PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TEL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

WORKING PAPER N° 2005 - 42

Concentration, agglomeration and the size of plants

Miren Lafourcade

Giordano Mion

JEL Codes : C21, L11, R12, R30, R34

Keywords : Concentration, spatial auto-correlation, plant size.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES – ÉCOLE NORMALE SUPÉRIEURE

Concentration, Agglomeration and the Size of Plants*

Miren Lafourcade† Giordano Mion‡

November 28, 2005

Abstract

This paper investigates whether the geographic distribution of manufacturing activities depends on the size of plants. Using Italian data we find, as in Kim (1995) and Holmes and Stevens (2002, 2004), that large plants are more concentrated than small plants. However, considering distance-based patterns via spatial auto-correlation, we find that small establishments actually exhibit a greater tendency to be located in adjacent areas. These apparently contradictory findings raise a measurement issue regarding co-location externalities, and suggest that large plants are more likely to cluster within narrow geographical units (concentration), while small establishments would rather co-locate within wider distance-based clusters (agglomeration). This picture is consistent with different size plants engaging in different transport-intensive activities.

JEL classification: C21, L11, R12, R30, R34.
Keywords: Concentration, Spatial Auto-Correlation, Plant Size.

*We would like to thank Giuseppe De Arcangelis, Jérome Carreau, Bart Jourquin, Renato Santelia and Stefano Usai for their critical help in collecting the data, and Antonio Ciccone, Andrew Clark, Esther Lamouroux, Thierry Mayer, Henry Overman, Frédéric Robert-Nicoud, Fabiano Schivardi, Eric Strobl, Stefano Usai and three anonymous referees for insightful comments. Mion gratefully acknowledges financial support from the CEPR Research Network on "The Economic Geography of Europe: Measurement, Testing and Policy Simulations", funded by the European Commission under the Research Training Network Programme (Contract No: HPRN-CT-2000-00069). The usual disclaimer applies.

†Université d'Évry (IUT - GLT Department) and Paris-Jourdan Sciences Economiques (UMR 8545 CNRS-EHESS-ENPC-ENS). Address: PSE, 48 boulevard Jourdan, 75014 Paris, France. lafourcade@pse.ens.fr. http://www.enpc.fr/ceras/lafourcade/.

‡Corresponding author. CORE-UCL and FNRS. Address: CORE, Voie du Roman Pays 34, B-1348 LLN, Belgium. mion@core.ucl.ac.be. http://www.core.ucl.ac.be/~mion.

1 Introduction

Economists, geographers and historians share a considerable interest in analyzing the causes of regional specialization. Among the myriad of determinants which have been explored, particular attention has been paid to regional endowments or raw-material intensity, comparative advantage, localization externalities, or, more recently, transport costs and market potential. In this paper, we focus on one particular aspect of this complex set of mechanisms which has remained relatively unexplored: the size of plants. Our main contribution is to investigate whether the geographic distribution of manufacturing activities is related to plant size, and in particular to look for a separate role for physical distance in explaining location.

We address this question using a number of years of Italian census data on manufacturing industries, at varying geographic and industrial scales. By extending the empirical focus to Europe, we complement the seminal empirical studies of Kim (1995) and Holmes and Stevens (2002, 2004), which both focused on North America. Kim (1995) reports a positive correlation between concentration and both average plant size and the intensity of raw materials, across U.S. manufacturing industries. Holmes and Stevens (2002) find strong evidence of the same phenomenon within industries: plants located in areas with higher industry concentration are larger, on average, than those located outside such areas, and this holds especially for the manufacturing sector. Holmes and Stevens (2002) extend Kim's findings by emphasizing that this positive relationship is robust to controlling for the establishment's own size effect on concentration.

We extend the analysis of the relation between plant size and spatial location patterns to include spatial dependence among geographic units. A number of recent papers (for example, Arbia, 2001a, Duranton and Overman, 2005, and Marcon and Puech, 2003) have underlined significant differences in the concentration patterns obtained from continuous distance-based measures compared to more standard indicators like the index proposed by Ellison and Glaeser (1997). Since labor productivity is positively related to employment density (Ciccone and Hall, 1996, and Ciccone, 2002), a clearer picture of how establishment density varies within industries and of how this depends on the number of employees, remains high on the regional policy agenda.

Although our approach builds on this recent revival of distance, we continue to work with a discrete vision of space as a finite number of spatial units (which can be aggregated to various degrees). We can therefore complement the Ellison and Glaeser concentration index, which measures the strength of co-location spillovers *within* each geographic unit, with an indicator of spatial autocorrelation (the Moran index) that accounts for possible distance-based co-location patterns *across* geographical units. Spatial auto-correlation exists if, for a particular industry, data on the location of plants in region i is 'linearly' informative about the location of other plants within the same industry in regions 'close' to i.[1] Spatial auto-correlation as a feature of location decisions has received relatively little attention compared to concentration, and can be related to the minimization of transport costs, as in the New Economic Geography (henceforth NEG) literature. In this paper we use the term "agglomeration", which is the term commonly used in the NEG to refer to distance-based location patterns, as a synonym for positive spatial auto-correlation.

Using Italian manufacturing data, we find strong evidence of a significant positive relationship between plant size and concentration, as in Kim (1995) and Holmes and Stevens (2002). We go further and examine the sensitivity of this result to distance, and find that, on average, small plants exhibit more pronounced agglomeration patterns. Large establishments are therefore more concentrated but less agglomerated (in the sense of spatial auto-correlation), while small establishments are by contrast more agglomerated and less concentrated. These apparently contradictory results raise a measurement issue linked to the spatial magnitude of co-location spillovers by plant size. This issue is likely related to the properties of the indices used, as discussed below.

Our results suggest that large plants cluster within narrow geographical areas such as local labor markets. By contrast, small plants would rather co-locate within wider areas in which a distance-based pattern emerges. One interpretation is that large manufacturing plants are more export-intensive and thus more oriented towards international markets, as shown for instance by Eaton, Kortum and Kramarz (2004). As they are less sensitive to domestic distances, they co-locate within few dense clusters of activities where they can benefit from particular features, such as Marshallian labor market externalities. By contrast, since small establishments mainly serve local demand, the need to save on domestic transport costs produces a spatial distribution of small plants which corresponds more closely to that of the Italian population (as shown in our data), which is itself spatially auto-correlated. Some exceptions arise, however, for most of the so-called "Italian Districts" industries, which are highly concentrated, but only weakly agglomerated, despite small plant size.[2]

Finally, we show that these results are robust to different definitions of space, of plant size, of industries, and of distance, and that they are also robust to industry characteristics. Furthermore, we find that concentration (agglomeration) has slightly decreased (increased) over the period 1981-1996, small plants exhibiting more movement.

The paper proceeds as follows. Section 2 presents the analytical framework relating plant scale, industry concentration and industry agglomeration. Section 3 describes the Italian data we use

[1]For a more precise definition of spatial auto-correlation and the difference between this and concentration see Section 2, page 3.

[2]This is consistent with a definition of districts based on two 'implicit' criteria: a high concentration of plants among which there is a consistent share of small establishments (Sforzi, 1990), and production mainly geared towards foreign markets (Bagella, Becchetti and Sacchi, 1998).

to investigate the geographic distribution of manufacturing and its relation to plant size. We also discuss briefly how we deal with some well known spatial issues such as the Modifiable Areal Unit Problem. Section 4 provides the results of the cross-section analysis of Italian Local Labor Systems and 3-digit manufacturing industries in 1996. Section 5 checks the robustness of the results and explores long-run trends. Finally, Section 6 concludes and suggests some new lines of research.

2 Analytical Framework

Various indices can be used to investigate the location patterns of economic activity.[3] A large group of indices used by economists, which we refer to as the "concentration" family, splits space into a certain number of geographic units and looks for relative differences in the number of activities within these units, abstracting from their relative position in space. The second family of measures, generally preferred by geographers, accounts for spatial interdependence between regions. We will call this second set "agglomeration" indices. We illustrate the difference between these two families with an example inspired by Arbia (2001b), in which we consider the distribution of 12 plants over the 9 cells of a 3x3 grid.

Figure 1: Agglomeration or concentration?

3	3	
3	3	

a

	3	
3		3
	3	

b

In Figure 1, the activity is unevenly distributed in two different ways. Case b illustrates *concentration*, a distribution concept which is not affected by the permutation of observations in space. Measures such as the Hoover-Balassa location quotient, the Gini coefficient or the Ellison and Glaeser (1997) index, indicate spatial concentration as they provide *a quantification of how much spatial variability a phenomenon exhibits with respect to some average*. As they treat data without considering their relative position in space (*i.e.* independent of distances between spatial units), their value is actually the same in both cases a and b in Figure 1. Case a reflects *agglomeration*[4] because it shows

[3]Holmes and Stevens (2004), Combes and Overman (2004), and Fujita, Henderson and Mori (2004) contain exhaustive presentations of the indices used in a large number of empirical studies in North America, Europe and Asia respectively.

[4]The exact terminology used by Arbia (2001b) is actually "polarization". However, we prefer "agglomeration"

a certain degree of spatial interdependence, in the left corner distance-based cluster. More precisely, there is spatial auto-correlation if, for each industry, the location of plants in region i is 'linearly' predictive of the location of other plants in the same industry in 'close' regions.

It is not always so simple to distinguish between concentration and agglomeration. As can be seen in Figure 1, spatial auto-correlation at a given geographic scale translates, to an extent, into concentration at a more aggregated level. This is particularly true when spatial auto-correlation is modeled as a process involving only contiguous locations. In a more complex space (like the real one), with a distance diffusion process (as we consider here), agglomeration is identified because it implies a distance decay pattern which is not obviously related to concentration. This issue of disentangling concentration from agglomeration for different partitions of space is related to the Modifiable Areal Unit Problem which we will discuss in Section 3.1 below.

The patterns of regional specialization produced in these two families of indices rarely concur, unfortunately. As both types of indices have their pros and cons, we look at the relationship between plant size and regional specialization in a framework combining both families.

2.1 Measuring concentration

Among the most popular measures of concentration are the location quotient (also known as Hoover's coefficient of localization) and the Gini coefficient. However, such indices are not always appropriate tools for the analysis of plant size and concentration: if, by chance, one area includes a very large plant, a positive correlation would emerge randomly, without revealing any real link between concentration and plant size. It is possible to adapt such indices to correct for random causality, as in Holmes and Stevens (2002) for the location quotient. A more theory-grounded framework has been proposed by Ellison and Glaeser (1997) to purge own plant size from concentration. Our concentration measure builds on their model (henceforth EG), as discussed now.[5]

Let M (S) denote the number of spatial geographic units (sectors), $s_i^s = emp_i^s / \sum_{i=1}^{M} emp_i^s$ area i's employment share in the manufacturing industry s, and $x_i = \sum_{s=1}^{S} emp_i^s / \sum_{i=1}^{M} \sum_{s=1}^{S} emp_i^s$ area i's share in total employment. We henceforth omit the industry superscript s, for simplicity.

The EG approach starts from the employment-based index $G_{EG} = \sum_{i=1}^{M} (s_i - x_i)^2$, and then controls for differences in industrial structure using the Herfindahl index of concentration. This index results from a rigorous probabilistic model of plant location. Let N denote the number of

as this term is now widely used in the field of New Economic Geography (NEG) to actually reflect the location process arising from the interaction between transport costs (and so distance) and increasing returns to scale. For a comprehensive review of NEG theory, see Fujita, Krugman and Venables (1999).

[5] A number of previous empirical studies have adopted Ellison and Glaeser's (1997) framework to study the geographic distribution of activities. See for instance Maurel and Sédillot (1999) for France, and Devereux, Griffith and Simpson (2004) for the UK. Our description of the EG model here builds on the simplified version proposed by Maurel and Sédillot (1999).

plants and $z_1, ..., z_j, ...z_N$, the shares of these plants in total industry employment. The fraction of sectoral employment in area i is therefore

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

where $u_{ji} = 1$ if business unit j locates in area i, and 0 otherwise. The u_{ji} are non-independent Bernouilli variables such that the probability for plant j to locate in area i is $P(u_{ji} = 1) = x_i$, which means that a random process of plant location choices will, on average, lead to a pattern of employment shares which matches the aggregate pattern (x_i), which is assumed to be exogenous as is plant size (z_j). More precisely, Ellison and Glaeser model the interaction between the locations of any pair of plants j and k belonging to the same industry as

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

where $\gamma \in [-1, 1]$ describes the strength of spillovers within the industry. The probability that business units j and k locate in the same area i does not depend on j and k:

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

The probability P that the two plants co-locate in *any* of the M locations is therefore a linear function of γ:

$$P = \sum_{i=1}^{M} P(i, i) = \gamma \left(1 - \sum_{i=1}^{M} x_i^2\right) + \sum_{i=1}^{M} x_i^2, \tag{4}$$

Data on plant location can be used to model P, and thus estimate γ.

One of the most appealing ways to interpret this model, as suggested by the authors is to think of plants as darts thrown on a dartboard. Imagine a two-stage process in which nature first chooses to weld some of the darts into clusters (representing groups of plants that are sufficiently interdependent that they will always locate together), and then each cluster is randomly thrown at the dartboard to choose a location. The importance of spillovers is then captured by the parameter γ, which can be viewed as the “fraction” of plants among which co-location occurs.

Ellison and Glaeser propose the following unbiased estimator of γ:

$$\hat{\gamma}_{EG} = \frac{\frac{G_{EG}}{1 - \sum_{i=1}^{M} x_i^2} - H}{1 - H}, \tag{5}$$

where $H = \sum_{j=1}^{N} z_j^2$ is a Herfindahl index which controls for industry differences both in the number and the size of plants.

2.2 Concentration and the size of plants

The EG index allows us to compare geographic concentration across industries because it is immune to the biases associated with differences in establishment structure. However, the EG location model neglects any possible correlation between plant size and concentration: within each industry, the probability $P(i,i)$ that two plants co-locate is independent of their size. In the same spirit as Holmes and Stevens (2002), we propose a simple test of non-random correlation between concentration and plant size: if establishment size does not depend on concentration, then all the variability in the geographic distribution of manufacturing should reflect differences in the number of plants.

By comparing the EG employment-based concentration measure to its plant-based counterpart, whose properties are presented in Maurel and Sédillot (1999), we can identify differences which result from plant size. The employment- and plant-based estimators are equal under the null of a random link between concentration and size, but different if plant size plays a role. The plant-based estimator proposed by Maurel and Sédillot (1999), henceforth labeled "Un-Weighted" (UW) because it treats all observations identically, is

$$\hat{\gamma}_{UW} = \frac{\frac{G_{UW}}{1-\sum\limits_{i=1}^{M} x_i^2} - \bar{H}}{1-\bar{H}}, \tag{6}$$

where $G_{UW} = \sum_{i=1}^{M} \left(\frac{n_i}{N}\right)^2 - \sum_{i=1}^{M} x_i^2$, and $\frac{n_i}{N}$ is the share of plants located in i. The Herfindahl index $\bar{H} = 1/N$, which accounts for differences in the number of plants, is the counterpart of that used in the employment-based EG index.

Maurel and Sédillot (1999) show that this plant-based concentration index is also an unbiased estimator of the spillover parameter γ. Furthermore, it is easy to show that $\hat{\gamma}_{UW}$ is statistically more efficient than its employment-based counterpart $\hat{\gamma}_{EG}$.[6]

A significant difference between $\hat{\gamma}_{UW}$ and $\hat{\gamma}_{EG}$ suggests a non-random relationship between plant size and location choice. If plant concentration does depend on establishment size, then the EG index reflects concentration in large plants, while its UW counterpart is more illustrative of small plants (of which there are a far greater number). We now turn to the issue of spatial auto-correlation and its link to plant size.

[6] As in the standard linear regression framework, both weighted and unweighted estimators are unbiased but, absent heteroscedasticity, the latter should have a smaller variance.

2.3 Measuring agglomeration

The picture we draw from a partition of space into independent cells will likely change if these latter are spatially linked. Accurate indices have been developed to capture the spatial phenomenon of agglomeration. Among such indicators are those proposed by Cliff and Ord (1981), Getis and Ord (1992), and Moran (1950); the latter is the one we actually use in this paper. Let us first define a $M \times M$ spatial weighting matrix W, whose generic element w_{il} is the relative weight of location l for location i and $w_{ii} = 0$. w_{il} may either rely on simple contiguity criteria (for instance, a first-order contiguity matrix will give weight one to all contiguous locations and zero otherwise, including to own location), or be inversely related to the distance d_{il} between i and l (with various specifications such as $d_{il}^{-\tau}$ or $\exp^{-\tau d_{il}}$). Moran's formula is:

$$I = \frac{M \sum\limits_{i=1}^{M} y_i \sum\limits_{l=1}^{M} w_{il} y_l}{S_0 \sum\limits_{i=1}^{M} y_i^2}, \tag{7}$$

where y_i is a measure of economic activity in location i and $S_0 = \sum_{i=1}^{M} \sum_{l=1}^{M} w_{il}$. As proposed by Anselin (1988), the weighting matrix can be row-standardized so that S_0 equals to M (each row is therefore divided by the sum of the row elements).

The most intuitive interpretation of Moran's I is in the context of regression. If we regress the spatially weighted variable Wy on y (where y is the vector of y_i), then I is the slope coefficient of this regression, which is the ratio of $cov(W_i y, y_i)$ to $var(y_i)$, where $W_i = (w_{i1}, ..., w_{il}, ..., w_{iM})$ is the i-related row of the weighting matrix W. Therefore, Moran's I is the correlation coefficient between y_i and that of its neighbors, which enables us to detect departures from spatial randomness and to determine whether neighboring areas are more similar than would be expected under the null hypothesis.

The issues related to the measurement of agglomeration are to an extent similar to those related to concentration. As in the EG model with no spillovers, if plants were distributed randomly in a way that reproduces, on average, the overall distribution of activities, then the largest regions should have more plants. A simple way to control for this location-size effect is to express the variable y relative to its mean, so that $y_i = s_i - x_i$, as in the EG model. Under this null, the mean value of the Moran index is $E[I] = -1/(M-1)$,[7] while under the alternative, it could be either positive or highly negative, depending on the sign of the spatial auto-correlation. We can therefore test for the absence of spatial auto-correlation using the variance of I under the null hypothesis.

[7]See Cliff and Ord (1981) for further details.

2.4 Agglomeration and the size of plants

In order to be consistent with the concentration framework presented in Section 2.2, we investigate the role of plant size by comparing Moran's I statistics calculated for two different (zero mean) variables, measuring employment and the number of plants: $y_i = s_i - x_i$ and $y_i = \left(\frac{n_i}{N}\right) - x_i$. We obtain the two following agglomeration indices:

$$I_W = \frac{(M/S_0)\sum_{i=1}^{M}(s_i - x_i)\sum_{l=1}^{M} w_{il}(s_l - x_l)}{\sum_{i=1}^{M}(s_i - x_i)^2} \quad \text{and} \quad I_{UW} = \frac{(M/S_0)\sum_{i=1}^{M}\left(\frac{n_i}{N} - x_i\right)\sum_{l=1}^{M} w_{il}\left(\frac{n_l}{N} - x_l\right)}{\sum_{i=1}^{M}\left(\frac{n_i}{N} - x_i\right)^2}, \tag{8}$$

For these two Moran indices, we face an issue similar to that in the concentration framework. To perform comparisons between sectors, we have to recognise that extreme agglomeration may occur because, in some industries, there are only a very small number of plants relative to the number of locations. For such industries, strong positive auto-correlation may indicate emptiness surrounded by emptiness. The inclusion of neighboring empty locations may lead to over-estimates of the real number of industrial clusters, so we will check whether the correlation between agglomeration and plant size is not due to industries with very few plants.

A specific issue, to which concentration indices are immune, and which is thus inherent to the Moran index, arises from its regression coefficient nature. As in all regression contexts, the presence of outliers may bias the Moran indicator towards agglomeration that is not representative of the majority of observations. To correct this over-estimation, Anselin (1995) proposes an identification procedure based on conditional randomization. For each location i, Anselin (1995) suggests first computing a *local* Moran statistic, I_i, measuring the correlation between a particular y_i and its neighbors, and then testing its local instability. Further, as the Moran index corresponds to the sample average of the $I_i's$, he also suggests using the sample variance of the $I_i's$ to identify outliers on a two-sigma rule basis. Extreme values under both tests are actually *spatial outliers* or, in the terminology of Anselin (1995), "hot spots". The reason for this double check is the need to test for local instability when the null is not randomization, but some degree of spatial correlation. Under spatial correlation, outliers are more likely to occur than under a random scheme. We will thus pay particular attention to "hot spots", to be sure that the agglomeration indices are representative of the majority of observations, and provide a robust measure of the correlation between agglomeration and plant size.

3 Data and methodological issues

We use data from the Italian Census of economic activities, provided by the Italian National Statistic Institute (ISTAT), which includes information on the location and employment of the whole population of Italian plants. The data is very detailed in its geographic coverage of manufacturing industries. The geographic scale of observation can be disaggregated up to the 8192 Italian communes and the industrial scale up to the 3-digit NACE nomenclature (revision 1) for the years 1981 and 1991, and up to the 5-digit category for 1996.

Contrary to most of the previous relevant empirical literature (Ellison and Glaeser, 1997, Maurel and Sédillot, 1999, and Holmes and Stevens, 2002 and 2004), we have no problem of missing data, the only limitation being that, in some cases, plant size had to be recovered from the size-range groups to which data are allocated. Nonetheless, in roughly 90% of the cases, plant size was directly identified and not estimated. Moreover, information on plant size is not necessary for $\hat{\gamma}_{UW}$ and I_{UW}, which are less demanding than $\hat{\gamma}_{EG}$ and I_W.

3.1 Partitioning space and industries

To calculate both the concentration and agglomeration indices, we have first to choose an adequate scale of industry aggregation and an appropriate geographic unit of analysis.

As Kim (1995) notes, the definition of industry aggregation depends on the subjacent phenomenon in question. Spillover effects and incentives for plants to co-locate can operate either within narrowly-defined industry categories such as the 3-digit nomenclature of activities, or more broadly-defined categories such as the 2-digit nomenclature. In the Italian census, the 3-digit category corresponds to 103 different sub-activities within manufacturing, whereas its 2-digit counterpart divides the latter into only 23 sub-industries. In Italy, the 3-digit category highlights industries which underpin some well-known districts, such as 'Preparation and spinning of textile fibres' (3-digit NACE number 171), 'Textile weaving' (172), 'Tanning and dressing of leather' (191), 'Watches and Clocks' (335), 'Manufacturing of Musical instruments' (363), and 'Ceramic tiles and flags' (263). Although we will emphasise the results obtained with the finer level of industrial disaggregation, to account for the district phenomenon, we will also use the 2-digit disaggregation as a robustness check.

The second issue we have to tackle is the Modifiable Areal Unit Problem (henceforth MAUP), which arises from the partition of space into an arbitrary number of geographic units. The problem, which is well described in Arbia (1989 and 2001b), concerns both the boundaries and the scale chosen. In Figure 2 below, we see that, enlarging the grid of squares in Figure 1 asymmetrically alters the picture of both agglomeration and concentration. Figure 2 leads to exactly the reverse conclusion to that observed in Figure 1: Case c, which is the counterpart of case a, now exhibits pure concentration, whereas case d, which is the counterpart of case b, shows agglomeration.

Figure 2: The MAUP problem

The first precaution we take to minimize the MAUP is to choose a partition of space that relies on real economic features. The partition we adopt is that of Local Labor Systems (henceforth LLS). The LLS spatial nomenclature, which covers both urban and rural areas, divides Italy into 784 geographic units. The average LLS spreads over 384 km^2, which is equivalent to splitting the U.S. continental territory into more than 25,000 units. The LLS grid is not thus dissimilar to the U.S. partition into 41,313 zip-code units. The boundaries of LLS were defined in 1991 by the Italian Statistic Institute on the basis of minimum daily commuting patterns, so as to maximize the coincidence between residential and working areas. The geographic scale of LLS is therefore far less arbitrary than a more standard partition based on simple administrative schemes.

Although the core of the paper will focus on LLS only (Section 4), we will also check that our results hold for other partitions. In section 5, we will check the robustness of the results to the adoption of a more aggregated partition of space. We choose the NUTS3 scale of aggregation (Italian “provincie”), which splits Italy into 95 geographic units. Equally, checking the robustness of the results to a finer partition than LLS - which is already roughly as disaggregated as U.S. zip-codes - also warrants consideration. However, as such a fine-grained partition would bias the regression-based Moran index,[8] we restrict robustness checks to wider geographic units only.

3.2 Partitioning the universe of plants

To explore the role of plant heterogeneity, we must split establishments into at least two groups: large and small. However, defining a clear frontier between the two is far from trivial. As in Holmes and Stevens (2002), pragmatism leads us to adopt the simple strategy of dividing the sample of

[8]A disaggregation into the 8192 municipalities, for instance, requires us to treat 103 x 8192 = 843,716 industry-space observations. As most of the industry-location pairs would attract zero observations, this may bias the Moran index (which was conceived for continuous rather than censored variables) in a way that we cannot correct.

plants according to number of employees. The choice of an employment threshold is arbitrary, but we can appeal to the features of Italian labor markets for help. Two minimum cut-off values arise naturally.

A first threshold of 20 workers makes sense with respect to both the fiscal and legal status of Italian firms. Italian firms with fewer than 20 workers ("piccole imprese") benefit from specific incentives such as tax credits, and lower social contributions and loan interest rates. Further, in order to have an employee board a firm must have at least 20 employees. Finally, the 20-employee cut-off, which was also retained by Holmes and Stevens (2002) for the U.S., allows us to compare our results with theirs, in addition to being the only threshold compatible with the older (1981 and 1991) census data. Table 1 shows summary statistics for the two sub-samples obtained when Italian plants are divided at the 20 worker threshold. Table 1 shows that this cut-off point roughly corresponds to the median plant, which is a further reason to use this division.

Table 1: Italian plants by plant size (20 employee threshold): summary statistics.

	Small plants	Large plants
Mean size	3.73	67.84
St. deviation	11.26	222.18
Coefficient of variation	3.02	3.27
Number of manuf. plants	549,747	41,363
% of manuf. plants	93.01	6.99
% of manuf. employment	42.20	57.80

Nonetheless, a second threshold of 15 workers conforms more closely to Italian dismissal law.[9] When a dismissal is judged to be illegal, a worker has the right to be re-employed by the firm if the latter has more than 15 employees, otherwise the right refers to monetary compensation. This distinction is crucial regarding labor costs because it favors small plants. Furthermore, the 15-employee threshold also appears in Italian legislation with respect to many other labor and fiscal issues, such as working overtime, the hiring of the disabled, training, and tax benefits.

Although the first threshold of 20 workers is preferable for cross-section and intertemporal comparisons, the second seems more reasonable from a legal point of view. We will first appeal to the 20-employee cutpoint (Section 4) and then turn to robustness checks. As information on employment is more detailed in 1996 than in earlier years, we can check in Section 5 whether the 1996 results are robust to the threshold size (15 *vs* 20) and to the particular division adopted (small/large *vs* small/medium/large plants).

[9] See "Statuto dei Lavoratori" art. 18.

3.3 Measures of distance and the weighting matrix

In order to compute the Moran statistic, we need a spatial weighting matrix. Following Harris (1954) and the large literature on gravity estimations recently surveyed by Disdier and Head (2005), we use the inverse of bilateral distance to measure the spatial interdependence of LLS.[10] The only distance measure available at the most disaggregated level of LLS is the great-circle distance.[11] Section 4 thus reports agglomeration indices derived from great-circle distance-based matrices.

However, the geographic scope of agglomeration is likely to depend upon *effective* rather than great-circle distance. For instance, if mountainous terrain impedes access to a location, as might be the case in the Italian Alps, firms may prefer to locate elsewhere. Likewise, the configuration of real transport networks may also affect firms' location choices. In order to guarantee fast delivery and implement "just-in-time" practices, for instance, plants may prefer to locate alongside highways, as illustrated in Arbia (2001a) for the San Marino Republic. Unfortunately, we have no better measure than great-circle distance at the scale of Italian LLS, which is very fine. Real road distances can be computed at the more aggregated geographic scale of the 95 Italian provinces, and will be used in Section 5 as a robustness check. The calculation of bilateral road distances is based on an original GIS provided by Bart Jourquin, that we apply using TRIPS transport modelization software. This allows us to calculate the distance corresponding to the fastest itinerary connecting any pair of Italian provinces through the real road transport network in 1996.[12]

4 Basic Results for Italy: LLS, 3-digit industries, 1996

Section 4.1 assesses the impact of plant size on the geographic distribution of activities by comparing the concentration and agglomeration indices computed on both employment and number of plants bases. As there are significant differences in the results, Section 4.2 considers separate samples of large and small plants. Section 4.3 considers extreme cases of concentration and agglomeration in order to gain a richer understanding of different location patterns.

[10] In Lafourcade and Mion (2003), we also experimented with first-order contiguity matrices. The results obtained being qualitatively similar, we do not report them here.

[11] The great-circle distance is the shortest bilateral distance between the centroids of two geographic units, assuming they would be on a sphere without any physical or network constraint between them. The average great-circle distance between Italian LLS is 467 km.

[12] For more details on the methodology, see Combes and Lafourcade (2005).

4.1 Differences between employment- and plant-based indices of concentration and agglomeration

Table 2 shows that the correlation between the employment- and plant-based measures of both the concentration and agglomeration measures is quite small.

Table 2: Concentration and agglomeration indices (784 LLS): All plants.

	Concentration		Agglomeration	
	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	I_W	I_{UW}
Average value	0.033	0.022	0.010	0.018
Average st. deviation	0.0115	0.0018	0.0025	0.0031
R^2	0.20		0.60	
R^2 ranks	0.54		0.67	
Number of manuf. plants	591,110		591,110	
Number of industries	103		103	
Number of spatial units	784		784	

With respect to concentration, both the weighted and unweighted indices suggest that Italian manufacturing activities are highly concentrated. On the two-sigma rule criterion,[13] $\hat{\gamma}_{EG}$ ($\hat{\gamma}_{UW}$) is significantly different from zero in 91% (97%) of industries.[14] However, the average weighted estimator is 50% larger than its unweighted counterpart and we estimate that around 60% (25%) of industries exhibit a significant positive (negative) differential.[15] Such discrepancies are large and suggest that concentration is more marked for large establishments (which are over-weighted in the employment-based indices) than for small establishments. The difference in average standard errors is also large, to such an extent (up to 15 times larger) that we suspect significant heterogeneity in the sample of plants. Further, correlations between the weighted and unweighted concentration indices are weak, for both values and ranks.

With respect to agglomeration, the two-sigma rule criterion[16] suggests that 66% (86%) of industries exhibit a significant tendency to agglomeration, as measured by I_W (I_{UW}). These results, which are reminiscent of those in Usai and Paci (2002), illustrate how important spatial auto-correlation is in the location decisions of manufacturing plants. Regarding differences between indices, the average

[13] The difference between the index and its expected value under the null of no spillovers (zero) has to be larger than twice its standard error for an industry to be concentrated.

[14] Ellison and Glaeser (1997) define the degree of concentration by classifying industries on a scale referring to both the mean and median $\hat{\gamma}$. They find that 25% of the U.S. manufacturing industries are highly concentrated, while 50% show only slight concentration. The corresponding values in Maurel and Sédillot (1999) and Devereux, Griffith and Simpson (2004) are respectively 27% (for France) and 16% (for the UK) of highly-concentrated industries, and 50% (for France) and 65% (for the UK) of slightly-concentrated industries. Our results show the same concentration ranges for Italy.

[15] The variance of $\hat{\gamma}_{EG}$ and $\hat{\gamma}_{UW}$ is available only under the null of no spillover effect ($\gamma = 0$), so that it is not possible to test properly for differences between two positive values of the estimators. However, assuming normality, we can use the variances to perform a test based on twice the sum of the standard errors.

[16] The difference between the Moran index and its expected value under the null hypothesis (-1/(M-1)) needs to be larger than twice its standard error for an industry to be agglomerated.

unweighted Moran index is 80% larger than its weighted counterpart, this difference being significantly positive (negative) for around 55% (5%) of the industries. Following from the results found for the concentration indices, this provides further evidence of heterogeneity within the sample of plants. Finally, we note that the correlations between the weighted and unweighted agglomeration indices (around 60%), although much larger than their concentration counterparts (for both values and ranks), are not as high as what we would expect if location choices were independent of plant size.

There are significant differences between the employment- and plant-based indices of both concentration and agglomeration. We therefore ask what lies behind this heterogeneity. The next section considers the role of plant size in explaining these differences.

4.2 Large *vs* small plants

Table 3 shows concentration and agglomeration by plant size (with the threshold at 20 employees).

Table 3: Concentration and agglomeration indices (784 LLS): small *vs* large plants

	Small plants				Large plants			
	Concentration		Agglomeration		Concentration		Agglomeration	
	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	I_W	I_{UW}	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	I_W	I_{UW}
Average value	0.024	0.022	0.016	0.018	0.036	0.033	0.007	0.009
Average st. deviation	0.0016	0.0010	0.0029	0.0032	0.0047	0.0020	0.0024	0.0025
R^2	0.90		0.89		0.81		0.70	
R^2 ranks	0.92		0.85		0.73		0.72	
Number of manuf. plants	549,747		549,747		41,363		41,363	

In both cases, the correlation between the weighted and the unweighted indices is far higher, underlining the importance of plant size. This is particularly true for the sub-sample of small establishments, for which the correlation reaches 0.90, for both the concentration and agglomeration indices. The higher correlation in small rather than large plants may result from their relative homogeneity. The ratio of the largest to the smallest in the sample of large plants is around one hundred, which is much higher than 19. We further see in Table 3 that the difference between the unweighted and weighted indices is much lower here than in Table 2. Last, the sharp fall in data variability is further evidence of a relation between the geographic distribution of manufacturing activities and plant size. For instance, while the unweighted concentration estimator has lower variance, consistent with the underlying EG model, the size of the variance gap likely now reflects simple statistical efficiency rather than plant size heteroscedasticity.

Once we recognize that splitting the sample into small and large plants yields consistent results for both indices, it seems reasonable to evaluate concentration and agglomeration patterns by plant size. We see that plant-based concentration indices are about 50% larger for large than for small establishments. More precisely, with a two-sigma rule for $\hat{\gamma}_{UW}$, we find that, in 60% of industries, large plants are significantly more concentrated than are small plants. The reverse is true for only 26% of industries. Holmes and Stevens (2002) find comparable results for the U.S.: the EG index for plants in the fourth quartile (268 employees on average) is twice as large as in the first quartile (25 employees on average). Our results therefore confirm the positive relationship between size and concentration found by Holmes and Stevens (2002) in the U.S. and by Barrios, Bertinelli and Strobl (2003) in Ireland.

On the contrary, small plants are more agglomerated than larger plants, with a mean I value 2-3 times larger for the former. Applying the two-sigma rule for I_{UW}, small plants are significantly more (less) spatially correlated than large plants in 52% (9%) of industries. Small plants are thus more sensitive to distance-based patterns. This result is robust to the exclusion of spatial outliers identified using Anselin's (1995) methodology.

Measures of concentration and agglomeration thus differ by plant size. We explain these apparently contradictory findings when extending our analysis to NUTS3 regions in Section 5.1. In Section 4.3, we instead present selected cross-industry comparisons based on extreme patterns of concentration and agglomeration. This exercise provides concrete industry examples, and yields some insights into the underlying mechanisms behind the spatial distribution of activities. This issue is likely to be key in the design of regional development policies.

4.3 Extreme concentration *vs* extreme agglomeration patterns

Tables 4 and 5, which correspond to detailed 3-digit LLS results available upon request, show the ten manufacturing industries with the highest and lowest indices of concentration and agglomeration for small and large plants respectively. The distinction between unweighted and weighted indices no longer being an issue once plant size is controlled, the following results are based on unweighted indices, because of their efficiency properties.

The first striking feature in Tables 4 and 5 is that extreme concentration and agglomeration patterns correspond to different industries.

The most significantly concentrated industries correspond to activities at the core of Italian districts: 'Preparation and spinning of textile fibres' (3-digit NACE number 171) and 'Textile weaving' (172) located in the 'Prato' LLS, 'Tanning and dressing of leather' (191) in both the LLS of 'Arzignano' and 'Santa Croce', 'Ceramic tiles and flags' (263) around the LLS of 'Sassuolo', 'Manufacturing of jewelery and related articles' (362) in the LLS of 'Alessandria', 'Arezzo' and 'Vicenza',

Table 4: The 10 most concentrated and agglomerated industries (784 LLS): small plants

The 10 most concentrated			The 10 most agglomerated		
NACE	3-digit industry	$\hat{\gamma}_{UW}$	NACE	3-digit industry	I_{UW}
172	Textile weaving	0.247	154	Manuf. of vegetable and animal oils and fats	0.125
171	Preparation and spinning of textile fibres	0.244	153	Process. and preserving of fruit and vegetables	0.104
191	Tanning and dressing of leather	0.204	158	Manuf. of other food products	0.082
296	Manuf. of weapons and ammunition	0.164	232	Manuf. of refined petroleum products	0.075
263	Manuf. of ceramic tiles and flags	0.117	266	Manuf. of articles of concrete, plaster and cement	0.073
192	Manuf. of handbags, saddlery, harness	0.067	281	Manuf. of structural and metal products	0.059
244	Manuf. of pharmaceuticals, med. chemicals, etc.	0.065	193	Manuf. of footwear	0.059
173	Finishing of textiles	0.061	203	Manuf. of builders' carpentry and joinery	0.057
160	Manuf. of tobacco products	0.059	265	Manuf. of cement, lime and plaster	0.051
363	Manuf. of musical instruments	0.049	287	Manuf. other fabricated metal products	0.047

Table 5: The 10 most concentrated and agglomerated industries (784 LLS): large plants

The 10 most concentrated			The 10 most agglomerated		
NACE	3-digit industry	$\hat{\gamma}_{UW}$	NACE	3-digit industry	I_{UW}
223	Reproduction of recorded media	0.393	153	Process. and preserving of fruit and vegetables	0.047
191	Tanning and dressing of leather	0.219	193	Manuf. of footwear	0.039
296	Manuf. of weapons and ammunition	0.216	160	Manuf. of tobacco products	0.038
263	Manuf. of ceramic tiles and flags	0.172	334	Manuf. of optical instruments, photo. equipment	0.034
355	Manuf. of other transport equipment	0.170	363	Manuf. of musical instruments	0.029
363	Manuf. of musical instruments	0.155	152	Process. and preserving of fish products	0.028
362	Manuf. of jewelery and related art	0.129	182	Manuf. of other wearing apparel and accessories	0.027
173	Finishing of textiles	0.089	295	Manuf. of other special purpose machinery	0.026
221	Publishing	0.079	232	Manuf. of refined petroleum products	0.025
244	Manuf. of pharmaceuticals, med. chemicals, etc.	0.078	293	Manuf. of agricultural and forestry machinery	0.022

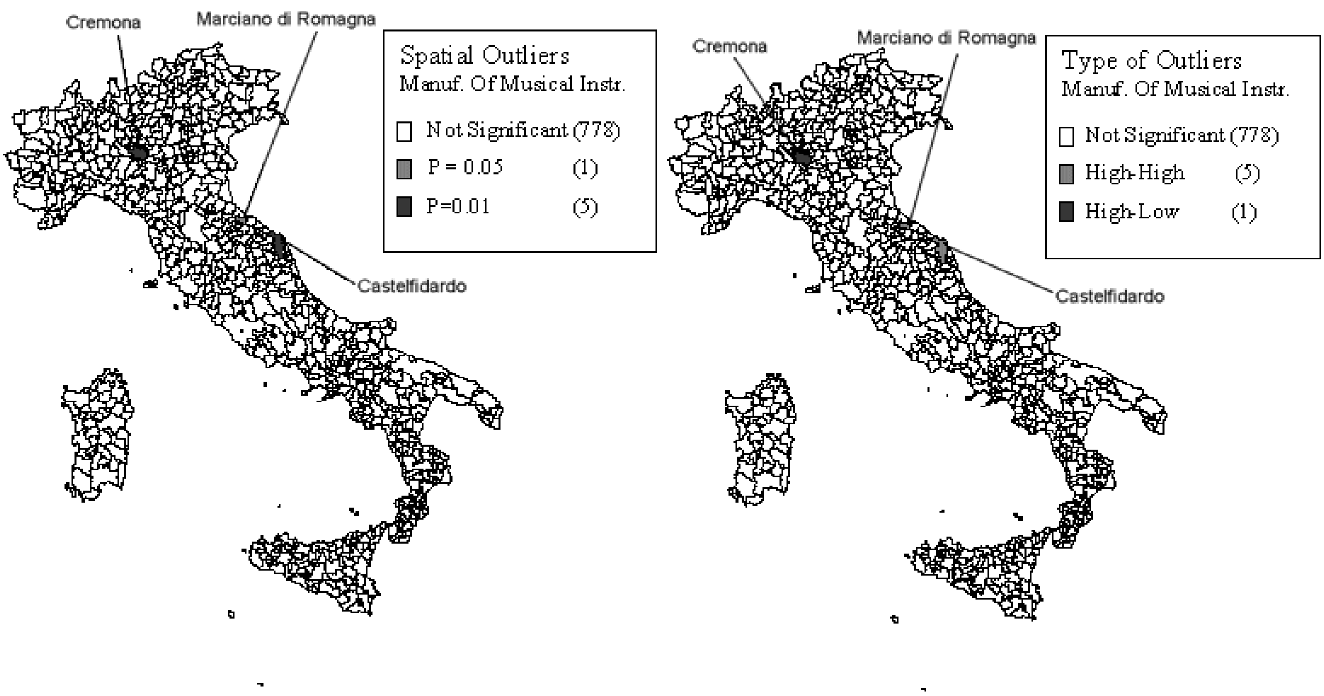
and 'Manufacturing of Musical Instruments' (363) in the LLS of 'Ancona'.[17] The districts related to textile industries (171 and 172) show high levels of concentration which mainly result from small plants (under 20 employees), while concentration in other districts seems to be more or less equally associated with both small and large plants (191, 263 and 363). This is consistent with a primary characterization of Italian districts by their proportion of small plants, as in Sforzi (1990). However, although concentrated, the industries behind the district structure are only weakly or not at all agglomerated, as shown by the insignificance of both the weighted and unweighted Moran indices. Despite particularly small plants, Italian districts are an exception to the rule that small plants locate according to national distance-based patterns. This apparently puzzling feature can be explained by

[17]For more details on the mapping of Italian Districts see Sforzi (1990).

the fact that the production of many Italian districts is mainly oriented towards foreign markets, as emphasized by Bagella, Becchetti and Sacchi (1998).

A clear exception to this trend is found for the 'Manufacturing of Musical Instruments', which is the only district industry that is both concentrated and agglomerated, independently of plant size (industry 363 figures in nearly all of the columns of Tables 4 and 5). However, separation between concentration and agglomeration can be restored by showing that agglomeration here is spurious, being triggered by spatial outliers. Figure 3 shows that the manufacturing of musical instruments exhibits some interesting cases of "hot spots", as identified by the procedure described in Section 2.4.

Figure 3: Spatial Outliers in the 'Manufacturing of Musical Instruments' Industry



The left-hand side of Figure 3 shows the number of spatial outliers in the industry,[18] while the right-hand side depicts the "type" of observed spatial correlation. From top to bottom, three non-white areas stand out as outliers. The first is located in the LLS of 'Cremona', where there is extreme concentration of plants virtually surrounded by nothing (High-Low). By contrast, the southern location of 'Marciano di Romagna', where plants are also very concentrated, is neighbored by LLS that share a small proportion of the same activity (High-High, with most of the activity concentrated in the central LLS). Finally, the well-known musical instruments district centered around 'Castelfidardo', which includes the neighboring labor markets of 'Ancona', 'Macerata', 'Osimo' and

[18]The shades of grey correspond to the significance of the local Moran statistic, and the probabilities to the test of local instability.

‘Recanati’, has plants which are very concentrated (also High-High, but with both the central and neighboring LLS highly concentrated). Excluding spatial outliers in this industry (which still leaves 110 LLS with non-zero employment), the ‘Manufacturing of Musical Instruments’ industry is no longer significantly agglomerated (*i.e.* the Moran index is not significant anymore). Therefore, the extreme agglomeration in the data here is due to only a few outliers (6 LLS out of 116 with non-zero unemployment), and does not reflect any general tendency for the musical instruments industry to be more agglomerated than average.

The most concentrated industries with respect to large plants (See Table 5) include ‘Manufacturing of other transport equipment’ (355). The Moran index for this industry is negative but not significant. Caution is thus needed in interpreting distribution patterns here: in 1996, there were only four Italian plants with at least 20 employees in the transport equipment industry. They were all located in the small Northern triangle of ‘Bergamo’, ‘Modena’, and ‘Imola’ LLS. This number is particularly small compared to the 784 LLS geographic units, meaning that the Moran index is biased (see Section 2.4). Mass-production activities such as ‘Publishing’ (221) or ‘Reproduction of recorded media’ (223), are very concentrated in the category of large plants, but not agglomerated.

These results suggest that extreme concentration patterns occurring in local labor markets, such as those above, reflect dense clusters of economic activity and may therefore generate productivity gains within the related industries and areas. These gains could result from either strong increasing returns to scale (there are few plants, most of which are large) or localized spillovers created through density. These may enable large plants to serve markets located far beyond the boundaries of neighboring markets. Eaton, Kortum and Kramarz (2004) show that larger firms are more productive and are thus able to target a greater number of foreign markets. This is consistent with (i) a geographic distribution of large plants that is highly concentrated but insensitive to domestic distances, and (ii) goods which are designed for the export market.

Food industries are over-represented (152, 153, 154, and 158) among activities with the strongest positive spatial auto-correlation. This is particularly true within the sample of small plants. Other activities related to the final stage of production, such as ‘Manufacturing of footwear’ (193), exhibit the same pattern. Theory suggests that agglomeration patterns are likely to prevail because plants want to save on transport costs (due to either perishability, volume, specific transport types or customers’ face-to-face requirements), and therefore locate in close proximity to potential buyers (consumers or downstream industries). Our data do indeed suggest that the distribution of small plants closely matches the distribution of the Italian population. The average (across 3-digit industries) correlation between the number of small (large) plants and population[19] in each LLS is 0.6106

[19]The data on population by LLS comes from the 1991 Census of Population provided by the Italian National

(0.4394), with population being itself significantly spatially auto-correlated (Moran's $I = 0.0084$).[20] High transport costs due to face-to-face involvement with domestic customers and demand linkages (final and intermediate) likely explain why small plants are more agglomerated than large plants, which are more export-oriented.

The sharpest exceptions to this trend are the Italian districts formed around small concentrated plants. The fact that such districts do not show any distance-based co-location patterns at the LLS scale is compatible with an underlying criterion defining districts by production oriented mainly towards foreign (and not domestic) markets.

5 Robustness checks and long-run trends

This section considers the robustness of the positive (negative) correlation found between concentration (agglomeration) and plant size, and suggests some explanations of these apparently contradictory findings in the light of previous work (Section 5.1). It further explores the time dimension of our panel data, investigating the evolution of concentration and agglomeration over the 1981-1996 period. Last, it shows that changes have been triggered by small rather than large plants and that results are robust the inclusion of industry characteristics (Section 5.2).

5.1 Robustness checks

We first investigate the robustness of the results presented in Section 4 to changes in geographic, industrial, and plant size definitions.

Controlling for the MAUP and distance bias: The case of Italian 'provincie'

Do the trends in Section 4 depend on the geographic partition of Italian space into Local Labor Systems? In other words, is the positive (negative) correlation found between concentration (agglomeration) and plant size robust to the MAUP? We answer by presenting concentration and agglomeration patterns at the scale of 95 Italian NUTS3 regions. The comparison of Table 6 to its LLS counterpart (Table 3), shows that both concentration and agglomeration indices rise with the geographical scale of the analysis.

The result that concentration is larger in wider spatial units is now well-established.[21] This suggests that concentration is likely to depend on other spillovers than those resulting, for instance, from labor market pooling. Spatial auto-correlation also seems to occur at a larger scale than

Statistic Institute (ISTAT).

[20]The average (across 3-digit industries) correlation between employment related to small (large) plants and population observed in each LLS is 0.5412 (0.3831).

[21]See among others Kim (1995), Ellison and Glaeser (1997), Maurel and Sédillot (1999), Pagnini (2003) and Barrios, Bertinelli, Strobl and Teixeira (2005).

Table 6: Concentration and agglomeration indices (95 'provincie'): small *vs* large plants

	Small plants				Large plants			
	Concentration		Agglomeration		Concentration		Agglomeration	
	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	I_W	I_{UW}	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	I_W	I_{UW}
Average value	0.034	0.030	0.024	0.033	0.049	0.046	0.007	0.013
Average st. deviation	0.0025	0.0015	0.0126	0.0133	0.0069	0.0030	0.0124	0.0125
R^2	0.89		0.81		0.82		0.65	
R^2 ranks	0.92		0.70		0.75		0.58	
Number of manuf. plants	549,747		549,747		41,363		41,363	

LLS. Nevertheless, as the difference between Moran indices calculated for LLS and provinces are rarely significant, we cannot draw any general conclusion from this result.[22] The overall pattern between concentration, agglomeration, and plant size is similar to that found within LLS. At the scale of Italian administrative NUTS3 regions, large plants concentrate more in regions where other large plants are already located. With respect to agglomeration, small plants display more spatial auto-correlation than do large plants.

These apparently contradictory results, which are valid for both LLS and NUTS3 regions, raise a measurement issue related to the spatial magnitude of the co-location spillovers of small *vs* large plants. This issue is likely to be inherent to the indices we use. Maurel and Sédillot (1999) have already shown that the EG index does not provide an accurate estimate of the parameter governing the co-location incentives of plants (γ) *at a given geographic scale* whenever there are spillovers working at different geographical levels. More intuitively, the EG index at an aggregate spatial level is a kind of "sum" of the co-location incentives operating at both that level (say γ_{NUTS3}) and at a finer level (say γ_{LLS}). The fact that the EG index increases with geographical aggregation in our data, as in previous work, fits in with this intuition.

In the data, we find that large plants are more concentrated than small plants at all geographic scales. However, this does not necessarily mean that the co-location spillovers operating at different spatial levels are greater for large plants. For instance, concentration at the NUTS3 level may just reflect a high value of γ_{LLS} at the scale of LLS. By contrast, spatial auto-correlation is more closely tied to co-location spillovers operating at a given geographical level as it directly measures the degree to which establishments locate close to each other. The greater spatial correlation of location decisions for small establishments across both LLS and NUTS3 regions suggests that co-location incentives are probably larger for big plants "within" LLS only. Large plants therefore benefit more

[22]The "real" increase in observed spatial auto-correlation when going from LLS to 'provincie' is underestimated by the change in the Moran index. The expected value of the Moran index, under the null of no spatial auto-correlation, is $E[I] = -1/(M-1)$, which equals -0.0013 (-0.0106) for LLS ('provincie'). However, when deflating the Moran index by its mean, the increase in spatial correlation at the 'provincie' level is significant for only a few industries.

from co-locating within narrow geographical units such as Local Labor Systems than do small plants. By contrast, small establishments gain from locating in wider distance-based clusters.

Another interpretation of our results relates to the continuous-space concentration analysis of Duranton and Overman (2005), who develop a methodology for evaluating concentration at any given distance by means of an observed density function of bilateral distances between plants. Contrary to the EG index, concentration is captured by an explicit function of distance. Our opposing findings for small *vs* large plants may simply show that the two corresponding concentration densities cannot be ranked according to stochastic dominance (*i.e.* no one density dominates the other over all distances). Duranton and Overman (2005) note that their index of localization (considering large plants only) is slightly above that in the baseline simulations (for all plants) over short distances. This result is consistent with large plants showing a higher propensity to co-locate within narrow areas. However, they cannot prove the significance of this finding within their framework.

Introducing real road distances

Table 7 shows that introducing real road distances between NUTS3 regions instead of bilateral great-circle distances, does not much change the results.

Table 7: Moran indices computed with real road distances (95 'provincie')

	Small plants		Large plants	
	I_W	I_{UW}	I_W	I_{UW}
Average value	0.032	0.040	0.011	0.016
Average st. deviation	0.0141	0.0148	0.0140	0.0141
R^2	0.79		0.75	
R^2 ranks	0.78		0.70	
Number of manuf. plants	549,747		41,363	

The most striking difference between Tables 6 and 7 is the rise in the mean Moran index, which is comparable to that between LLS and NUTS3 regions with great-circle distances. The differences between employment- and plant-based Moran indices fall slightly for the sample of large plants (the correlations in both the levels and ranks rise). However, the agglomeration indices are still around three times larger for small plants than for large establishments. The similarity of the results based on the road and great-circle distances is not surprising: Combes and Lafourcade (2005) show that, for France, great-circle distance is a good proxy for road distance in *cross-section analysis*. As the distance bias worsens when passing from cross-section to time-series analysis, we will have to be cautious in interpreting the long-run trends in Section 5.2.

Comparison between 2-digit and 3-digit industries

To test whether the results are sensitive to the definition of manufacturing products, we recalculate our indices using 2-digit industry classification.[23] Although the qualitative relationship between concentration, agglomeration and plant size is not affected, there are changes in the level of these indices. At the LLS scale, the unweighted index of concentration (agglomeration) falls (rises) by 58% (36%) at this broadly-defined level of manufacturing industries. The fact that concentration falls with the level of industry aggregation suggests that spillovers in manufacturing are less likely to operate between 3-digit industries than within each industry. This result does not depend upon the definition of regions. This is not surprising, and has already been underlined by Ellison and Glaeser (1997), and Maurel and Sédillot (1999), in the perspective of concentration. Our analysis complements these findings by showing that the distribution of plants within the broad 2-digit category is characterized by stronger distance-based patterns (more agglomeration). However, the variances of the 2-digit indices are so large with respect to the 3-digit classification that these differences are rarely significant.

Plant-size sensitivity

The positive (although reduced) gap between employment- and plant-based indices remaining after correcting for plant size (Table 3) suggests that more complex partition schemes than small-large may be preferable. We therefore split Italian plants into three size categories instead of two: small (under 15 employees), medium (15 to 100) and large (over 100). This has the advantage of respecting the cut-off of 15 employees which, as noted above, may be more relevant for Italy.

Table 8: Concentration and agglomeration in 3 sub-samples of plants (784 LLS)

	Concentration ($\hat{\gamma}_{UW}$)			Agglomeration (I_{UW})		
	Small	Medium	Large	Small	Medium	Large
Average value	0.022	0.032	0.053	0.019	0.013	0.005
Average st. deviation	0.0016	0.0020	0.0200	0.0032	0.0027	0.0024
Number of manuf. plants	534,427	51,298	5,385	534,427	51,298	5,385
Number of manuf. workers	1,783,799	1,600,103	1,471,875	1,783,799	1,600,103	1,471,875

Table 8, the counterpart of Table 3 in Section 4, shows that a finer analysis by plant size leaves the results virtually unchanged: large plants are more concentrated than medium establishments, which are in turn more concentrated than small plants; the reverse trend holds for agglomeration. Our explanation of the correlation between concentration, agglomeration and plant size thus remains unchanged to that in Section 4. The results above suggest however that it may be better to use a continuous framework: The larger (smaller) the employment scale of plants, the greater the tendency to concentrate (agglomerate).

[23]The results for 2-digit industries (for both LLS and NUTS3 regions) are available upon request.

5.2 Changes in concentration and agglomeration over time

Table 9 presents the result of the following industry fixed-effects panel regression: for the three census years 1981, 1991 and 1996, the concentration and agglomeration indices are regressed on size=ln(average size of establishments) in each industry, with u^s being the industry fixed-effect.

Table 9: Panel regression of indices on Size=ln(average size of plants) (784 LLS)

	Coefficients or Test values	
	Concentration ($\hat{\gamma}_{UW}$)	Agglomeration (I_{UW})
Size	0.0248* (0.0039)	-0.0055* (0.0012)
Constant	-0.0437* (0.0106)	0.0327* (0.0033)
R^2	0.17	0.14
Number of observations	309	309
Number of years	3	3
Industry dummies	Yes	Yes
Time dummies	Yes	Yes

Note: Standard errors in brackets. * denotes significance at the 1% level.

The effect of size on concentration (agglomeration) is positive (negative) and significant at the 1% level, as expected. Plant scale has an impact on the spatial distribution of establishments which is not simply driven by (time-invariant) sector specific characteristics such as factor endowments or raw materials.

With respect to changes in concentration and agglomeration over time, Table 10 (in the Appendix) shows that mean concentration slightly fell, while mean agglomeration rose over the period 1981-1996. Kim (1995) finds a similar declining pattern of concentration after the Second World War in the U.S. More recent studies for Europe, such as Brülhart (2001) or Midelfart, Overman, Redding and Venables (2002) lead to more mixed results, with some industries exhibiting increases and others falls in concentration.[24] Disentangling changes over time by plant size (See Tables 11 and 12 in the Appendix), the dynamics mainly result from small plants, with larger plants being more stable.

6 Conclusions

This paper has analyzed the spatial distribution of manufacturing in Italy with respect to two different features. The first, concentration, can be defined as the degree of variability across data for a given partition of space. The second, agglomeration, explicitly considers distances between observations and thus their spatial dependence. Although much work has focused on concentration, agglomeration has received far less attention.

[24] Both articles relate changes in concentration to industry characteristics such as increasing return to scale, skills and R&D intensity.

Examining the influence of plant size on both concentration and agglomeration is useful in a number of ways. While the fact that large plants are more concentrated than smaller plants has already been emphasised, the finding that small plants are more agglomerated than large plants is innovative. Large plants therefore benefit from co-locating within fine geographic units such as labor markets; small establishments, by way of contrast, rather gain from locating in wider distance-based clusters. These results are robust to different partitions of space, industries, plants, and distance. In addition, controlling for time-invariant sector specific characteristics, such as raw materials or natural resources, does not change the results. Our findings highlight some of the underlying economic mechanisms behind plant location choice. Differences in the intensity of transport costs between small and large plants is a plausible explanation of the different location choices. A key determinant of the distance-based patterns found for industries with small plants, such as food processing, is the need to save on transport costs by locating close to domestic demand. By way of contrast, industries with large plants, or which underpin Italian districts, exhibit a higher proportion of international activities. Last, changes over time in the geographic distribution of Italian establishments reveal declining (increasing) concentration (agglomeration), the dynamics being mainly driven by small plants.

With respect to policy, our findings suggest that regional policies may have different impacts on large and small plants, which we show behave differently with respect to location. For instance, Baldwin and Okubo (2005) develop a model in which they combine plants of heterogeneous size and productivity with geography. They show that a policy which increases the share of industry in peripheral regions will drive the largest and most productive firms to move to the core. This result, which they call 'sorting of firms', may explain why modest production subsidies have very little impact on regional welfare, as they only attract a few small and inefficient firms.

A number of further lines of research deserve attention in future work. A first valuable contribution would be to measure both the narrow and large scope spillovers within the same integrated theoretical framework, instead of combining the two. This would enrich the analysis by enabling different types of externalities to be identified. It would also be useful to examine the causality of the relationship between plant size, concentration and agglomeration. As circular causation may be at work, structural econometrics is called for to distinguish between rival theoretical models. Finally, our results underline the importance of models dealing explicitly with both plants' characteristics and location. The emerging literature on heterogeneous firms and international trade seems to be the most promising framework to investigate this issue.

References

Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anselin, L., 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27(2), 93-115.

Arbia, G., 1989. *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Kluwer, Dordrecht.

Arbia, G., 2001a. Modelling the geography of economic activities on a continuous space. *Papers in Regional Science*, 80, 411-424.

Arbia, G., 2001b. The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems*, 3, 271-281.

Bagella, M., Becchetti, L. and Sacchi, S., 1998. The positive link between geographical agglomeration and export intensity: The engine of Italian endogenous growth?. *Economic Notes*, Monte dei Paschi di Siena, 27(1), 1-34.

Baldwin, R.E., and Okubo, T., 2005. Heterogeneous firms, agglomeration and economic geography: Spatial selection and sorting. Forthcoming, *Journal of Economic Geography*.

Barrios, S., Bertinelli, L., and Strobl, E., 2003. Geographic Concentration and Establishment Scale: Can Panel Data tell us more?. *CORE Discussion Papers*, 36.

Barrios, S., Bertinelli, L., Strobl, E., and Teixeira, A.C., 2005. The Dynamics of Agglomeration: Evidence From Ireland and Portugal. Forthcoming, *Journal of Urban Economics*.

Brülhart M., 2001. "Evolving Geographical Specialisation of European Manufacturing Industrie". *Weltwirtschaftliches Archiv*, 137(2), 215-243.

Ciccone, A., 2002. Agglomeration effects in Europe. *European Economic Review*, 46, 213-227.

Ciccone, A., and Hall, R.E., 1996. Productivity and the density of economic activities. *American Economic Review*, 86(1), 54-70.

Cliff, A.D., and Ord, J. K., 1981. *Spatial Processes: Models and Applications*. *Pion*, London.

Combes, P-P., and Lafourcade, M., 2005. Transport Costs: Measures, determinants and regional policy implications for France. *Journal of Economic Geography*, in press.

Combes, P.-P., and Overman, H.G., 2004. The Spatial Distribution of Economic Activities in the European Union. In *Handbook of Regional and Urban Economics*, Vol.4, J.V. Henderson and J.-F. Thisse (eds.), Elsevier-North Holland, Amsterdam.

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

Disdier, A. C., and Head, K., 2004. The Puzzling Persistence of the Distance Effect on Bilateral Trade. Mimeo. http://pacific.commerce.ubc.ca/keith/Papers/meta.pdf.

Duranton, G., and Overman, H.G., 2005. Testing for Localisation Using Micro-Geographic Data. Forthcoming, *Review of Economic Studies*.

Eaton, J., Kortum, S., and Kramarz, F., 2004. An Anatomy of International Trade: Evidence from French Firms. Mimeo, New York University.

Ellison, G., and Glaeser, E., 1997. Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105, 889-927.

Fujita, M., Krugman, P., and Venables, A.J., 1999. *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge, Massachusetts, 367p.

Fujita, M., Henderson, J.V., and Mori, T., 2004. Spatial Distribution of Economic Activities in Japan and China. In *Handbook of Regional and Urban Economics*, Vol.4, J.V. Henderson and J.-F. Thisse (eds.), Elsevier-North Holland, Amsterdam.

Getis, A., and Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189-206.

Harris, C.D., 1954. The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, 44, 315-348.

Holmes, T.J., and Stevens, J.J., 2004. Spatial Distribution of Economic Activities in North America. In *Handbook of Regional and Urban Economics*, Vol.4, J.V. Henderson and J.-F. Thisse (eds.), Elsevier-North Holland, Amsterdam.

Holmes, T.J., and Stevens, J.J., 2002. Geographic concentration and establishment scale. *Review of Economics and Statistics*, 84, 682-690.

Kim, S., 1995. Expansion of Markets and the Geographic Concentration of Economic activities: The Trends in U.S. Regional Manufacturing structure, 1860-1987. *Quarterly Journal of Economics*, 110(4), 881-908.

Lafourcade, M., and Mion, G., 2003. Concentration, Agglomeration and the Size of plants: Disentangling the Source of Co-Location Externalities. *Core Discussion Paper* 91, UCL.

Marcon, E., and Puech, F., 2003. Evaluating the Geographic Concentration of Industries Using Distance-Based Methods. *Journal of Economic Geography*, 3(4): 409-428.

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

Midelfart, K.H., Overman, H.G., Redding, S.J., and Venables, A.J., 2002. Integration and industrial specialisation in the European Union. *Revue Économique*, 53, 469-481.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena, *Biometrika*, 37, 17-23.

Pagnini, M., 2003. Misura e Determinanti dell'Agglomerazione Spaziale Nei Comparti Industriali in Italia. *Mimeo*, Bank of Italy.

Sforzi, F., 1990. The quantitative importance of Marshallian industrial districts in the Italian economy. In *Industrial districts and inter-firm co-operation in Italy*, F. Pyke, G. Becattini and W. Sengerberger (eds.), 75-107, Geneva, International Institute for Labor Studies, ILO.

Usai, S., and Paci, R., 2002. Externalities and local economic growth in manufacturing industries. In *European Regional Growth*, Fingleton B. (ed.), Springer-Verlag.

Appendix: additional tables

Table 10: Concentration and agglomeration over time: all plants

	Concentration ($\hat{\gamma}_{UW}$)			Agglomeration (I_{UW})		
	1981	1991	1996	1981	1991	1996
Average value	0.026	0.023	0.022	0.014	0.016	0.018
Average st. deviation	0.0004	0.0011	0.0018	0.0027	0.0029	0.0031
Number of industries		103			103	
Number of spatial units		784			784	
Number of manuf. plants	622,353	592,753	591,110	622,353	592,753	591,110

Table 11: Concentration and agglomeration over time: small plants

	Concentration ($\hat{\gamma}_{UW}$)			Agglomeration (I_{UW})		
	1981	1991	1996	1981	1991	1996
Average value	0.026	0.022	0.022	0.017	0.018	0.019
Average st. deviation	0.0009	0.0010	0.0010	0.0030	0.0032	0.0033
Number of industries		103			103	
Number of spatial units		784			784	
Number of manuf. plants	579,676	550,103	549,747	579,676	550,103	549,747

Table 12: Concentration and agglomeration over time: large plants

	Concentration ($\hat{\gamma}_{UW}$)			Agglomeration (I_{UW})		
	1981	1991	1996	1981	1991	1996
Average value	0.032	0.032	0.033	0.010	0.009	0.010
Average st. deviation	0.0016	0.0023	0.0020	0.0025	0.0025	0.0025
Number of industries		103			103	
Number of spatial units		784			784	
Number of manuf. plants	42,677	42,650	41,363	42,677	42,650	41,363