. Proposition 2.2) that the process governing the evolution of the economy admits a unique long-run equilibrium (i.e. a unique invariant limit distribution) so that a probability π is assigned to each possible configuration $\boldsymbol{n} = \{n_1, \ldots, n_L\}$ where n_l is the number of agents in

---

[3] For a detailed comparative cross-country overview concerning firms turnover c.f. Bartelsman et al. (2005). On the Italian case, see e.g. Quarterly Reports by Unioncamere, "Movimprese: Dati Trimestrali sulla Nati-Mortalità delle Imprese", *Uffici Studi e Statistica Camere di Commercio*, Italy, various years, available on line at the url: http://www.starnet.unioncamere.it. Clearly the extent to which the assumption of zero net entry is realistic depends on the level of aggregation. At higher level of disaggregation one should in fact allow for (possibly endogenous) entry-exit processes with positive or negative net entry flows.

[4] On the employment turnover rates in Italy cfr. Contini (2002) and more generally Davies and Haltiwanger (1999) for international comparisons.

location l. This limit distribution $\pi(\boldsymbol{n}; \boldsymbol{a}, \boldsymbol{b})$ is analytically characterized as a function of the set of parameters of the model, the L-tuples of the geographic attractiveness $\boldsymbol{a} = (a_1, \ldots, a_L)$ and of the agglomeration strength $\boldsymbol{b} = (b_1, \ldots, b_L)$ of the L different locations. By varying the relative strength of geographical attractiveness and of agglomeration positive feedbacks this model is able to reproduce a rich variety of different patterns of spatial concentration. At one extreme, when agglomeration forces are very low, different locations attract on average a number of agents that is proportional to their geographical attractiveness, a_l. At the other hand, when agglomeration forces are very strong this model implies the emergence of highly polarized distribution, where few locations capture the great majority of agents.

To sum up, the dynamics governing the model does generate sharp empirically testable implications, in terms of the probability of finding the economy in a given state $\pi(\boldsymbol{n}; \boldsymbol{a}, \boldsymbol{b})$. Notice that this equilibrium (limit) distribution *does not* necessarily depict a long-run (limit) state associated to some ‘old’ or ‘mature’ industry. Since each entry/exit decision made by any one firm constitutes one time-step in the model, the invariant distribution describes the state of the system after a sufficient large number of spatial reallocation events have taken place (which may well imply a relatively short period of real time). Invariant distributions can then be directly compared with cross-section empirical data as far as they describe a system which is, on average, near its stochastic equilibrium state.

4 Testable Instances of the Model

The most general version of the model described in the previous section does contain a quite large number of free parameters. More precisely, one has to deal with two parameters for each location l: its geographic attractiveness a_l and the local strength of agglomeration b_l. In order to estimate such a model against empirical observations one would need longitudinal data on the number of firms in every single location. Unfortunately, we do not have such information. Indeed, in the following we apply the model to a dataset, described in Section 2, which contains only one observation per location per industrial sector. This forces us to explore less general models containing a lower numbers of parameters. Consequently, in estimating our model on empirical data, we will mainly employ the marginal distribution of the number of firms in a given location $\pi(n; a, b, \boldsymbol{a}, \boldsymbol{b})$, the latter being the probability to find n firms in a location characterized by coefficients (a, b). Bottazzi and Secchi (2007) shows that $\pi(n; a, b, \boldsymbol{a}, \boldsymbol{b})$ can be easily obtained from $\pi(\boldsymbol{n}; \boldsymbol{a}, \boldsymbol{b})$.

Let us then present different instances of our general model, starting with a simple (and, as we will see, utterly unrealistic) example, characterized by “homogeneous” space and constant returns to agglomeration, and progressively introducing more general models that differentiate locations and sectoral dynamics.

4.1 Model 0: Homogeneous Locations without Agglomeration Effects

Let us start with the simplest model where the agglomeration strength parameter is set to zero in any location, i.e. $b_l = 0$, $\forall l$, and all locations possess the same geographic attractiveness $a_l = a$, $\forall l$, where a is a positive constant. Consider this case as a sort of "null hypothesis" benchmark whereby neither spatial specificities nor agglomeration processes play any lasting role. In this extreme setup, firms choose locations totally at random. The limit distribution $\pi(n; a, b, \boldsymbol{a}, \boldsymbol{b})$ will then be multinomial, while the probability to find n firms in any given location is

$$\pi(n; N, L) = \binom{N}{n} \left(\frac{1}{L}\right)^{n} \left(1 - \frac{1}{L}\right)^{N-n}, \tag{4.1}$$

that is a binomial distribution with N trials and probability $1/L$. Therefore, in a homogeneous-space model without agglomeration economies, the stationary distribution does not depend on the common geographic attractiveness a. The underlying intuition is that the asymptotic occupancy of a location is driven by its relative attractiveness rather than its absolute one. In this case, whatever the value of the common parameter a, the locations are all and always equally attractive. Notice also that, given the full symmetry of the model, the marginal distribution is the same for all locations.

4.2 Model 1: Homogeneous Locations with Agglomeration Effects

Next, let us consider a model where locations are homogeneous and share the same geographic attractiveness $a_l = a > 0$, but one allows for agglomeration economies in the form of an industry-wide agglomeration parameter $b_l = b > 0$.

In analogy with the simpler Model 0 discussed above, also in this case all locations are identical with respect to the geographic attractiveness and the model is perfectly symmetric. The marginal distribution of the number of firms in a location $\pi(n)$ does not depend on the particular chosen location and can be shown to follow a Polya distribution (Bottazzi et al., 2007):

$$\pi(n; a, b, N, L) = \binom{N}{n} \frac{\Gamma(La/b)}{\Gamma(La/b + N)} \frac{\Gamma(a/b + n)}{\Gamma(a/b)} \frac{\Gamma((L-1)a/b + N - n)}{\Gamma((L-1)a/b)} \tag{4.2}$$

In this case the marginal distribution in (4.2) depends on the total number of firms N, the total number of locations L and the two parameters a and b.

As an illustration, we report in Fig. 1 the Polya distributions for different values of the parameter b. All distributions are computed according to (4.2), by setting $a = 1$ and with the same values for the parameters N and L (the latter values are chosen to be similar to the ones found in the subsequent empirical analyses). As shown, for small values of the parameter b the

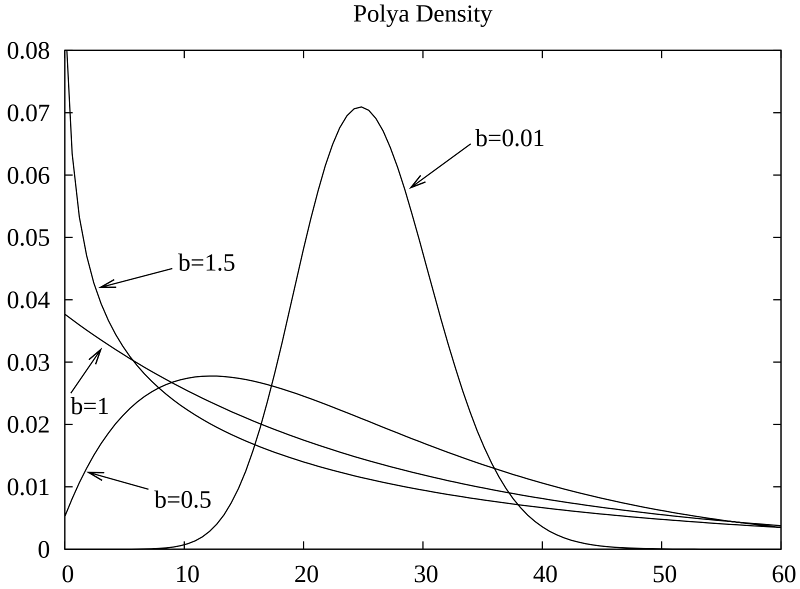

Figure 1: Polya marginal distributions for different values of b. All distributions are computed for $N = 20000$, $L = 784$, and geographic attractiveness $a = 1$. Note that values for N and L are set to be similar to values empirically found in our subsequent analyses.

Polya distribution is similar to the Binomial distribution, with a positive modal value and its well-known "bell" shape. When the parameter b increases, the mode of the distribution moves towards $n = 0$ and the upper tail becomes noticeably fatter. In tune with the intuition on the properties of agglomeration economies, an increase in the agglomeration strength parameter b yields a stronger "clusterization" of firms, i.e. a large number of firms in few locations (hence the fat tail), leaving, at the same time, more locations empty (hence the modal value of zero).

4.3 Model 2: Heterogeneous Locations with Agglomeration Effects

Let us now relax the assumption of homogeneity among locations and consider different geographic attractiveness a_l for each different location l. The strength of the agglomeration economy is still represented by an industry-specific parameter b, equal for all locations.

In this case locations do, in general, differ and are characterized by their specific attractiveness parameter a_l. As it happens to Model 1, also in this case the marginal distribution of the number of firms in a location with geographic attractiveness a can be shown to follow a Polya distribution, given by

$$\pi(n; a, A, b, N, L) = \binom{N}{n} \frac{\Gamma(A/b)}{\Gamma(A/b + N)} \frac{\Gamma(a/b + n)}{\Gamma(a/b)} \frac{\Gamma((A - a)/b + N - n)}{\Gamma((A - a)/b)} \qquad (4.3)$$

where $A = \sum_{h=1}^{L} a_h$ (cfr. Bottazzi et al., 2007). The marginal distribution in (4.3) depends, for a given location with attractiveness parameter $a_l = a$, on the total number of firms N, the total number of locations L, the global parameter b and the location-specific parameters a_l through their sum A.

5 Empirical Analysis

To recall, the model presented in Section 3 describes the localization pattern of a single sector economy wherein the number of firms is kept constant and the economy is governed by a steady entry/exit process capturing both the flow of firms to and from the industry, and a reallocation process by incumbents across locations. As mentioned, the empirical flows in and out industries are quite high. Hence it is not implausible to assume that the actual observations tell us something about the underlying invariant distribution $\pi(\boldsymbol{n}; \boldsymbol{a}, \boldsymbol{b})$. Of course, this does not rule out the possibility that in the long-term the nature and intensity of agglomeration drivers may well change. Such longer-term modifications may be captured by corresponding changes in the a and b coefficients (eventually detectable by comparing estimates across, say, different decades). However, since our database contains information on one single year, we can only compute the occupancy value for a given location and a given sector at a given point in time. This means that neither a direct verification of the dynamic process described in Section 3 nor a maximum-likelihood estimation of the equilibrium distribution in $\pi(n; \boldsymbol{a}, \boldsymbol{b})$ are possible. We have therefore to resort to some derived statistics in order to fit our models. In this way, we are able to exploit the rich cross-sectional information stemming from the presence of multiple sectors.

Let $n_{j,l}$ be the number of BUs in LSLM l operating in sector j, where $1 \leq j \leq 22$ and $1 \leq l \leq 784$ (cfr. Section 2). Denote with $N_{\cdot,l} = \sum_j n_{j,l}$ the total number of BUs operating in location (LSLM) l and with $N_{j,\cdot} = \sum_l n_{j,l}$ the total number of BUs belonging to the j-th sector.

For each sector j, we can build the occupancy frequency $f_j(n)$, counting the number of locations that contain exactly n firms operating in sector j. For instance, $f_3(0)$ is the number of locations that contain no firms of sector 3, $f_3(1)$ is the number of locations that contain exactly 1 firm operating in sector 3, and so on. The formal definition is

$$f_j(n) = \sum_{l=1}^{L} \delta_{n_{j,l},n} \tag{5.1}$$

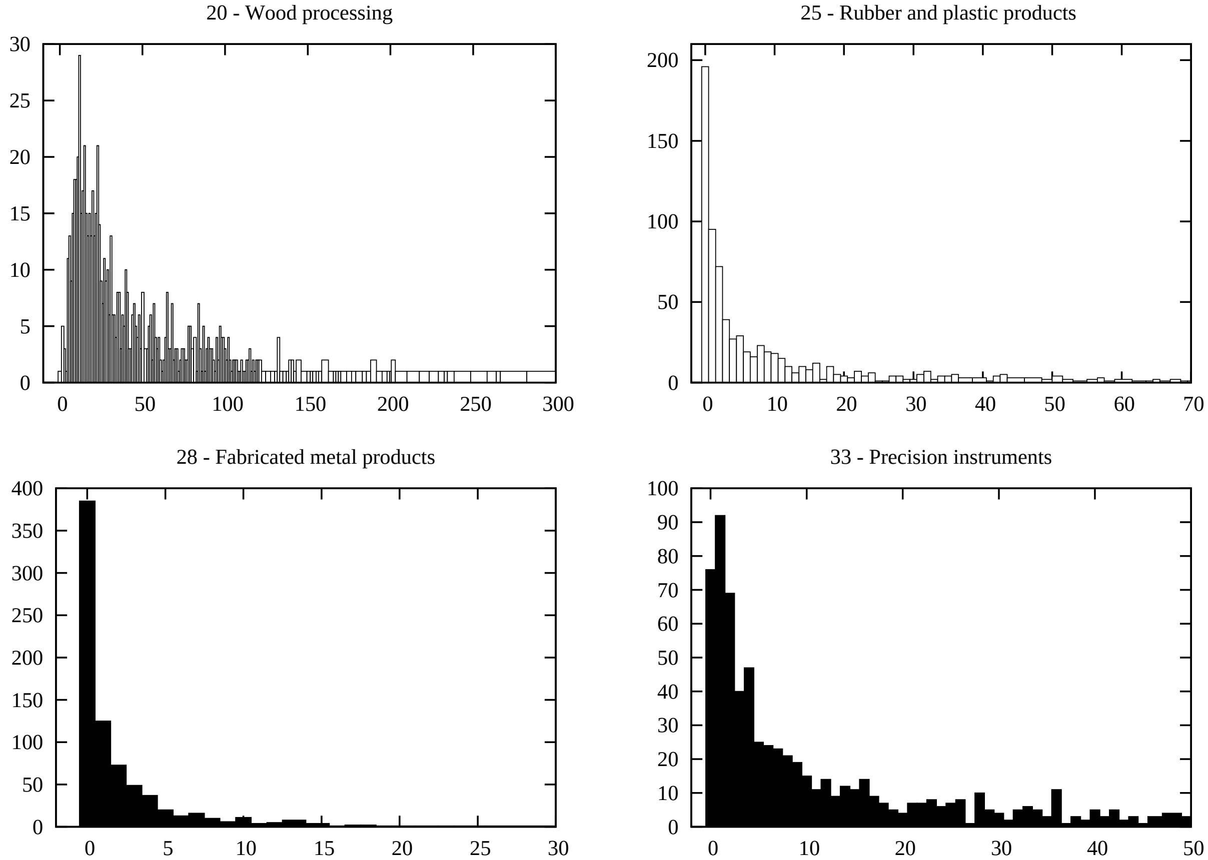

Figure 2: Occupancy frequency in four different sectors. The largest locations have been removed in order to better focus on the behavior of the distributions near the origin.

where $\delta_{n_{j,l},n}$ is the Kronecker delta. From (5.1) it is obvious that[5]

$$\sum_{n=0}^{+\infty} f_j(n) = L \ \ \forall j \in \{1, \ldots, 22\} \ \ .$$

In Figure 2 we show, as an example, the occupancy frequencies in four different sectors. Sectoral specificities are striking: both the shape of the distributions and the scales on the x and y axis are, indeed, very different. For instance, consider ATECO 15 sector (Food products): there are few locations which do not contain any firm belonging to this sector and the majority of locations contains $10 - 20$ firms operating in it. In the case of ATECO 21 sector (Pulp and paper) the picture changes. Here the number of empty locations is quite large, around 320, i.e. 40% of the total. For this sector, a location with 25 firms is a "crowded" one, and indeed $n = 25$ belongs to the upper tail of the frequency distribution. For sector

---

[5] Note that one sets infinity as the upper bound of the summation even if, clearly, such a summation stops with the number of firms in the most populated location. For instance, if sector 3 has a location with 5000 firms and no locations with a larger occupancy, we get $f_3(5000) = 1$ and $f_3(n) = 0$, $\forall n > 5000$ so that the summation effectively stops at 5000 .

Class	C_0	C_1	C_3	C_4	C_5	C_6
Range	0	$1-2$	$3-6$	$7-14$	$15-30$	$31-62$
Class	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
Range	$63-126$	$127-254$	$255-510$	$511-1022$	$1023-2046$	$2047-4094$

Table 2: Definition of the the first 12 occupancy classes.

20 (Wood processing), on the other hand, observing around 100 firms in a location is a quite common event.

In general, a frequency distribution with an high modal value around 0 and long tails represent a sector where the majority of firms is clustered in few places and the remaining locations are basically empty. On the contrary, a "bell-shaped" distribution is associated with a sector where the large part of the firms is evenly distributed in a relatively large number of locations.

In the rest of this section we will use the empirical occupancy frequency, defined in (5.1), to study the degree of agreement of the empirical data with the theoretical models presented in Section 4. Indeed, if $\pi(n)$ is the marginal distribution derived from a theoretical model and associated with a given sector, say j, the theoretical prediction for the occupancy frequency is $\pi(n)\, N_{j,\cdot}$.

Since the support of the empirical occupancy frequency is in general large, due to the presence of few extremely populated locations and many (almost) non-populated ones, instead of using each occupancy number we consider occupancy classes (analogous to the often-used size classes) defined, for each sector, as the number of locations having a number of firms belonging to that same sector inside a given range. We define classes with ranges following a geometric progression

$$C_k = \left[2^k - 1, 2^{k+1} - 1\right) \quad k = 0, 1, 2, \dots \quad , \tag{5.2}$$

and we report in Table 2 the first 12 occupancy classes as an example. The frequency of the different occupancy classes $f_j(C_h)$ for $h \in \{1, \dots, 12\}$ can then be easily computed from (5.1). We have

$$f_j(C_h) = \sum_{n \in C_h} f_j(n), \tag{5.3}$$

where the sum spans over the integers belonging to each class range.

Model 0

Let us start with the simplest benchmark provided by Model 0, described in Section 4.1, where all locations are assumed to be homogeneous and the agglomeration strength is set to zero ($b = 0$). In this case no estimation procedure is necessary. Indeed, the marginal distribution only depends on the number of locations L and the number of firms N operating in the sector (see Section 4.1).

For each sector j we can obtain a theoretical prediction for the class frequency directly from (4.1). One has

$$f_j^{\mathrm{th}}(C_h) = \sum_{n \in C_h} L\ \pi(n; N_{j,\cdot}, L) \tag{5.4}$$

where $L = 784$ (i.e., the number of LSLM contained in our database).

Figure 3 plots the empirical class frequency (5.3) together with the theoretical prediction (5.4) for two sectors quite representative of all of them. The agreement is basically nil for all sectors. The theoretical frequency is proportional to the binomial distribution, and thus displays a bell-like shape with almost all the weight being distributed in few central classes. This pattern, however, is never observed in empirical data. Note that this negative result is indeed an important one in that it falsifies any notion of random attribution of business units over a homogeneous space with null returns to agglomeration (see also Rysman and Greenstein, 2005).

Model 1

Next, let us start to investigate the relevance of agglomeration economies by considering Model 1, described in Section 4.2, in which we allow for a non-zero agglomeration strength parameter $b > 0$. In this case, the marginal distribution of the model, defined in (4.2), depends on the parameters ratio a/b. This means that the model is insensitive to re-scaling, by a common factor, of both the locational geographic attractiveness a and the agglomeration strength b. Without loosing in generality, in the following analysis we set $a = 1$, and, for each sector, we estimate the best fit by varying the parameter b. For this purpose, we use the Chi-Squared statistics with the occupancy classes C_h as categories. For each sector, starting from the marginal distribution in (4.2), we can build the observed classes frequency $f_j(C_h)$ and also the theoretical classes frequency as

$$f_j^{\mathrm{th}}(C_h) = \sum_{n \in C_h} N_{j,\cdot}\ \pi(n; N_{j,\cdot}, L, 1, A, b) \quad . \tag{5.5}$$

We then consider the Chi-Squared statistics

$$\chi_j^2(b) = \sum_h \frac{\left(f_j(C_h) - f_j^{\mathrm{th}}(C_h)\right)^2}{f_j^{\mathrm{th}}(C_h)} \tag{5.6}$$

defined, for each sector j, as a function of the parameter b. Finally, we estimate the sectoral-specific optimal value b_j^* according to

$$b_j^* = \underset{b \in R^+}{\arg\min}\ \chi_j^2(b) \quad . \tag{5.7}$$

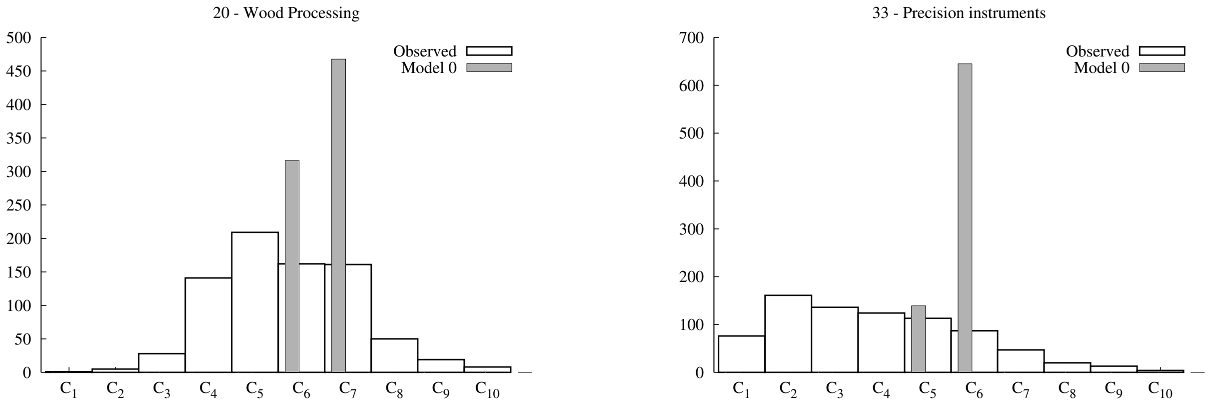

Figure 3: Occupancy class frequencies computed on observed data (white bars) and estimated using Model 0 (gray bars).

The resulting b_j^* for different sectors are reported in Table 3 together with the average absolute deviation (AAD) that represents a measure of the agreement between the empirical and the theoretical frequencies and is defined as

$$AAD_j = \frac{1}{K_j} \sum_h \left| f_j(C_h) - f_j^{\text{th}}(C_h) \right| \quad , \tag{5.8}$$

where K_j denotes the number of classes in sector j. From Table 3, is apparent the high degree of sectoral heterogeneity in both the strength of the agglomeration forces and in the ability of Model 1 to reproduce empirical distributions in different sectors. This is weel illustrated by Figure 4 showing, for six different sectors, the theoretical class frequencies obtained using (4.2) with the estimated value b_j^* (gray bars). Visual inspection of these plots reveals that the degree of accordance with the data dramatically improves as compared with Model 0. In particular, the agreement with empirical frequencies (white large bars) is, in general, good in the central part of the distribution while the fit remarkably worsens at the two extremes: in some sectors (for instance ATECO sectors 20, 26 and 28) Model 1 largely overestimates the number of locations with few firms. In other sectors (for instance in sectors 20, 25 and 32), the model does a good job in describing the nearly empty locations but fails to capture the upper tail of the distribution, underestimating the occurrences of very “busy” sites.

Model 2

Ultimately Model 1, while significantly improving the ability to reproduce the observed patterns, seems unable to describe the tails of the empirical distributions, in particular when the latter displays both a large number of locations containing a relatively small number of firms and a few locations with an high number of firms. These tail effects cannot be replicated by varying the parameter b alone. Indeed if the value of b is large, Model 1 predicts the existence of several locations with a huge number of firms but, at the same time, predicts that all other

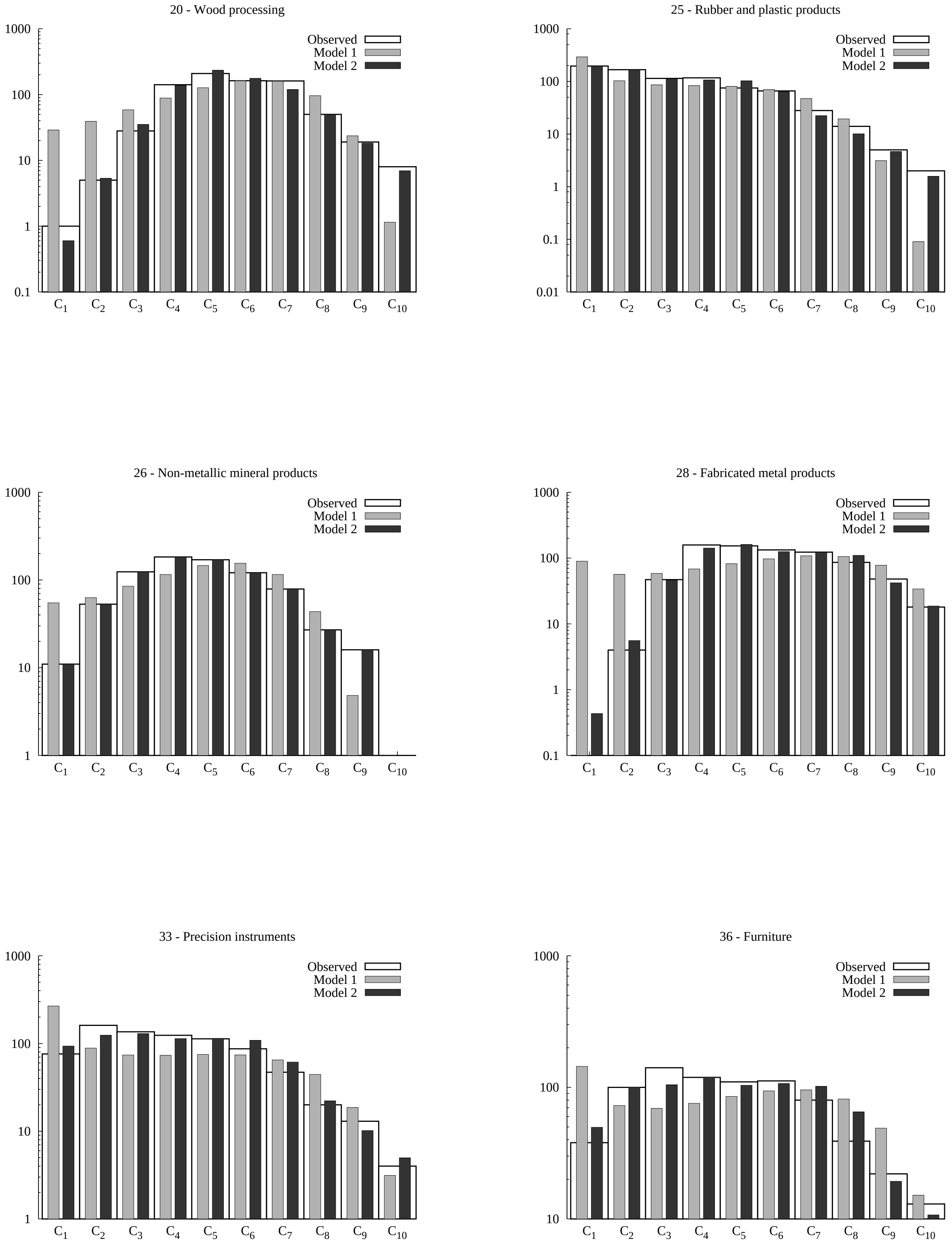

Figure 4: Occupancy class frequencies computed on observed data (white bars) and estimated using Model 1 (gray bars) and Model 2 (black bars).

locations are essentially empty. Conversely, a small value of b accounts for a large number of locations with few firms, but reduces the probability of finding large clusters close to zero. This difficulty can be partly tackled with Model 2, wherein different locations are allowed to have different geographic attractiveness, so that the observed clusterization of a large number of firms in a single location can be explained by the presence of a relative high geographic attractiveness, even if the sector is characterized by a mild value of the agglomeration parameter b.

The drawback of Model 2 as presented in Section 4.3, however, rests in its large number of parameters: one should specify the value of the sector-specific parameter b and the value of the parameter a for each location. Hence, one cannot hope to obtain the values of all these parameters from a Chi-Squared minimization procedure, undertaken on each sector separately, as in (5.6). Indeed, in our case the number of parameters is equal to the number of observations plus one (i.e. 785). In order to overcome this problem, we exploit the double disaggregation (by sector and by location) of our database.

First of all, let us make the following

Assumption 1 (Urbanization effect)**.** *The geographic attractiveness* $a_{j,l}$ *of location* l *for firms operating in sector* j *is proportional to the number of firms located in* l *belonging to all the sectors except* j

$$a_{j,l} \sim \alpha_j + \beta_j \frac{N_{-j,l}}{N_{-j,\cdot}} \tag{5.9}$$

where, with the usual notation

$$\begin{aligned} N_{-j,l} &= \sum_{i \neq j} n_{i,l} \\ N_{-j,\cdot} &= \sum_{l} N_{-j,l} \quad . \end{aligned}$$

As noted in Section 3, the geographic attractiveness coefficient a controls for all geographical factors that are not related with the sector under study. We can think to all such factors as both exogenous "geographical" and infrastructural ones, but also general demand-induced externalities, market proximity effects, etc.

The linear relation in (5.9) depends on two sectoral parameters α_j and β_j. The parameter β_j represents a measure of the overall "pull" exerted by all business units from all other sectors on the locational decisions of firms belonging to sector j. Parameter β_j captures what we call "urbanization effect": the overall installed base of production units in a particular location brings about a stronger attractive strength in sectors with a higher value of β.

The stationary distribution of Model 2 depends only on the ratios a_l/b so that, again, we can rescale all the parameters $\boldsymbol{a}$ and b by the same factor without affecting the distribution. In order to obtain values for b comparable with the ones found when estimating Model 1, where

	# of firms		Model 1		Model 2 - All sites			Model 2 - No metropolis		
Sector	All sites	No Urban	b^*	AAD	β	$b^{*,(a)}$	AAD	β	$b^{*,(a)}$	AAD
15 Food products	75420	62751	1.17	0.0364	0.00	1.17	0.0364	0.00	0.95	0.0303
17 Textiles	36217	32043	6.05	0.0530	834.83	0.00	0.0108	0.00	6.76	0.0579
18 Apparel	49782	38137	3.42	0.0388	820.35	0.00	0.0084	0.00	2.48	0.0308
19 Leather products	25451	19791	6.57	0.0469	0.00	6.57	0.0469	0.00	5.57	0.0465
20 Wood processing	50662	42322	1.36	0.0366	652.54	0.06	0.0121	0.00	0.95	0.0342
21 Pulp and paper	5268	3794	5.63	0.0301	795.67	0.48	0.0144	0.01	3.50	0.0155
22 Publishing and printing	28183	16402	9.02	0.0785	954.12	0.51	0.0655	813.43	0.00	0.0154
23 Coke, refined petroleum and nuclear fuel	825	617	3.67	0.0111	786.78	0.46	0.0039	233.82	2.47	0.0067
24 Organic and Inorganic chemicals	7593	4941	7.43	0.0525	812.40	0.29	0.0160	871.17	0.00	0.0104
25 Rubber and plastic products	14626	11324	4.49	0.0330	847.53	0.00	0.0071	854.77	0.00	0.0091
26 Non-metallic mineral products	30709	25140	1.55	0.0401	715.51	0.07	0.0058	697.11	0.08	0.0064
27 Basic metals	4034	3010	6.16	0.0297	1.42	6.16	0.0297	0.00	4.65	0.0199
28 Fabricated metal products	94771	74340	2.67	0.0465	784.81	0.00	0.0076	774.05	0.00	0.0096
29 Industrial machinery and equipment	42984	33157	3.47	0.0331	830.67	0.00	0.0065	832.34	0.00	0.0080
30 Office machinery	592	331	12.02	0.0091	0.00	12.02	0.0091	856.86	3.77	0.0039
31 Electrical machinery	17312	11906	6.44	0.0478	844.53	0.00	0.0093	849.32	0.00	0.0122
32 Radio, TV and TLC devices	9773	6546	3.53	0.0415	825.08	0.00	0.0131	0.00	2.37	0.0284
33 Precision instruments	28280	18713	4.80	0.0556	827.10	0.00	0.0134	818.89	0.00	0.0152
34 Motor vehicles and trailers	2261	1619	8.13	0.0297	961.57	0.35	0.0036	0.05	4.37	0.0076
35 Other transport equipment	4514	3500	5.43	0.0138	0.75	5.43	0.0138	0.34	5.12	0.0097
36 Furniture	59627	46460	3.26	0.0449	822.64	0.00	0.0139	0.00	3.28	0.0416
37 Recycling	2061	1568	3.19	0.0124	729.96	0.63	0.0043	1.04	2.67	0.0032

Note: (a) Values smaller than 10^{-4} are reported as 0.0.

Table 3: Summary statistics of estimates from models 1 and 2, by sector (estimates are based on the *number of firms*)

	# of employees		**Model 2 - All sites**			**Model 2 - No metropolis**		
Sector	All sites	No Urban	β	$b^{*,(a)}$	AAD	β	$b^{*,(a)}$	AAD
15 Food products	434515	357838	728.85	0.03	0.0104	0.00	1.69	0.0235
17 Textiles	345338	317929	0.00	6.45	0.0275	0.00	6.86	0.0299
18 Apparel	346387	292519	791.87	0.00	0.0102	0.00	3.19	0.0226
19 Leather products	230543	190829	398.67	4.81	0.0240	0.00	8.05	0.0250
20 Wood processing	170294	146997	0.00	1.57	0.0300	693.84	0.05	0.0122
21 Pulp and paper	85424	68215	0.00	7.72	0.0079	838.57	0.50	0.0099
22 Publishing and printing	175012	90325	680.44	2.87	0.0529	0.00	3.64	0.0288
23 Coke, refined petroleum and nuclear fuel	24147	15058	0.00	16.21	0.0117	1928.72	0.80	0.1080
24 Organic and Inorganic chemicals	209242	120570	843.29	0.31	0.0192	0.00	6.69	0.0114
25 Rubber and plastic products	198401	155614	0.00	6.07	0.0121	4.11	5.05	0.0067
26 Non-metallic mineral products	250824	216898	756.22	0.06	0.0072	1.28	3.76	0.0335
27 Basic metals	136123	108682	872.11	0.96	0.0062	1004.81	2.80	0.0146
28 Fabricated metal products	621642	502906	594.54	0.80	0.0273	0.00	2.90	0.0302
29 Industrial machinery and equipment	554105	430467	290.72	3.71	0.0223	0.00	4.58	0.0231
30 Office machinery	18609	9359	1083.65	15.05	0.0123	997.39	19.65	0.0098
31 Electrical machinery	205797	136008	821.10	0.15	0.0203	0.00	5.08	0.0122
32 Radio, TV and TLC devices	103161	53877	556.44	2.77	0.0373	0.00	6.79	0.0352
33 Precision instruments	129448	79972	660.57	0.81	0.0345	0.00	4.23	0.0365
34 Motor vehicles and trailers	185748	100842	0.00	15.09	0.0080	0.00	13.47	0.0073
35 Other transport equipment	100780	63304	0.00	11.59	0.0084	1218.21	3.79	0.0117
36 Furniture	309911	260270	0.00	4.66	0.0344	770.49	0.80	0.0487
37 Recycling	8327	6364	2.30	5.71	0.0114	144.00	5.14	0.0108

Note: (a) Values smaller than 10^{-4} are reported as 0.0.

Table 4: Summary statistics of estimates from models 1 and 2, by sector (estimates are based on the *number of employees*).

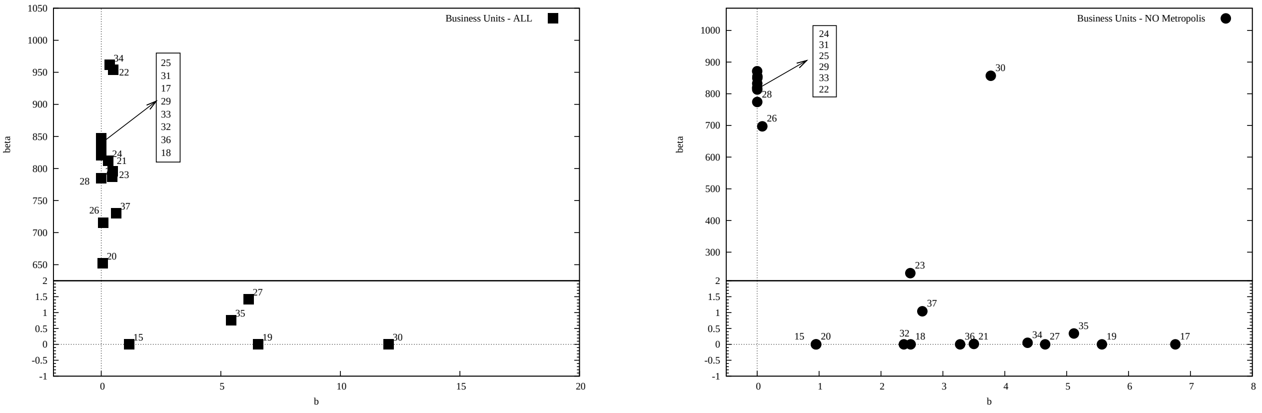

Figure 5: Scatter plot of the b and β parameters estimated from model 2 for different sectors with(left panel) and without(right panel) the metropolitan areas (estimates are based on the *number of firms*).

we assumed $a = 1$, we impose[6] the further requirement that the average value of a is 1, i.e.

$$\frac{1}{L}\sum_{l} a_{j,l} = 1$$

so that (5.9) reduces to a one parameter relation

$$a_{j,l} = 1 + \beta_j \left(\frac{N_{-j,l}}{N_{-j}} - \frac{1}{L} \right) \quad . \tag{5.10}$$

Substituting (5.10) in the marginal distribution (4.3) one can compute the theoretical prediction for the occupancy class frequency

$$f_j^{\mathrm{th}}(C_h) = \sum_{n \in C_h} \sum_{l} \pi(n; N_{j,\cdot}, L, \beta_j, A, b) \quad . \tag{5.11}$$

Notice that in (5.11) a summation over l is required since different locations now possess different geographic attractiveness and, consequently, are characterized by different marginal distributions.

Finally, following the same approach described in the previous section, one can obtain an estimate for (b, β) as

$$(b_j^*, \beta_j^*) = \underset{b,\beta \in R^+}{\arg\min}\, \chi_j^2(b, \beta) \quad , \tag{5.12}$$

where χ^2 is defined as in equation (5.7).

Let us start by noting that moving from Model 1 to Model 2, one observes an unambiguous improvement of the ability of model to reproduce the empirical observations: this is clear from visual inspection of Figure 4, where one observes a very good accordance of predicted

---

[6] This assumption is made only for comparability purposes and does not significantly affect our results.

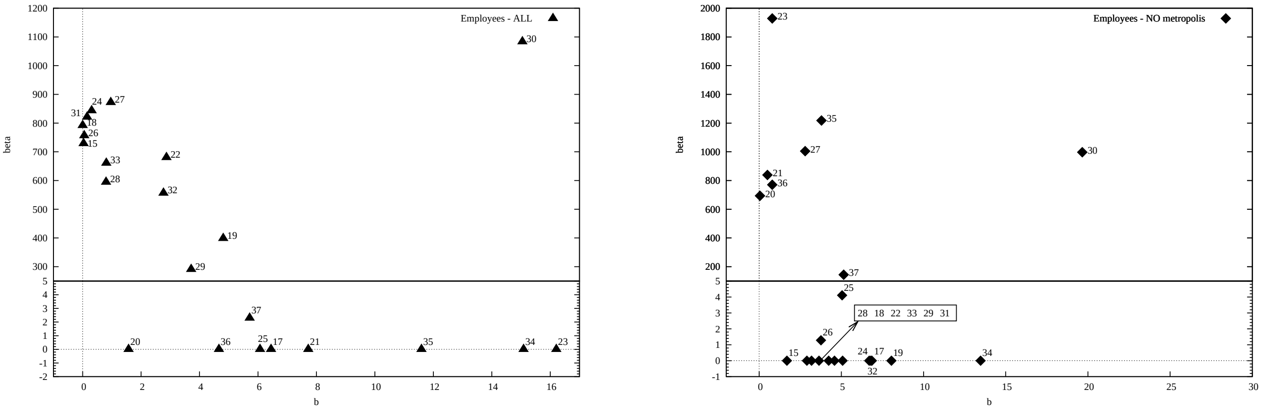

Figure 6: Scatter plot of the b and β parameters estimated from model 2 for different sectors with(left panel) and without(right panel) the metropolitan areas (estimates are based on the *number of employees*).

frequencies (black bars) with the observed ones and is confirmed by the reduction in the average absolute deviation (AAD), Table 3. Indeed, Model 2 seems able to overcome, at least in a first approximation, the inability of the previous one to capture the tails behavior of the empirical distributions.

The estimates of the values of (b_j^*, β_j^*) together with the AAD (defined in equation 5.8) are reported in Table 3 and illustrated in Figure 5 (left panel). A rather striking feature of the plot is the apparent polarization between a group of sectors which shows a nearly exclusive impact of "urbanization effect" and another one wherein sector-specific agglomeration effects dominate. In that, the attribution of individual sectors to the two groups turns out to be somewhat puzzling (for example "17 - Textiles" and "19 - Apparel" appear to belong, counterintuitively, to the former group). Such a puzzling evidence, in fact, may be largely the outcome of a sort of "horizontal pull" of metropolitan areas which tend to exert what we could call a *more-of-everything* effect (including more of the activities which are traditionally associated with sector-specific agglomeration phenomena, such as the mentioned textiles and apparel). In fact by removing the metropolitan areas[7] the picture significantly changes: cfr. Figure 5 (right panel) and Table 3. When they are present, agglomeration effects tend to be mostly of a sector specific nature. Note that, even in those sector where β is positive, urbanization tends to explain a relatively small part of the inter-site variation in locational intensities.[8]

In the previous analyses agglomeration has been measured by considering only the number of firms present in each location, and not their (relative) size. Further precious information, stemming from firm size distribution in different locations, may be obtained by estimating our

---

[7] The Italian Statistical Office identifies 11 (out of 784 LSLM) Metropolitan areas around the cities of Bari, Bologna, Cagliari, Firenze, Genova, Milano, Napoli, Palermo, Roma, Torino, Venezia.

[8] Rough but illustrative evidence comes from the low goodness of fit of the estimate of the relation

$$n_{j.l}/N_j = \gamma_0 + \gamma_1 \; N_{-j,l}/N_{-j} \quad . \tag{5.13}$$

model on employment data. That is, instead of using data on the number of firms belonging to any given sector that are present in each location, we can apply the model to to the number of firm employees, per location and per sector. In this case "agglomeration" also captures the effect of increasing returns and internalization of productive activities within single firms. So, for example, in employment-based estimates the "strength of agglomeration" of a location with say one firm with a thousand employees is taken to be equivalent to another one with 100 firms of 10 employees each (which of course would not be the case in the previous estimation procedures).

The estimates of the (b_j^*, β_j^*) are presented in Table 4 and illustrated in Figure 6 (left panel). Again the analysis of the universe of locations tend to be affected by the rather special agglomerative pull of metropolitan areas (cfr. Figure 6, left panel). If one excludes them, the picture, Figure 6 (right panel), is relatively similar to the one stemming from firms locational patterns. Sectoral agglomeration effects seem to dominate.[9] And, of course, given the somewhat expansive notion of agglomeration, the estimates now capture also the effects of the location patterns of few but large firms (cfr. for example, the sector "34 - Motor vehicles")

6 Conclusions

The purpose of this work has been to offer relatively general and empirically applicable formal tools able to assess the importance of agglomeration phenomena, in general, and to distinguish between their location-wide and sector-specific drivers. Despite its simple structure, the model is indeed able to generate testable implications on the whole shape of the distribution of firms locational or employment choices in any given sector (indeed an improvement over the majority of existing models which only provide insights on agglomeration indices: cfr. for instance Ellison and Glaeser (1997)). The outcomes are quite encouraging.

First, the evidence from the locational patterns of Italian manufacturing industry adds very robust statistical support to the old claim that the spatial dimension provides structure to the distribution of production activities. Our results, indeed, strongly reject any hypothesis that observed locational patterns are explained by purely random factors for every 2-digits manufacturing sectors.

Second, our model allows to disentangle the relative importance of the "pull" of particular locations themselves from the agglomeration forces associated with each particular sectors. The former include inter-sectoral linkages via technological and demand flows and other location-wide externalities. Together, they make what we have called the attractiveness of a location. When one allows for heterogeneity of such attractiveness across locations, one does indeed find that such forces appear to matter in particular for metropolitan areas. In other

---

[9] Also in this case the goodness of fit of the estimate of the relation 5.13 is relatively low.

terms large metropolitan agglomeration forces exert a powerful pull upon locational patterns irrespectively of the characteristics of many sectors. This pull is horizontal in the sense that it tends to join together all activities. However, when one excludes very few big urban centers the impact of "horizontal" agglomeration forces appears to be significantly more seeable (another way of describe the same phenomenon could be by saying that the effect of "urbanization" appears to be highly non linear in the size of urban sites themselves). Correspondingly, one is able to detect also the important role played by the very history of locational decision *within each sector*. This has to do with some form of *dynamic increasing returns* such that the number of production units belonging to one particular sector of production at a particular time influences the probability that an additional unit will be located there, too. In this respect we do find important sector-specific forces of agglomeration which, interestingly, vary a good deal across sectors.

As such our findings are somewhat at odds with the almost exclusive emphasis of new economic geography on location-wide externalities, and plausibly hint at sector specific and localized forms of knowledge accumulation, spin-offs and formation of new firms. Such conclusions are indeed strengthened by the application of the model to the dynamics of employment: again, sector specific agglomeration forces appears to be powerfully at work, sometimes closely resembling the agglomeration profiles of (plausibly district-type) firms and some other time internalized within the employment strategies of relatively fewer but bigger firms.

The foregoing model can be extended in different ways. First, one might explicitly take into account interdependencies between locations and industries. In its present version, our model does not include the possibility that firms locational choices may be influenced by the choices made by firms belonging to different sectors, possibly located in neighboring regions. One might think to an extended version of the model where locations are positioned over some metric space, e.g. a two-dimensional lattice, and firms decisions (entry and exit) are somewhat correlated in space. Similarly, one might introduce urbanization economies whose advantages spill over to neighboring regions (unlike being concentrated in a given region).

Second, as discussed in Section 3, our assumption of a zero net entry rate may be justified, at least at the aggregation level at which we pursue our analysis, by empirical evidence. However, if one wants to extend the model to consider also lower levels of aggregation, an endogenous process of entry and exit might be possibly required (e.g., by assuming that the probability of exit is related to the number of firms in a region).

Third, as briefly discussed also in Bottazzi et al. (2006), one ought to explore the importance of the specificities of technological knowledge underlying the activities of each industry in explaining the observed intensities of agglomeration forces.

Finally, an interesting challenge involves the incorporation into the model of a non-linear account of location probabilities.

References

Arthur, W. (1994). *Increasing returns and path-dependency in economics.* University of Michigan Press, Ann Arbor.

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

Bottazzi, G., G. Dosi, and G. Fagiolo (2006). On sectoral specificities in the geography of corporate location. In S. Breschi and F. Malerba (Eds.), *Cluster, networks and innovation.* Oxford University Press, Oxford.

Bottazzi, G., G. Dosi, G. Fagiolo, and A. Secchi (2007). Modeling industrial evolution in geographical space. LEM Working Paper 2007-07, Scuola Speriore Sant'Anna.

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

Brenner, T. (2003). An identification of local industrial clusters in germany. Papers on Economic and Evolution No. 0304, Max Planck Institute, Jena, Germany.

Combes, P. and H. Overman (2004). The spatial distribution of economic activities in the european union. In V. Henderson and J. Thisse (Eds.), *Handbook of Urban and Regional Economics*, Volume 4. Amsterdam, North Holland. Forthcoming.

Contini, B. (2002). *Osservatorio sulla mobilità del lavoro in Italia.* Il Mulino, Bologna.

Davies, S. and J. Haltiwanger (1999). Gross job flows. In O. Ashenfelter and D. e. Card (Eds.), *Handbook of Labor Economics.* Elsevier, Amsterdam.

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

Dosi, G., Y. Ermoliev, and M. Kaniovski, Yu. (1994). Generalized urn schemes and technological dynamics. *Journal of Mathematical Economics 23*, 1–19.

Dosi, G. and Y. Kaniovski (1994). On 'badly behaved' dynamics. *Journal of Evolutionary Economics 4*, 93–123.

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

Ellison, G. and E. Glaeser (1997). Geographical concentration in u.s. manufacturing industries: A dartboard approach. *Journal of Political Economy 105*, 889–927.

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

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

Garibaldi, U. and M. Penco (2000). Ehrenfest urn model generalized: an exact approach for market participation models. *Statistica Applicata 12*, 249–272.

Garibaldi, U., M. Penco, and P. Viarengo (2002). An exact physical approach for market participation models. In R. Cowan and N. Jonard (Eds.), *Heterogeneous Agents, Interactions, and Economic Performance.* Lecture Notes in Economics and Mathematical Systems, Berlin-Heidelberg, Springer-Verlag.

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

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

Ottaviano, G. and G. Thisse (2004). Agglomoration and economic geography. In V. Henderson and J. Thisse (Eds.), *Handbook of Urban and Regional Economics*, Volume 4. Amsterdam, North Holland. Forthcoming.

Overman, H. and G. Duranton (2002). Localisation in u.k. manufacturing industries: Assessing non-randomness using micro-geographic data. CEPR Discussion Paper 3379, Centre for Economic Policy Research.

Rysman, M. and S. Greenstein (2005). Testing for agglomeration and dispersion. *Economics Letters 86*, 405–411.

Sforzi, F. (2000). Local development in the experience of italian industrial districts. In CNR (Ed.), *Geographies of Diverse Ties. An Italian Perspective.* Rome, CNR-IGU.