Regional Science and Urban Economics 29 (1999) 575–604

A measure of the geographic concentration in french manufacturing industries

Françoise Maurel*, Béatrice Sédillot

INSEE, timbre G201, 15 Boulevard Gabriel Péri, BP 100, 92244 Malakoff Cedex, France

Received 1 June 1996; received in revised form 1 August 1998; accepted 17 May 1999

Abstract

The purpose of this paper is to offer an empirical investigation of the geographic concentration of French industries. The index of concentration is derived from a location model in the line of Ellison and Glaeser (1994, 1997) and can be interpreted as the correlation between the location decisions of two business units in the same industry. Along with extractive and traditional industries, some high technology industries are highly localized, which supports the view that technological spillovers may be important. Besides, the identification of the most and least localized industries reveals similar patterns in France and in the U.S. © 1999 Elsevier Science B.V. All rights reserved.

JEL classification: R12; R32

Keywords: ; Location model; Geographic concentration; Agglomeration spillovers

1. Introduction

Models of economic geography can lead to different predictions regarding the location of economic activity. In many theories however, plants should locate near to each other because of agglomeration spillovers or local amenities. Besides, empirical evidence brings out that jobs and industries are highly clustered in a limited number of regions. The U.S. manufacturing belt offers a famous example

*Corresponding author. Tel.: +33-141-17-6067; fax: +33-1-41-17-6045.

E-mail address: francoise.maurel@insee.fr (F. Maurel)

0166-0462/99/$ – see front matter
PII: S0166-0462(99)00020-4

of industry concentration. To a lesser extent, the case of France is illustrative. More than one-third of manufacturing workers are employed in only two regions (Paris and Lyon areas), even though specific industries appear to have less concentrated location patterns than in the United States. There is no automotive industry cluster comparable to Detroit for example. At the national level, Krugman (1991) argued that the four major European countries had less specialized industry structure than the U.S. regions.

Several empirical investigations on high density areas in the United States and Japan have also concluded that agglomeration had positive effects on productivity (Ciccone and Hall, 1996; Fujita and Tabuchi, 1997), or that dynamic spillovers contributed positively to employment growth (Henderson et al. (1995)). But a precise diagnosis on the degree of agglomeration of industrial activity remains to be done. Our concern in this paper is therefore to offer an empirical investigation of the geographic industrial concentration for the French case, which can be compared with recent similar work for the United States by Ellison and Glaeser (1997). In a first section, we briefly review the main theories of localization and their explanation of agglomeration forces. In a second section, we propose an index of geographic concentration to explore the agglomeration of French industries in 1993. This index relies on a location model in the line of Ellison and Glaeser. Although it slightly differs from the one suggested by these authors, it has the same attractive features. First, it controls for differences in the size distribution of plants and thus provides a measure of the localization beyond the sole concentration of the employment. Hence, one industry will not be regarded as localized only because its employment is concentrated in a small number of plants. Second, this indicator allows for comparisons between industries. In a third section, the index of concentration is computed for French 4-digit industries. In a fourth section, we use different industry definitions to capture inter-industry spillovers and explore the geographic scope of localization whereas the last section compares, as far as possible, the results obtained for France with those of Ellison and Glaeser for the United States.

2. The main theories of localization

By allowing a decrease in the mean cost of production, returns to scale induce industries to concentrate their production in a small number of business units. In basic industries where fixed costs are high, the location is strongly influenced by the access to raw materials (ore, coal, quarry products). In some industries, the location may result from historical accident that led some industry to develop in one single region (Krugman, 1991).

This tendency towards polarization is reinforced by external economies that create interdependence between firms' location choice. Relying on the classification proposed by Hoover (1936), one can distinguish two types of spillovers:

localization economies (Marshall, 1890, Arrow, 1962, Romer, 1984), that benefit firms in the same industry, and *urbanization economies* that are common to all firms. Both localization and urbanization economies are dynamic and strengthen the attracting power of specific areas by a snowball effect mechanism (Fujita, 1989, Fujita and Thisse, 1996).

Intra- or inter-industry spillovers lead to different predictions regarding the organization of space (Glaeser et al., 1992). When localization economies dominate, space tends to be structured in specialized industrial poles. Conversely, when spillovers are common to all industries, polarization goes along with highly industrially diversified areas (Jacobs, 1969). The empirical work by Glaeser et al. (1992) on the growth of industrial employment in U.S. cities supports the view that spillovers across industries are more important than knowledge spillovers within one industry: employment growth is higher in highly diversified cities. However, others papers find opposite or less conclusive results. In Henderson et al. (1995), specialization speeds up employment growth whereas city diversity and specialization both contribute to the growth of French cities in Maurel (1996). From another perspective, empirical results of Ciccone and Hall (1996) show that productivity is positively related to spatial density and that more than half the variance of labor productivity across U.S. states can be explained by differences in the density in economic activity.

3. The index of geographic concentration

Although the debate on the nature of external economies is still pending from an empirical point of view, this rapid look on models of location choice stresses the importance of interdependence of firms' location choices, in particular through spillover mechanisms that contribute to centripetal forces. But a precise diagnosis of the importance of agglomeration forces in specific industries or countries remains to be done. Although there has been a growing literature on agglomeration over the recent years, it is not straightforward to derive measures of agglomeration economies or geographic externalities that allow for consistent comparisons between industries. With the Gini index[1], proposed by Krugman (1991), inter-industry comparisons appear to be very sensitive to the characteristics of the industry and results are highly dependent on the concentration of production within the industry. Using such an index, an industry will be regarded as localized as soon as its employment is concentrated in a small number of plants located in a limited number of geographic areas and even though the industry plants' location decisions are independent.

The index considered here aims to tackle this problem and improve the

[1] This index is based on the comparison between the geographic patterns of employment for one industry and in the aggregate.

measurement of the degree to which industries are geographically concentrated. In the line of Ellison and Glaeser (1994, 1997), it relies on a narrow definition of localization that focuses on the plant's location decision. What is important here is to determine whether plants' locations are influenced by such factors as the access to raw materials or industry technological spillovers. Returns to scale can also influence the geographic concentration within industries, but since we consider the number and the size of business units as given, we focus here on the correlation between firms' location decision.

3.1. The model of localization by Ellison and Glaeser

The index is based on a location model suggested by Ellison and Glaeser (1997) although it is slightly different from the one proposed by these authors. Ellison and Glaeser propose the following model of the location decision of business units in an industry. When location decisions are not independent, plants can choose their location to benefit from the *natural advantage* of one particular geographic area (access to raw materials, good climatic conditions) or *spillovers* generated by the proximity of other plants in the industry. The two models of natural advantage and spillovers are observationally equivalent. Hence, we only detail the spillover model here.

Let N denote the number of industry plants and $z_1 \dots z_N$ the share of each plant in industry employment. Let M be the number of geographic areas (e.g. departments in the French case) and $x_1 \dots x_M$ the fraction of each area in aggregate employment. The fraction of industry employment located in geographic area i is therefore:

$$s_i = \sum_{j=1}^{N} z_j \, u_{ji} \tag{1}$$

where $u_{ji} = 1$ if the business unit j locates in area i, 0 otherwise. u_{ji} are non-independent Bernouilli variables such that $P(u_{ji} = 1) = x_i$, which means that the random location process will on average lead to a pattern of employment shares matching the one that prevails in the aggregate. More precisely, Ellison and Glaeser propose to model the interaction between the location decisions of any pair of plants by:

$$\mathrm{Corr}(u_{ji}, u_{ki}) = \gamma \text{ for } j \neq k \tag{2}$$

where γ is a parameter lying between -1 and 1 that describes the strength of spillovers within the industry. In that case, the probability that two business units j and k locate in the same area i is independent from j and k and writes simply:

$$P(i,i) = E(u_{ji}u_{ki}) = \mathrm{Cov}(u_{ji}u_{ki}) + E(u_{ji})E(u_{ki}) = \gamma x_i(1 - x_i) + x_i^2 \tag{3}$$

Finally, the probability that pairs of plants locate in any same region is:

$$p = \sum_i P(i,i) = \gamma \, (1 - \sum_i x_i^2) + \sum_i x_i^2 \tag{4}$$

3.2. A natural estimator of geographic concentration

Using the linear relationship between p and γ, we can exhibit, for any industry, a simple estimator of the spillover parameter γ derived from a natural estimator of the probability p. We suggest to select the frequency estimator weighted by the size of plants. Whereas a simple frequency estimator would compare the number of pairs of plants located in each geographic area to the total number of pairs of plants in the country, the weighted estimator weighs each plant by its share in the industry employment (denoted by z_j). The choice of the latter estimator, that has our preference, means that we try to assess the frequency of the event {two workers belong to plants located in the same department} rather than the event {two plants are located in the same department}. Both estimators are unbiased but the weighted one is consistent with an Herfindahl measure of productive concentration (in terms of employment) that gives a higher weight to large business units, as will appear below. The reader should refer to Appendix A for a comparison of the simple and weighted estimators.

We thus suggest to use the weighted estimator that writes:

$$\hat{p} = \sum_i \frac{\sum_{\substack{j,k \in i \\ j \neq k}} z_j z_k}{\sum_{\substack{j,k \\ j \neq k}} z_j z_k} \tag{5}$$

$j,k \in i$ denoting the event {the business units j and k are located in region i}

Simple calculation of the sums in the formula (See Appendix A) leads to:

$$\hat{p} = \frac{\sum_i s_i^2 - H}{1 - H} \tag{6}$$

where $H = \Sigma_j z_j^2$ is the (employment) Herfindahl index of the industry, hence

$$\hat{\gamma} = \frac{\hat{p} - \sum_i x_i^2}{1 - \sum_i x_i^2} = \frac{\dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2} - H}{1 - H} \tag{7}$$

3.3. Comparison with the estimator by Ellison and Glaeser

Our estimator $\hat{\gamma}$ of γ is slightly different from the one suggested by Ellison and Glaeser (1994, 1997). In their papers, the estimator of γ derives from the *a priori* definition of a raw geographic concentration index G_{EG}[2]. This index is based on a comparison between the fraction of employment located in geographic area i for one industry (measured by s_i) and in the aggregate (measured by x_i):

$$G_{\mathrm{EG}} = \frac{\sum_i (s_i - x_i)^2}{1 - \sum_i x_i^2} \tag{8}$$

From this definition, Ellison and Glaeser build the following estimator of γ:

$$\hat{\gamma}_{\mathrm{EG}} = \frac{G_{\mathrm{EG}} - H}{1 - H} = \frac{\dfrac{\sum_i (s_i - x_i)^2}{1 - \sum_i x_i^2} - H}{1 - H} \tag{9}$$

It is straightforward to show that our estimator differs from Ellison and Glaeser's by a term whose expectation is equal to zero. Thus, both estimators are unbiased.

However, our estimator $\hat{\gamma}$ has a more natural specification than $\hat{\gamma}_{\mathrm{EG}}$ since it derives directly from the probability model. In particular, the Herfindahl index H that shows up in the expression of $\hat{\gamma}$ comes directly from the writing of the frequency estimator $\hat{p}$. The only difference between the estimators $\hat{\gamma}$ and $\hat{\gamma}_{\mathrm{EG}}$ lies in the measure of the raw concentration. With our estimator, we have

$$\hat{\gamma} = \frac{G_A - H}{1 - H} \text{ with } G_A = \frac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2}$$

whereas, in Ellison and Glaeser, they are given by Eqs. (8) and (9). Both G_A and G_{EG} can be interpreted as a measure of the raw geographic concentration of an industry since they are based on the comparison between the geographic patterns of employment for one industry (measured by s_i) and in the aggregate (measured by x_i), as well as other indices like the Gini index.

For both measures, it can be checked that: $E(G_m) = H + \gamma(1 - H)$, $m \in \{\mathrm{EG}, A\}$. This equation provides a new interpretation for the index γ that captured up to now the extent of spillovers or the correlation between business units' location choices. γ can also be interpreted as the excess of raw geographic concentration (G_m) on productive concentration (H) and therefore can be regarded as an index of the industry geographic concentration, controlling for the size distribution of plants. With this index, an industry will not be considered as localized only because its employment is concentrated in a small number of plants: an industry with a random distribution of plants across regions will have an expected γ index

[2] In their working paper, Ellison and Glaeser (1994) used the definition of G of Eq. (8), but in their published paper of 1997 they use the proportional and equivalent measure $G_{\mathrm{EG}} = \Sigma_i (s_i - x_i)^2$. We refer here to the definition of G of 1994.

equal to 0, regardless of the value of its Herfindahl index. This of course is not true for the raw geographic concentration indices G_m in either Ellison and Glaeser's or our definition. In what follows, we will refer to γ either as a spillover or a concentration index.

The model of natural advantage leads to the same probability of joint location for any pair of plants. The models of natural advantage and spillovers cannot therefore be identified separately[3]. In both models γ has an easy interpretation. When business units' location choices are independent, the expectation value of γ is zero ($E(\hat{\gamma}) = 0$). Thus, a value of γ greater than zero in one industry can be interpreted as a geographic concentration in excess of the one that would prevail if the location choices were independent between plants (no spillover) and random among regions (no natural advantage). The industry is therefore regarded as localized. The index also allows us to classify industries according to the strength of agglomerative forces.

4. The concentration of French manufacturing industries

4.1. The data

The computation of the geographic concentration index of French manufacturing industries relies on the Annual Business Survey (Enquête Annuelle d'Entreprise) launched at the business unit level by the French Ministry of Industry. This data set provides information on manufacturing employment[4], fields of activity (corresponding approximately to the U.S. 2- and 4-digit levels) and location measured along the two geographic subdivisions: regions (22 regions in France) and departments (95 departments). A total of 44,428 manufacturing plants have been investigated in 50 2-digit industries and 273 4-digit industries. The index of geographic concentration γ is computed for the year 1993. The availability of the whole data at the plant level ensures that aggregate variables for different geographic subunits are consistent. In Ellison and Glaeser (1997), for instance, employment data was only available at the level of the geographic unit and Herfindahl index had to be taken from another data source.

4.2. The localization of 4-digit manufacturing industries

We computed the index γ for each of the 4-digit industries at the department level. For almost all industries (270 out of 273), the index γ is statistically

[3] Natural advantages and spillovers can operate simultaneously. In this case the overall concentration measure is $\gamma = \gamma^s + \gamma^{na} - \gamma^s \gamma^{na}$ (Ellison and Glaeser, 1997).

[4] The data only covers the productive plants that are attached to a manufacturing firm employing more than 20 workers.

significant at 95% confidence level ($\gamma \neq 0$)[5]. Moreover, 211 industries (77%) display positive spillovers ($\gamma > 0$). French manufacturing industries appear therefore to be very localized and in most industries, the plants location decisions cannot be regarded as independent. Negative values for γ were found in 38 industries. Recalling that γ measures the correlation between the location decisions of two plants in the same industry, a negative value for γ means that dispersion forces dominate clustering forces. In other words, plants in the same industry try to be as scattered as possible.

As previously observed by Ellison and Glaeser for the United States, the degree of localization varies greatly from industry to industry (see Fig. 1): half of the industries display a low degree of concentration ($\gamma < 0.02$) while 23% have moderate concentration levels ($0.02 \leq \gamma \leq 0.05$) and 27% are very localized ($\gamma > 0.05$). The distribution of γ is also quite skewed, with a mean of 0.06 and a median of 0.01.

For purposes of comparison, the classification of industries with respect to concentration index γ was taken identical to Ellison and Glaeser's. The trigger values of 0.02 and 0.05 are somewhat arbitrary, but their magnitudes are discussed in Ellison and Glaeser (1997).

At the department level, the most localized 4-digit industries are extractive industries in which location decisions are highly influenced by the availability of raw materials (iron ore and coal, uranium ore, minerals for chemical industry and fertilizers). As could be expected, shipbuilding industries are also highly localized in departments that have access to the sea while the location of traditional industries is determined by the historical specialization of some regions: cotton and wool mills; knitting industry; footwear; leather products; watch-making; toys; sport equipment.

As displayed in Tables 1 and 2, a high degree of geographic concentration can

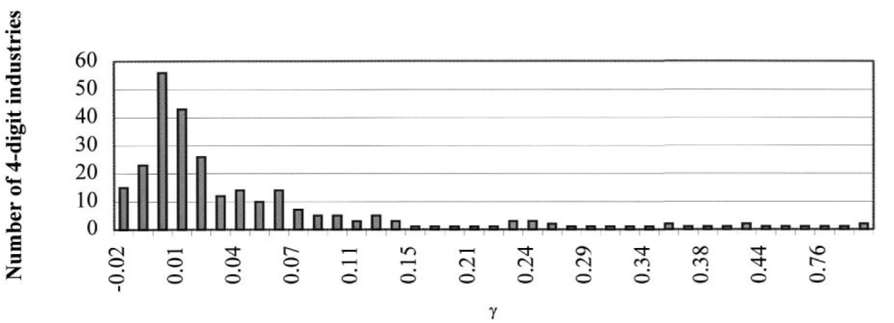

Fig. 1. Histogram of γ at the department level.

[5]The statistical tests uses the variance of the estimator $\hat{\gamma}$ under H_0 ($\gamma = 0$). The expression for this variance is computed in Appendix A.

Table 1
Most localized industries

4-digit industry (French NAF 700)	Index of geographic concentration γ	Share of the industry employment in the first department
Extraction of slate	0.88	95.1
Extraction of iron ore	0.88	96.5
Made-to-measure clothing	0.80	89.8
Extract of minerals for chemical industry and fertilizers	0.76	91.7
Steel pipe and tubes	0.69	88.6
Extraction of coal	0.53	77.6
Combed wool spinning mills	0.44	68.7
Vehicles hauled by animals	0.42	72.5
Wool preparation	0.42	61.8
Periodicals	0.40	62.1
Watch-making	0.38	62.6
Flat glass	0.37	77.9
Screw cutting	0.36	60.6
Lawn and garden equipment	0.36	68.5
Carded wool weaving mills	0.34	56.0
Essential oils	0.32	59.6
Book publishing	0.30	56.2
Extraction of uranium ore	0.29	57.0
Cutlery	0.28	53.4
Carded wool spinning mills	0.25	45.7
Small arms	0.25	42.2
War vessels	0.24	62.9
Sound recording	0.24	46.9
Cotton spinning mills	0.24	43.6

Table 2
Geographic concentration

2-digit industry (French NAF 100)	% of 4-digit industries with		
	$\gamma < 0.02$	$0.02 \leq \gamma \leq 0.05$	$\gamma > 0.05$
Shipbuilding	0	0	100
Textile yarns and fabrics	0	8	92
Knit fabric and articles	0	33	67
Radio and television communications equipment	0	50	50
Railroad equipment	0	100	0
Soap, perfumes and cleaning products	0	100	0
Office machinery and data processing	0	100	0
Jewelry and musical instruments	0	67	33
Sporting goods, toys and miscellaneous industries	20	20	60
Textile goods	29	14	57
Iron and steel	25	38	38
Leather goods and footwear	33	0	67
Hard coal mines, extraction of uranium ores, peat	33	0	67
Photographic and optical instruments, watches and clocks	33	0	67
Aircraft and space equipment	33	67	0
Pharmaceutical goods	33	33	33
Other machines of particular use	34	44	22
Apparel, clothing accessories and fur goods	37	25	38
Industrial chemicals, man-made and synthetic fibers	37	50	13
Extraction of quarry products and minerals	38	37	25

Extraction of metalliferous ores	50	0	50
Arms and ammunition	50	0	50
Industrial services for metal work	50	17	33
Printing and publishing	50	17	33
Motorcycles and other transportation equipment	50	25	25
Mechanical equipment	50	33	17
Medicinal products	50	25	25
Boilers and tanks	50	50	0
Building materials	53	34	13
Ceramic products and floor tiles	57	14	29
Metal products	56	22	22
Paper and board products	56	44	0
Scientific and controlling instruments and apparatus	60	0	40
Processing of glass and glass products	67	0	33
Power generating machinery and equipment	67	33	0
General industrial machinery	71	29	0
Furniture	75	0	25
Non-ferrous metals	75	13	12
Electrical equipment	80	0	20
Processing of plastic	80	20	0
Machine tools	80	20	0
Wood products	83	17	0
Rubber products	100	0	0
Electronical components	100	0	0
Foundries	100	0	0
Metal work for building	100	0	0
Motor vehicles	100	0	0
Farm machinery	100	0	0
Sound recording and reproducing apparatus and equipment	100	0	0
All industries	**50**	**23**	**27**

also be found for clothing industry and book publishing (in Paris), fur goods, iron and steel. Finally, several high technology industries appear to be localized, such as the radio and television communication equipment that is mainly located in Paris suburbs.

The least localized products are motor vehicles, sound recording and reproducing apparatus, farm machinery, electronical components, rubber products, metal work for construction and non-ferrous metals. Other fine products (peat, ceramic and pottery products) display a very low level of geographic concentration. For most of these products, it appears very insightful to distinguish the measure of geographic concentration from the sole concentration of the production, related to returns to scale. Indeed, a low degree of geographic concentration must not be interpreted as the fact that the industry is actually scattered all over the country. In most cases, products are regarded as not localized only because their geographic concentration is largely lower than what could have been expected from the high level of concentration of their production.

If the index of concentration at the 4-digit level seems to support the idea of a high correlation between firms' location decisions in the same industry, the results should be interpreted carefully. Clearly, our index provides a static and unconditional measure of concentration that tends to overweigh the past and is not really fitted to measure dynamic externalities. High levels of concentration can therefore correspond to different localization strategies. In particular, the high degree of concentration in traditional industries should be the result of past static externalities whose effect still prevails today although the current dynamic may tend to reverse this process by favoring the growth of more diversified areas. On the other hand, the high level of concentration in high technology industries may derive from strong current dynamic knowledge spillovers.

5. The scope of spillovers is sensitive to industry and geographic definitions

5.1. Intra- and inter-industry concentration

The previous sections highlighted the role of technological spillovers in influencing the location decision of industry plants. However, the scope of these spillovers were limited to firms belonging to the same industry. In this section, we want to relax this assumption in order to determine whether a business unit only finds advantage to locate near other plants in the same subindustry or benefits from the proximity of plants working in related fields of activity.

To measure the interactions between the location choice of business units, one can assume that there exists not only spillovers between business units in the same 4-digit subindustry (measured by the index γ corresponding to the 4-digit classification) but also spillovers between business units in different 4-digit subindustries within the same 2-digit industry. These latter spillovers figured by the parameter γ_0, that measures the correlation between the location choice of business units in different 4-digit subindustries within the same 2-digit industry. It can be shown (Ellison and Glaeser, 1997) that the index of geographic concentration in a 2-digit industry accounts for both intra- and inter-industry spillovers and is equal to:

$$\gamma^{100} = \frac{\overbrace{\sum_{j=1}^{r} \gamma_j^{700} \omega_j^2 (1 - H_j^{700})}^{\text{Intra-industry concentration}} + \overbrace{\gamma_0 \left(1 - \sum_{j=1}^{r} \omega_j^2 \right)}^{\text{Inter-industry concentration}}}{1 - \sum_{j=1}^{r} \omega_j^2 H_j^{700}} \tag{10}$$

where r denotes the number of 4-digit subindustries within the 2-digit industry, ω_j the share of the j^{th} subindustry in the 2-digit industry employment, γ_j^{700} the geographic concentration index and H_j^{700} the Herfindahl index (i.e. productive concentration) in this subindustry.

The formula (10) tells us that the geographic concentration index γ^{100} for the 2-digit industry is a weighted mean of the inter- and intra-industry concentration indices. We computed the concentration index γ for the French 2-digit and 4-digit industry definitions (see Table 3). Agglomeration forces appear to have a stronger effect at the finest industry level (4-digit subindustries), with an estimated index of 0.058 compared to 0.036 for the 2-digit level. This result is consistent with the assumption that spillovers are stronger when business units work in the same field of activity.

In most cases, the most localized 2-digit industries encompass the 4-digit subindustries that were found to be most localized in Section 4.2 (extractive industries, textile industry, leather goods and footwear, shipbuilding, optical instruments and watches, printing and publishing, iron and steel)[6]. Furthermore, there are more high technology industries that display a high degree of geographic concentration at the 2-digit level. This is notably the case of the scientific and controlling instruments, aircraft and space equipment, pharmaceutical industry, office machinery and data processing. This result gives some support to the idea that business units in different subindustries may find advantage to locate in the same place to benefit from research spillovers or highly qualified workers.

In the case of extractive industries, the 2-digit concentration only comes from the joint location decision of plants in the same 4-digit subindustries (coal, uranium ore, lignite, metalliferrous ores). The fraction of inter-industry concentration is not significant or even negative. This should not be surprising since

[6] When there is no difference between 2- and 4-digit industries, 2-digit industries are ruled out for the computation. This is notably the case for sound recording and reproducing apparatus as well as railroad equipment.

Table 3
Geographic concentration of NAF 100 industries and share of between-industries concentration

French NAF 100 (2-digit industries)	Index γ of geographic concentration	Fraction of intra-industry concentration (%)
Hard coal mines, extraction of uranium ores, peat	0.441	0
Extraction of metalliferous ores	0.300	−4
Textile yarns and fabrics	0.079	79
Photographic and optical instruments, watches and clocks	0.074	−1
Knit fabric and articles	0.055	43
Pharmaceutical goods	0.049	14
Soap, perfumes and cleaning products	0.044	36
Shipbuilding	0.042	15
Office machinery and data processing	0.039	−4
Textile goods	0.037	70
Printing and publishing	0.037	74
Iron and steel	0.034	33
Scientific and controlling instruments and apparatus	0.033	61
Radio and television communication equipment	0.030	34
Leather goods and footwear	0.027	−6
Aircraft and space equipment	0.024	54
Jewelry and musical instruments	0.023	11
Apparel, clothing accessories and fur goods	0.018	68
Industrial chemicals, man-made and synthetic fibers	0.015	69

Sporting goods, toys and miscellaneous industries	0.014	29
Household appliances	0.010	10
Ceramic products and floor tiles	0.010	−22
Boilers and tanks	0.010	13
Medicinal products	0.009	45
Electronical components	0.008	35
Motorcycles and transportation equipment	0.008	−100
Power generating machinery and equipment	0.007	51
Other machines of particular use	0.007	64
Processing of glass and glass products	0.006	12
Motor vehicles	0.006	52
Mechanical equipment	0.006	6
Electrical equipment	0.005	24
General industrial machinery	0.005	31
Industrial services for metal work	0.005	3
Arms and ammunition	0.004	−96
Paper and board products	0.003	50
Processing of plastic	0.003	42
Extraction of quarry products and minerals	0.002	−265
Foundries	0.001	170
Farm machinery	0.001	234
Non-ferrous metals	0.001	−123
Machine tools	−0.002	90
Metal products	−0.002	152
Furniture	−0.003	124
Wood products	−0.003	110
Metal work for building	−0.003	80
Rubber products	−0.004	29
Building materials	−0.004	107

the location decision in these industries is mainly influenced by the availability of resources that are specific to each subproduct. Inter-industries spillovers are also low for optical instruments and watches, leather goods and footwear, motorcycles and office machinery. These results are consistent with the 4-digit analysis that showed that the corresponding subindustries were not localized in the same departments (Table 1).

For printing and publishing or textile and apparel industries, subindustry plants find advantage to locate near plants in other subindustries: the inter-subindustry concentration accounts for more than two-thirds of the overall concentration. In these industries, it may lower transportation costs and improve productive efficiency to gather in the same area business units that perform different steps of the production process (spinning–weaving–finishing; typesetting–printing–bookbinding–publishing) or plants that share the same inputs (yarn and fabrics for clothing and textile industries, for instance).

But this approach only captures part of the interactions across industries in location choice. In particular, it does not allow us to measure the likely technological spillovers between 4-digit industries in different 2-digit industries when both industries share specialized inputs such as highly qualified workers or fundamental research. A better approach would require to gather industries according to criteria more relevant to location decisions, like upstream or downstream relationships as in Ellison and Glaeser (1997).

5.2. The geographic scope of spillovers

When the geographic scope of natural advantage or spillovers is very limited, the density of agglomeration should be high and the measure of localization not very sensitive to definition of the geographic area (region or department)[7]. On the other hand, when the effects of spillovers decline slowly with distance, the choice of the geographic level should matter in determining the degree of localization of an industry.

To address this question, we computed the index γ for two geographic subdivisions corresponding to the French region and department. The concentration is substantially higher at the region level with a mean value of 0.09 against 0.06 for department concentration. This indicates that the scope of spillovers seems to go beyond the limit of the department. In order to confirm this intuition, and answer the question of the geographic scope of spillover, we suggest a more formalized approach, modifying slightly the location model suggested by Ellison and Glaeser.

Thus, we assume here that the location decision is a two-stage process: in the first stage, business units choose the region in which to locate according to the spillovers (or natural advantages parameter) γ_0 at the region level; in the second

[7] Ellison and Glaeser (1997) show that the index is invariant with spatial aggregation.

Table 4
Shares of region-level and intra-region spillovers in the department-level concentration

French NAF 100 (2-digit industries)	Concentration index γ_1	Fraction of region-level spillovers (%)	Fraction of intra-region spillovers (%)	Fraction of the cross product (%)
Radio and television communication equipment	0.046	109.2	−2.1	−7.1
Machine tools	0.005	105.7	−1.4	−4.3
Jewelry and musical instruments	0.031	100.9	4.6	−5.5
Other machines of particular use	0.011	91.7	3.6	4.7
Electrical equipment	0.007	89.1	5.9	5.0
General industrial machinery	0.010	82.4	13.1	4.5
Sporting goods, toys and miscellaneous industries	0.012	81.1	17.8	1.1
Scientific and controlling instruments and apparatus	0.060	76.2	7.5	16.3
Printing and publishing	0.044	75.0	14.3	10.7
Soap, perfumes and cleaning products	0.065	70.7	8.9	20.4
Medicinal products	0.018	70.1	21.5	8.4
Textile goods	0.017	65.2	22.5	12.3
Pharmaceutical goods	0.077	55.8	18.0	26.2
Office machinery and data processing	0.077	54.1	15.7	30.2
Textile yarns and fabrics	0.051	53.0	21.5	25.5
Iron and steel	0.026	51.0	34.5	14.5
Photographic and optical instruments, watches and clocks	0.074	50.4	16.4	33.2
Electronical components	0.013	46.7	42.0	11.3
Household appliances	0.012	46.6	55.7	−1.7
Mechanical equipment	0.004	39.9	54.3	5.8
Industrial services for metal work	0.004	39.1	56.2	4.7
Leather goods and footwear	0.022	38.7	45.7	15.6
Knit fabric and articles	0.030	31.7	49.9	13.4
Industrials chemicals, man-made and synthetic fibers	0.011	27.0	69.1	3.9
Processing of glass and glass products	0.006	23.8	75.9	0.3
Motor vehicles	0.005	4.1	95.1	0.8
Apparel, clothing accessories and fur goods	0.023	2.2	96.7	1.1

stage, they choose to locate in one specific subregion (department in the French case) within the region, according to the intra-regional spillovers γ_i, specific to department i. It can be shown that the resulting index of geographic concentration at the department level is a weighted mean of spillovers γ_0 at the region level, intra-regional spillovers γ_i and the cross product $\gamma_0\gamma_i$ (see Appendix B for the detailed computation):

$$\gamma_1 = \overbrace{\lambda_0\gamma_0}^{\text{spillover at regional level}} + \overbrace{\sum_i \lambda_i\gamma_i}^{\text{intra-regional spillover}} + \overbrace{\sum_i \mu_i\gamma_0\gamma_i}^{\text{cross-product}} \tag{11}$$

with $\lambda_0, \lambda_i, \mu_i$ such as $\lambda_0 + \sum_i (\lambda_i + \mu_i) = 1$

The ranking of industries derived from the two-stage model is roughly the same as the one obtained with the one-stage model, at the department level (see Table 3 for the one-stage model and Table 4 for the two-stage model). But we are now able to separate the contribution of broader and closer spillovers. As displayed in Table 4, the contribution of regional spillovers to concentration at the department level is often significant. This result supports the view that the benefits of agglomeration go beyond the limits of the department. This conclusion seems particularly appealing for some overall concentrated industries, like high technology industries (communication equipment, scientific and controlling instruments) and, to a lesser extent, for office machinery and electronical components, which exhibits smaller geographic concentration. Although face-to-face communication requires small distance, the small size of French regions and the good quality of transportation infrastructure may explain the benefits of gathering industries at the wider region level.

6. Comparison of different indices on French data and comparison with U.S. results

The index of geographic concentration used in this paper is slightly different from the one proposed by Ellison and Glaeser. Both have the attractive feature of measuring the localization beyond the sole concentration of production and provide unbiased estimators of the spillover or natural advantage parameter. It is not possible to compare the theoretical properties of the two estimators since it would require a full modelling of the location decision. But we can at least compare the empirical results derived from the two estimators in the case of France to assess the robustness of the two measures of geographic concentration. Moreover, whereas Ellison and Glaeser put forward the advantage of their index over a Gini index, they do not perform theoretical or empirical comparison of the two indices. Here we also computed Gini indices for 2-digit industries in France and compare the three indices.

Measures of geographic concentration are also useful to identify industries where agglomeration forces are important. If they are relevant, one should empirically find similar patterns of geographic concentration in different developed countries. Using the most comparable industry definition between France and the U.S., we thus attempted to compare our results with those of Ellison and Glaeser for the United States.

6.1. Comparison of different indices for France

The measures of concentration obtained with the Ellison and Glaeser's estimator do not lead to significant changes in the identification of the most localized industries[8]. Our estimator seems therefore sufficiently robust for these industries. The most important differences are observed for the least-localized industries. For instance, glass and motor vehicle industries display a lower degree of concentration with Ellison and Glaeser's index whereas quarry products and wood industries appear more concentrated (see Table 5).

Conversely, when one compares the Gini measure of geographic concentration with ours, the hierarchy of industries is broadly different: the rank correlation between the two indices only amounts to 0.5. The Gini index is indeed correlated to the concentration of production (the rank correlation between Gini and Herfindhal indices is 0.8). Hence, the main differences between the two indices are observed for industries displaying a high concentration of production but low correlated plants location decisions (arms and ammunition, rubber products, farm machinery) or for industries with low concentration of production but highly correlated firms location decisions (printing and publishing, boilers and tanks, pharmaceutical goods, scientific and controlling instruments) (Table 5).

6.2. Comparison of the results on French data with American results

It is not easy to compare the results of Ellison and Glaeser for the United States with our results for France. First, as already mentioned, our index of concentration is slightly different. But this is not the major point as the previous paragraph showed that there were only minor differences between the two indices on French data. Second, industry definitions are not identical in the two countries. Actually, what we abusively called 2-digit classification has 50 manufacturing[9] industries and the finest level in France, which we called 4-digit has 273 industries, so our

[8] The rank correlation between the two indices amounts to 0.9.

[9] Manufacturing industries are defined in a narrower sense in France. They exclude food processing and energy production.

Table 5
Geographic concentration at the department level according to different indices (Ellison and Glaeser index, Gini index, our index)

French NAF 100 (2-digit industries)	Ellison–Glaeser index γ_{EG}	Rank according to Gini index	Rank according to γ
Hard coal mines, extraction of uranium ores, peat	0.439	2	1
Extraction of metalliferous ores	0.304	1	2
Photographic and optical instruments, watches and clocks	0.083	6	4
Textile yarns and fabrics	0.066	10	3
Knit fabric and articles	0.062	9	5
Shipbuilding	0.054	3	8
Leather goods and footwear	0.040	14	15
Pharmaceutical goods	0.036	24	6
Soap, perfumes and cleaning products	0.033	19	7
Printing and publishing	0.032	42	11
Textile goods	0.030	27	10
Iron and steel	0.026	13	12
Aircraft and space equipment	0.025	7	16
Scientific and controlling instruments and apparatus	0.025	32	13
Jewelry and musical instruments	0.024	12	17
Radio and television communication equipment	0.022	18	14
Ceramic products and floor tiles	0.020	11	22
Apparel, clothing accessories and fur goods	0.020	33	18
Sporting goods, toys and miscellaneous industries	0.020	26	20
Office machinery and data processing	0.019	16	9
Household appliances	0.017	8	23
Electronical components	0.015	21	26

Extraction of quarry products and minerals	0.014	31	38
Motorcycles and transportation equipment	0.014	5	25
Foundries	0.014	22	39
Wood products	0.012	28	45
Processing of plastic	0.010	40	37
Industrial chemicals, man-made and synthetic fibers	0.010	30	19
Mechanical equipment	0.010	35	30
Industrial services for metal work	0.010	44	34
Farm machinery	0.009	15	41
Electrical equipment	0.008	39	32
Furniture	0.008	37	44
Rubber products	0.008	17	48
Paper and board products	0.007	38	36
Other machines of particular use	0.007	41	27
Boilers and tanks	0.006	45	21
Chemical products	0.006	36	24
Power generating machinery and equipment	0.006	25	28
Metal work for building	0.004	43	46
Metal products	0.004	46	43
Machine tools	0.004	29	42
Processing of glass and glass products	0.004	20	31
Building materials	0.003	47	47
General industrial machinery	0.002	48	33
Non-ferrous metals	0.002	23	40
Motor vehicles	0.000	34	29
Arms and ammunition	−0.052	4	35

2-digit (resp. 4-digit) is intermediate between 2 and 3-digit (resp. 3 and 4-digit) in U.S. definitions.

When we perform the comparison at the finest level (4-digit for the U.S. and NAF 700 for France) for the industries with closest definition in the two countries, we can find some convergence in the identification of highly localized industries (clothing, leather industries) and least-localized ones (rubber, electrical equipment, paper and board). But a precise comparison of the two results is much harder, since industry definitions are not comparable enough. To compare more precisely the ranking of industries in the two countries with respect to geographical concentration, we built for France a grouping of industry as close as possible to the intermediate grouping 2-digit for the U.S. (complete 3-digit equivalent was out of scope), starting from the French level NAF 100. Then we computed concentration indices for these groupings on French data. The results for the 17 manufacturing industries are presented in Table 6 together with Ellison and Glaeser's results for the U.S.

The overall correlation between the indices in the two countries amounts to 0.6 (for rank correlation as well as level correlation) which is not so low. Textile and leather products appear to be the more concentrated industries in the two countries whereas industrial machinery, stone, clay and glass products and fabricated metal products are the least localized. The main differences in industries hierarchy can be observed for furniture and fixtures, and transportation equipment that are both much more concentrated in U.S. and for printing and publishing that are more

Table 6
Geographic concentration in United States and France according to Ellison–Glaeser index

2-digit industries (U.S. definition)	USA		France	
	γ	Rank	γ	Rank
Textile mill products	0.127	1	0.036	2
Leather and leather products	0.029	2	0.039	1
Furniture and fixtures	0.019	3	0.008	10
Lumber and wood products	0.018	4	0.012	8
Primary metal industries	0.018	5	0.010	9
Instruments and related products	0.018	6	0.018	5
Transportation equipment	0.016	7	0.000	17
Apparel and other textile products	0.016	8	0.020	4
Miscellaneous manufacturing ind.	0.012	9	0.014	6
Chemicals and allied products	0.009	10	0.012	7
Paper and allied products	0.006	11	0.007	11
Electronic and other electrical equipment	0.005	12	0.004	13
Printing and publishing	0.005	13	0.032	3
Fabricated metal products	0.005	14	0.003	14
Rubber and misc. plastics	0.004	15	0.006	12
Stone, clay and glass products	0.004	16	0.003	15
Industrial machinery and equipment	0.003	17	0.002	16

Table 7
Geographic concentration in United States and France for specific 3-digit industries

	3-digit industries	USA γ	France γ
251	Household furniture	0.058	0.010
252	Office furniture	0.027	0.006
271	Newspapers	0.002	0.001
272	Periodicals	0.067	0.371
273	Books	0.025	0.292
278	Blankbooks and bookbinding	0.011	0.018
371	Motor vehicles and equipment	0.083	0.000
372	Aircraft and parts	0.024	0.018
373	Ship and boat building and repairing	0.018	0.044
374	Railroad equipment	0.123	0.045
375	Motorcycles, bicycles and parts	0.010	0.007
376	Guided missiles, space vehicles, parts	0.196	0.039

localized in France. To further investigate these differences, we compared the French and American results for the 3-digit industries belonging to these groups[10]. Whereas aircraft displays roughly the same index of concentration in U.S. and France, motor vehicles, and railroad equipment and space industries make the difference (see Table 7). It is also interesting to note that high tech industries (instruments and related products, electronic and other electric equipment, aircraft industries) display roughly the same ranking with the two measures (Tables 6 and 7).

However, this similarity between geographic concentration in France and the U.S. for specific industries should not be interpreted too rapidly, as the proposed measure of geographic concentration has two different interpretations in Ellison and Glaeser's modelling and is not based on structural modelling of plant's location decisions. Overall similarity can arise from different histories in the two countries.

7. Conclusion

Our empirical investigation confirms the interdependence of firms' location choice. It allowed us to identify three types of highly localized industries. A first group is composed of extractive industries whose localization seems mostly determined by access to raw materials or more generally industries depending on physical geography like shipbuilding. Traditional industries (textile and leather)

[10] The authors kindly provided their 3-digit indices, not published in Ellison and Glaeser (1994, 1997).

belong to the second group. For these industries, the initial location choice often dates from the industrial revolution but subsequent external effects may have contributed to reinforce it. Finally, a third group includes high technology industries for which knowledge spillovers seem to be high within industries.

While our modelling does not say much about urbanization externalities, which may also be important for high technology industries, the results suggest that agglomeration effects can exist also between different industries, based on sectoral grouping. Spillovers may also be important within a relatively wide area. Lastly, comparison with results for the United States confirms the identification of the most and least localized industries, with the notable exception of motor vehicles on the one hand and of printing and publishing on the other. High technology industries also display similar rankings in the two countries. On the whole, empirical results do not seem to be very sensitive to the precise choice of concentration index, provided that it takes into account the distribution of plant size.

Appendix A. The index of geographic concentration: a comparison with Ellison and Glaeser

Ellison and Glaeser (1994, 1997) propose an index of geographic concentration γ derived from a model of the location choice of N industrial business units across M geographic areas. To estimate this parameter γ, the authors suggest to select an indicator based on a pre-defined normalized measure G referred as the raw geographic concentration of the industry. While this estimator has a number of attractive features such as being unbiased for both probability models of natural advantage and spillovers, its computation is somehow ‘ad hoc’.

A rigorous derivation of the estimator would require a joint modelling of the random variables representing the location choice of the N plants. This would be very intricate and is out of the scope of this paper. However, even when we limit our modelling to the location choice of any pair of plants, there exists more natural unbiased estimators of the parameter that are also identical in the models of natural advantage and spillovers.

In the model of spillovers described in the text , the probability that two business units locate in the same geographical area is:

$$p = \sum_{\text{region } i} P(i,i) = \gamma \left(1 - \sum_{\text{region } i} x_i^2\right) + \sum_{\text{region } i} x_i^2 \tag{A.1}$$

We can now propose a natural estimator of p from which we will derive the estimator of γ. This estimator consists in approximating p by the frequency of the event, weighting the business units by their size z_j.

This estimator is:

$$\hat{p} = \sum_i \frac{\sum_{\substack{j,k \in i \\ j \neq k}} z_j z_k}{\sum_{\substack{j,k \\ j \neq k}} z_j z_k} \tag{A.2}$$

$j,k \in i$ denoting the event {the business units j and k are located in region i}.

It is straightforward to see that

$$\sum_{\substack{j,k \in i \\ j \neq k}} z_j z_k = s_i^2 - \sum_{j \in i} z_j^2,$$

where $s_i = \Sigma_{j=1}^{N} z_j u_{ji}$ denotes the fraction of the industry employment located in geographic area i. It comes that:

$$\hat{p} = \frac{\sum_i s_i^2 - H}{1 - H} \tag{A.3}$$

where $H = \Sigma_j z_j^2$ is the Herfindahl index.

The estimator of γ is thus:

$$\hat{\gamma} = \frac{\hat{p} - \sum_i x_i^2}{1 - \sum_i x_i^2} = \frac{\dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2} - H}{1 - H} \tag{A.4}$$

This estimator is similar to the one proposed by Ellison and Glaeser where:

$$\hat{G} = \frac{\sum_i (s_i - x_i)^2}{1 - \sum_i x_i^2}$$

is replaced by

$$\hat{G}_A = \frac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2}$$

$\hat{\gamma}$ is an unbiased estimator of γ since:

$$E(\hat{\gamma}) = \frac{\dfrac{E(\sum_i s_i^2) - \sum_i x_i^2}{1-\sum_i x_i^2} - H}{1-H} = \frac{\dfrac{\sum_i E((\sum_j z_j u_{ji})^2) - \sum_i x_i^2}{1-\sum_i x_i^2} - H}{1-H}$$

$$E(\hat{\gamma}) = \frac{\dfrac{\sum_i (x_i H + (1-H)(\gamma x_i(1-x_i) + x_i^2)) - \sum_i x_i^2}{1-\sum_i x_i^2} - H}{1-H} = \gamma$$

This estimator, just as the G-based estimator of Ellison and Glaeser, measures the geographic concentration in excess of productive concentration. Our estimator differs from the Ellison and Glaeser's estimator by the term

$$\frac{\sum_i x_i(s_i - x_i)}{(1-H)(1-\sum_i x_i^2)}$$

whose expectation is equal to zero.

We have also:

$$E(\hat{G}_A) = E(G) = H + \gamma\,(1-H). \tag{A.5}$$

A comparison of the properties of these two estimators would require a joint modelling of the variables u_{ji} for $i=1,2\ldots,M$ and $j=1,2,\ldots,N$. We can nonetheless compute the variance of G_A under the assumption H_0 of random localization ($\gamma=0$) as in Ellison and Glaeser.

After computation, we have:

$$\mathrm{Var}(G_A) = \mathrm{Var}(G) + \frac{4}{\left(1-\sum_i x_i^2\right)^2}\left((2\sum_j z_j^3 - H)(-\sum_i x_i^3 + (\sum_i x_i^2)^2)\right) \tag{A.6}$$

where

$$\mathrm{Var}(G) = \frac{2}{\left(1-\sum_i x_i^2\right)^2}\left(H^2\left(\sum_i x_i^2 - 2\sum_i x_i^3 + (\sum_i x_i^2)^2\right) - \sum_j z_j^4\left(\sum_i x_i^2 - 4\sum_i x_i^3 + 3\,(\sum_i x_i^2)^2\right)\right) \tag{A.7}$$

We could also consider an unweighted estimator of γ. If n_i denotes the number of industry plants in region i, it is straightforward to show that the frequency estimator writes in that case:

$$\hat{p} = \sum_i \frac{n_i(n_i - 1)}{N(N-1)} = \frac{\sum_i \left(\frac{n_i}{N}\right)^2 - \frac{1}{N}}{(1 - \frac{1}{N})} \tag{A.8}$$

Eq. (A.8) is identical to Eq. (A.3) after replacing s_i by $\frac{n_i}{N}$ and H by $\frac{1}{N}$. The first two terms both measure the fraction of region i in the industry, in terms of employment for s_i and in terms of number of plants for $(\frac{n_i}{N})$. It is interesting to note that $\frac{1}{N}$ can also be interpreted as an index of concentration.

We also have:

$$\hat{\gamma} = \frac{\hat{p} - \sum_i x_i^2}{1 - \sum_i x_i^2} = \frac{G_B - \frac{1}{N}}{1 - \frac{1}{N}} \quad \text{with} \quad G_B = \frac{\sum_i \left(\frac{n_i}{N}\right)^2 - \sum_i x_i^2}{1 - \sum_i x_i^2}$$

A straightforward calculation shows that

$$E(G_B) = \frac{1}{N} + \gamma(1 - \frac{1}{N}) \tag{A.9}$$

Hence, $\hat{\gamma}$ is also an unbiased estimator of γ. Moreover, Eq. (A.9) says that this estimator can still be interpreted as the geographic concentration in excess of the productive concentration when one measures the latter by the inverse of the number of plants instead of the Herfindahl index.

Appendix B. A sequential model of location choice

Let us consider two levels of geographic units:

1. the region level ($i = 1,2, \ldots ,M$), that will be composed for instance of the French regions
2. the subregion level $m(i) = 1,2, \ldots ,r_i$, that will be composed of the French departments

The business units' location decision is a two stage process:

In the first stage, business units take into account the natural advantages of geographic areas and the intra-industry spillovers to choose the region in which to locate. γ_0 denotes intra-regional spillovers.

In the second stage, business units choose a more specific location, within each region i, according to the spillover model at the region level where g_i denotes intra-industry spillovers in region i.

Some definitions

x_i is the share of the region i in total industrial employment

x_{im} is the share of the subregion $m(i)$ in total industrial employment

y_{im} is the share of the subregion $m(i)$ in regions i's employment

We have $x_{im} = x_i\ y_{im}$ and $\Sigma_{i,m}\ x_{im}^2 = \Sigma_{im}\ x_i^2 y_{im}^2 = \Sigma_i\ x_i^2 \Sigma_m\ y_{im}^2 = \Sigma_i x_i^2 \Omega_i$ where $\Omega_i = \Sigma_{m=1}^{r_i}\ y_{im}^2$

The probability model

In the first stage, if n_j is the random variable that describes the location of business unit j (i.e. $n_j = i$, $i = 1,2, \ldots ,M$ if j is located in region i, 0 otherwise) and u_{ji} the associated Bernouilli variable (0,1), it follows from the spillover model that:

$$E(u_{ji}) = x_i, V(u_{ji}) = x_i\ (1 - x_i) \text{ and } \mathrm{Corr}(u_{ji},u_{ki}) = \gamma_0 \text{ for } j \neq k.$$

In the second stage, the spillover model still applies but is **conditional to the first stage**: If w_{jim} is the Bernouilli variable associated to the location of business unit j in (i,m), we have:

$$E(w_{jim}/u_{ji} = 1) = y_{im}, V(w_{jim}/u_{ji} = 1)$$
$$= y_{im}(1 - y_{im}) \text{ and } \mathrm{Corr}(w_{jim}, w_{kim}/u_{ji} = 1,u_{ki} = 1) = \gamma_i$$
$$\text{for } j \neq k$$

Computation of the estimator of concentration at the finest geographic level (subregion-level): Just as in the one stage model, the estimator of concentration is derived from the probability that two business units choose to locate in the same subregion.

From the sequential model above, the probability that two business units (j and k) locate in the same subregion m of region i is the product of the probability that the business units locate in $m(i)$, knowing that they previously chose to locate in region i, by the probability that they simultaneously decided in the first stage to locate in region i. For $j \neq k$, it writes:

$$P(j \in m(i), k \in m(i)) = P(j,k \in m(i)/j,k \in i).P(j,k \in i)$$

From Section 3, this can be rewritten:

$$P(m(i),m(i)) = (\gamma_0 x_i(1 - x_i) + x_i^2)(\gamma_i y_{im}(1 - y_{im}) + y_{im}^2)$$

The probability (independent on j and k) that pairs of plants locate in the same subregion is:

$$p = \sum_{i,m} P(m(i), m(i)) = \sum_{i,m} (\gamma_0 x_i (1 - x_i) + x_i^2)(\gamma_i y_{im}(1 - y_{im}) + y_{im}^2)$$

or $p = \Sigma_i (\gamma_0 x_i (1 - x_i) + x_i^2)(\gamma_i (1 - \Omega_i) + \Omega_i)$

It follows that:

$$p = \gamma_0 \sum_i x_i (1 - x_i) \Omega_i + \sum_i \gamma_i x_i^2 (1 - \Omega_i) + \sum_i \gamma_0 \gamma_i x_i (1 - x_i) \Omega_i + \sum_i x_i^2 \Omega_i \tag{A.10}$$

In the case where the business units' decision process is one-stage with spillovers represented by the parameter γ_1 at the department level, the probability that two business units locate in the same department write:

$$p = \gamma_1 \left(1 - \sum_{i,m} x_{im}^2 \right) + \sum_{i,m} x_{im}^2 \tag{A.11}$$

Comparing the two latter formulas, it can be seen that the sequential localization model corresponds to a one-stage location process with a concentration index γ_1 at the department level such that:

$$\gamma_1 = \frac{\gamma_0 \sum_i x_i (1 - x_i) \Omega_i}{\left(1 - \sum_i x_i^2 \Omega_i \right)} + \frac{\sum_i \gamma_i x_i^2 (1 - \Omega_i)}{\left(1 - \sum_i x_i^2 \Omega_i \right)} + \frac{\sum_i \gamma_0 \gamma_i x_i (1 - x_i)(1 - \Omega_i)}{\left(1 - \sum_i x_i^2 \Omega_i \right)}$$

or

$$\gamma_1 = \lambda_0 \gamma_0 + \sum_i \lambda_i \gamma_i + \sum_i \mu_i \gamma_0 \gamma_i$$

$$\text{with } \lambda_0, \lambda_i, \mu_i \text{ such that } \lambda_0 + \sum_i (\lambda_i + \mu_i) = 1 \tag{A.12}$$

This specification has a straightforward interpretation. Just recall that the probability that two business units locate in the same region is a linear function of γ in the simplest one-stage spillover model. In the sequential timing considered here, the non-conditional probability is the cross product of two linear probabilities, that is the cross product of the linear function γ_0 and the linear function γ_i.

It can be checked that if $\gamma_i = 0$, $\forall i$ (there is no intra-regional spillovers, so that the location decision boils down to the region choice) then

$$\gamma_1 = \frac{\gamma_0 \sum_i x_i (1 - x_i) \Omega_i}{\left(1 - \sum_i x_i^2 \Omega_i \right)} < \gamma_0.$$

The concentration at the department level γ_1 is lower that the concentration γ_0 at the region level.

If $\gamma_i = 1\ \forall i$, so that plants are in theory clustered in one single department within each region,

$$\gamma_1 = \gamma_0 + (1 - \gamma_0)\frac{\sum_i x_i^2 - \sum_i x_i^2 \Omega_i}{1 - \sum_i x_i^2 \Omega_i}.$$

Since Ω_i is close to 1, the last term should be small. The department-level concentration is therefore close to the region-level concentration and the scope of spillovers is very limited.

Replacing γ_0 and γ_i by their 'natural' estimators (see Section 3 and Appendix A) in Eq. (A.12), we obtain an estimator of the geographic concentration in the sequential model that allows for spillovers beyond the limits of the department.

References

Arrow, K., 1962. The economic implications of learning by doing. Review of Economic Studies 29, 155–173.

Ciccone, A., Hall, R., 1996. Productivity and the density of economic activity. American Economic Review 86, 54–70.

Ellison, G., Glaeser, E., 1994, Geographic concentration in U.S. manufacturing industries: a dartboard approach. NBER Working Paper 4840

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

Fujita, M., 1989. Urban Economic Theory, Cambridge University Press, Cambridge.

Fujita, M., Tabuchi, T., 1997. Regional growth in postwar Japan. Regional Science And Urban Economics 27 (6), 643–670.

Fujita, M., Thisse, J.-F., 1996. Economics of Agglomeration. CEPR Discussion Paper 1344

Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A., 1992. Growth in cities. Journal of Political Economy 100, 1126–1152.

Henderson, J., Kunroco, A., Turner, M., 1995. Industrial development in cities. Journal of Political Economy 103, 1067–1090.

Hoover, E., 1936. Location Theory and the Shoe and Leather Industries, Harvard University Press, Cambridge, MA.

Jacobs, J., 1969. The Economics of Cities, Vintage, New York.

Krugman, P., 1991. Geography and Trade, Leuven University Press–MIT Press, Cambridge, MA.

Marshall, A., 1890. Principles of Economics, Macmillan, London.

Maurel, F., 1996. Evolutions locales de l'industrie 1982–1992 et convergence régionale: quelques résultats empiriques sur données françaises. Document de travail de la Direction de la prévision No. 96-6.

Romer, P., 1984. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.