Concentration, Spatial Clustering and the Size of Plants: Disentangling the Sources of Co-location externalities*

Miren Lafourcade† Giordano Mion‡

May 6, 2004

Abstract

Following Ellison and Glaeser (1997), we develop a framework to test for the link between concentration, spatial clustering and the size of plants. Concentration is an a-spatial concept of variability that can be measured with the Ellison and Glaeser (EG) index. By contrast, spatial clustering being directly concerned with distances, a two-dimensional measure such as the Moran index allows to identify specific agglomeration patterns. We argue that, if the size of plants were independent on both concentration and agglomeration, as in the standard monopolistic competition framework, all the variability in those indexes should be accounted for by the variation in the number of plants. Using the Italian 1996 census year data, we compare the values and significance of both the EG and Moran indexes computed on an employment and a number of plants basis. Our results indicate that, for the majority of Italian manufacturing industries, big plants have larger concentration incentives than small ones, even once controlled for size effects, and that concentration and size simultaneously influence each other. By contrast, small plants exhibit stronger positive auto-correlation, whose scope might extend to large spatial scales, size and clustering being also endogenous. Different externalities might thus drive concentration and agglomeration patterns according to a size-related basis, which casts some doubt on the relevance of standard monopolistic frameworks to empirically account for large scope "pecuniary" externalities instead of spatially bounded spillovers.

JEL classification: C21, L11, R12, R30, R34.
Keywords: Concentration, Spatial Auto-Correlation, Size of Plants, Pecuniary Externalities.

*We are very grateful to Giuseppe De Arcangelis and Renato Santelia for kindly providing their help in gathering data, and to Stefano Usai and Eric Strobl for fruitful discussions and comments.

†University of Evry (IUT - GLT Department) and CERAS-ENPC. CERAS, 48 Bd. Jourdan, 75014 Paris, France. lafourca@enpc.fr. http://www.enpc.fr/ceras/lafourcade/.

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

1 Introduction

The concentration of production, which is one of the most striking feature of the geography of economic activities, is probably also the most direct evidence of the pervasive need for firms to draw benefits from the presence of externalities. Disentangling the sources of co-location decisions is a challenge that a long tradition of economists has tried to take up. To this purpose, the intensity of regional specialization in particular industries, and, conversely, the level of industrial concentration in particular locations, have been used as complementary evidences for the existence and the significance of externalities. As explanatory variables to concentration or specialization patterns, determinants such as factor endowments, increasing returns to scale, or spatially bounded technological externalities arising from the close proximity of agents (Marshallian labor market pooling, urbanization or localization economies) have been largely explored.[1] By contrast, studies measuring benefits coming from non spatially bounded "pecuniary" externalities are quite scarce and this paper dedicates to the issue of disentangling the source of co-location decisions between pecuniary and localized externalities according to their respective spatial scope.

New Economic Geography models, with those "Market potential" reduced specifications that can be derived from them, offer a seducing theoretical background to account for spillovers whose geographical scope overlaps the borders of local labor markets, and more particularly for "pecuniary" externalities arising from demand and input-output linkages. The standard Dixit - Stiglitz (1977) - Krugman (1980) monopolistic competition model (DSK)[2] has been the most successful framework in explaining how large scope externalities might drive concentration and specialization patterns, in the presence of increasing returns to scale and transport costs. Therefore, most empirical studies structurally address the quantification of these effects, as the measure of their geographical scope of propagation, within this tractable framework (Hanson, 1998, Forslid, Haaland and Midelfart-Knarvik, 2002, Redding and Venables, 2004, Mion, fothcoming). However, a few recent models (Combes and Lafourcade, 2001, Duranton and Puga, 2001, Melitz and Ottaviano, 2004, Holmes and Stevens, fothcoming) actually depart from the main stream in proposing an alternative framework designed to overcome one of the major drawbacks of the standard DSK structure: its difficulty to account for differences in the size of plants as a determinant of industrial concentration or regional specialization.

The framework we develop here can help to shed a new light on the question of which models are

[1]See, among the studies tackling this issue Jaffe, Trajtenberg and Henderson (1993), or Henderson, Kunkoro and Turner (1995), Combes (2000), Brülhart (2001), Holmes and Stevens (2002).

[2]See Fujita, Krugman and Venables (1999) for a detailed presentation of the main variants of the DSK theoretical framework.

the most relevant according to empirical issues. We disentangle the concentration and specialization processes according to three different aspects indeed: 'internal' (the size of plants), 'localized' (natural advantages, factor endowments, technological spillovers) and 'large spatial scope' (non spatially bounded pecuniary externalities). Following the probabilistic-based approach of Ellison and Glaeser (1997), we study the link between concentration, spatial clustering and the size of plants, paying particular attention to the causal direction of this relation. Concentration is referred to as an a-spatial concept of variability that can be measured with the standard locational Gini or the more sophisticated Ellison and Glaeser (EG) index. By contrast, spatial clustering or "agglomeration" is directly concerned with distances. Therefore we also use a two-dimensional measure (the Moran index) to identify some specific distance-based patterns. In this respect, our approach turns towards a trend of studies designed to develop new measures accounting for the spatial content of concentration patterns, as Duranton and Overman (2002), Paci and Usai (2003), Marcon and Puech (2003) or Burchfield, Overman, Puga and Turner (2003).

We argue that, in a world where the size of establishments would be independent on both concentration and spatial agglomeration, as the standard DSK framework, all the variability in these measures should be accounted for by the variation in the number of plants. Using the Italian 1996 census year data on manufacturing industries, we therefore compare the values and significance of both the EG and Moran indexes computed on an employment and number of plants basis.

A first result arising from the comparison of the standard EG index to its plant alternative measure is the existence of huge discrepancies, the correlation being poor with respect to both values and ranks. However, comparisons based on two distinct samples (one for big and the other for small plants) and inspired by the work of Holmes and Stevens (2002), indicate that, once controlled for the size of plants, the EG model is consistent with data, with the two indicators of concentration giving similar results. In particular, we show that big plants exhibit much stronger co-location incentives (50% on average) than small ones, suggesting that they are more sensitive to 'localized' externalities. Moreover, as a further explorative exercise, we identify those industries whose concentration is mainly driven by big *vs* small units. Interestingly, many historical Italian Districts belong to the second group,[3] while the first one includes mass-production activities as industries belonging to the final stage of production.

As regards the relation between size and agglomeration, our main finding is that small firms display a stronger spatial pattern than big ones. This is, as far as we know, the first study that attempts to deal with this issue. Contrary to the case of concentration, small units thus seem to be more

[3] Thus confirming the relevance of the definition of these districts that is based on two 'implicit' criteria: a consistent share of small establishments and exports towards foreign markets.

sensitive to final markets accessibility and input-output linkages (pecuniary externalities). Finally, we characterize those industries whose agglomeration is fostered by big *vs* small units. Upstream and final product industries display more spatial correlation, with the latter being disproportionately more represented in the sample of small establishments.

The paper is therefore structured as follows. Section 2 computes standard employment-based indexes of concentration and specialization and compares them to their plant-based alternatives. Discrepancies between both types of measures being large, Section 3 then turns to the more sophisticated plant-based location model of Ellison and Glaeser (1997) and focuses on its robustness to the size of plants. Section 4 is designed to measure the geographical scope of spillovers, by constructing a spatial Moran index that is consistent with the Ellison and Glaeser model of location choices. Moreover, this index allows to disentangle local sources of concentration from a broader non spatially bounded agglomeration process, according to the size pattern of firms. Finally, Section 5 concludes and proposes directions for further research.

2 Measuring Specialization and Concentration

This section dedicates to the study of standard measures of concentration and specialization. It more particularly focuses on the comparison of employment-based and plant-based indexes, in order to reveal some specific patterns that could stem from differences in the size of establishments.

2.1 Employment-based Indexes

One of the most-often used measure of specialization is the employment Location Quotient (LQ), also known as the Hoover-Balassa coefficient. With respect to this measure, a particular location is defined as specialized in an industry (for instance manufacturing) if that location share of employment in the industry exceeds its national share. Complementary to the previous index, one can also define an employment "Industry Quotient" (IQ) along the idea that an industry should be concentrated in a particular location if its share of employment in the location exceeds the corresponding national share.

Let M (S) denote the number of spatial units (sectors), $s_i^s = emp_i^s / \sum_{i=1}^{M} emp_i^s$ location i share of employment in the manufacturing sector s, and $x_i = \sum_{s=1}^{S} emp_i^s / \sum_{i=1}^{M} \sum_{s=1}^{S} emp_i^s$ its share of total employment. The simplest way to measure how much location i is specialized in industry s is therefore to compute

$$LQ_i^s = \frac{s_i^s}{x_i}. \tag{1}$$

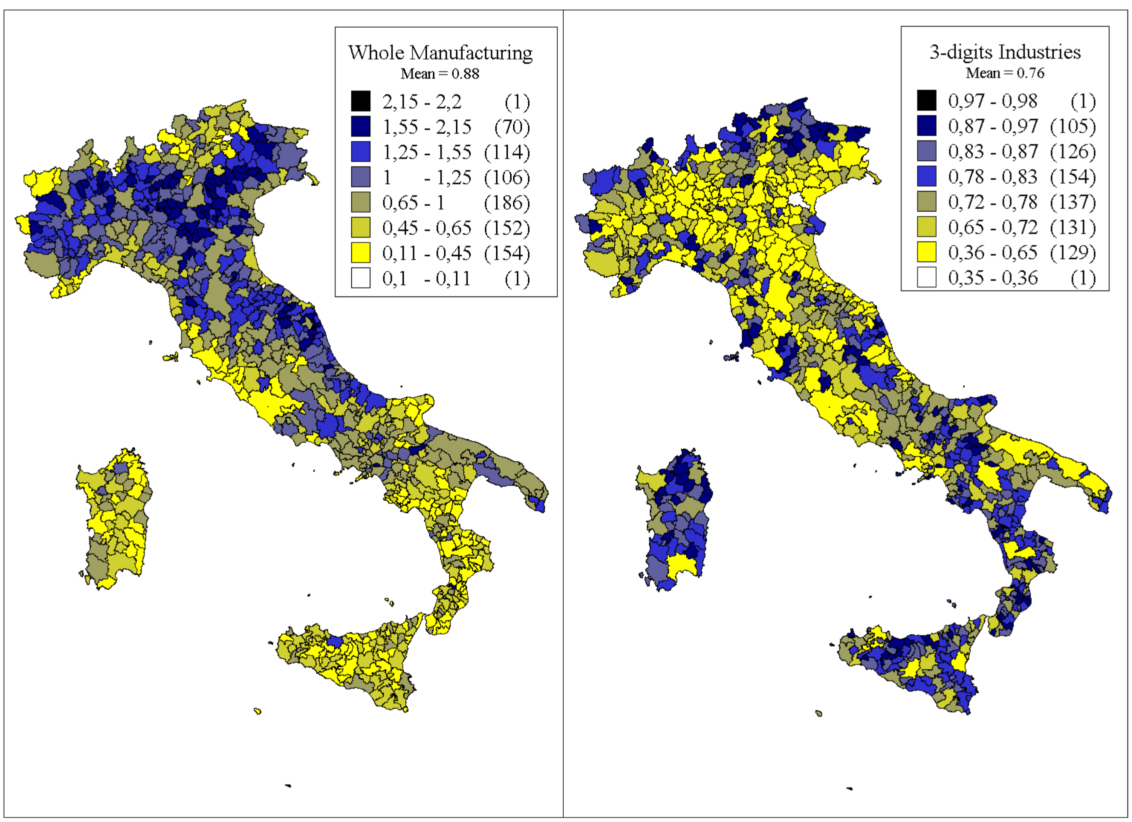
Reciprocally, let $c_i^s = emp_i^s / \sum_{s=1}^{S} emp_i^s$ denote the industry s share of employment in location i, and $x^s = \sum_{i=1}^{M} emp_i^s / \sum_{i=1}^{M} \sum_{s=1}^{S} emp_i^s$ its share of total employment. It is straightforward to see that LQ, as defined in equation (1), also measures the industry s concentration in location i, IQ. We have indeed

$$LQ_i^s = \frac{s_i^s}{x_i} \quad \equiv \quad IQ_i^s = \frac{c_i^s}{x^s}. \tag{2}$$

Values above 1 mean that the location (industry) is relatively specialized (concentrated) in the industry (location), as it has relatively more employment than it would be predicted based on its aggregate employment share.

Let illustrate this definition with a broad picture of the concentration and specialization patterns obtained for the manufacturing industry in Italy.[4] Figure 1 left depicts the distribution of employment LQs across the 784 Italian Local Labor Systems (LLS), whose boundaries were defined in 1991 by the Italian Statistic Institute on the basis of minimum daily commuting patterns.[5] The

Figure 1: Employment-based LQ (left) - Locational Gini (right): 1996



[4]The reader will find in Holmes and Stevens (2004), Combes and Overman (2004) and Fujita, Henderson and Mori (2004), very complete surveys of studies related to this issue for a large set of North-American, European and Asian countries.

[5]Extreme values of LQs can be traced back thanks to the white and black extreme classes of the legend.

mean concentration index, at $0.88 < 1$, reveals that manufacturing as a whole is less concentrated than total employment in Italy (yellow LLS dominate over blue ones indeed). However, the spatial variability exhibited for the concentration of manufacturing is strong, with LQs ranging from 0.11 (in the North-Center winter sports resort of 'Canazei', in white) to 2.20 (in the Middle-East leather product district of 'Monte San Pietrangeli', in black). This maximum value of 2.20 means that the share of employment in the manufacturing industry is 120 percent greater in this area than in Italy as a whole. Among the four Italian LLS presenting the highest absolute shares in manufacturing employment in 1996, 'Milano', 'Torino', 'Roma' and 'Napoli' (with relative LQs of respectively 0.88, 1.12, 0.40 and 0.74), only Torino exhibits some low manufacturing concentration.

To make inter-sectoral comparisons easier, it is convenient to synthesize this spatial distribution within a single measure of concentration. The 'Industrial' Gini coefficient is the LQ-based index that is most often used to measure the overall degree of an industry concentration within a set of different locations. More precisely, it measures the degree to which the percentage distribution of an industry employment across locations corresponds to the percentage distribution of the national employment across those same locations.[6] It is then derived from ordering locations by increasing values of employment LQs and cumulating the s_i^s and x_i shares over the ordered locations. Let denote $S_i^s = \sum_{l=1}^{i} s_l^s$ the cumulative share of sector s employment, and $X_i = \sum_{l=1}^{i} x_l$ the cumulative share of aggregate employment, in the ordered locations i. The 'Industrial' Gini is

$$Gini^s = 1 - \sum_{i=1}^{M} (X_i - X_{i-1})(S_i^s + S_{i-1}^s). \qquad (3)$$

It takes values in the range $[0, 1]$. A value of 0 means that all locations share the same employment proportion of the industry s, while, on the contrary, a value of 1 denotes extreme inequality, the whole industry employment being concentrated in a single location. The value obtained for the Italian manufacturing industry as a whole in 1996, at 0.225, is large compared to other sectors[7] and confirms the first glance impression of strong inequalities given by Figure 1-left.

Regarding the 'specialization' complementary perspective, we also compute a 'Locational Gini' coefficient by ordering employment IQs according to their increasing values and cumulating the corresponding shares c_i^s and x^s ranked over a sample of manufacturing sub-industries. Whereas the industrial gini uses the variability in regional concentration for a given industry, this specialization index uses, by contrast, the variability in the industrial structure of locations. This index therefore

[6]Indeed, it is based on the Lorenz cumulative frequency curve, that compares the distribution of employment in a specific location with the uniform distribution that would represent equality to the national average. This equality distribution is represented by a diagonal line, and the greater the deviation of the Lorenz curve from this line, the greater the inequality in the specialization patterns of employment.

[7]The corresponding Gini for non public services in 1996 is 0.138.

summarizes the degree of specialization of LLS with respect to diverse manufacturing sub-industries, a value of 1 meaning that the area is specialized in a single sub-activity. Figure 1-right exhibits the spatial distribution of Locational Gini's computed for the NACE revision 1 three-digits classification of manufacturing industries.

It is straightforward to see that the pattern of specialization is inversely related to the level of concentration, as depicted in Figure 1-left. The areas exhibiting the lowest concentration of manufacturing employment, as for instance northern border, west-southern and insular LLS, clearly reveal a strong specialization of these workers in only a few manufacturing activities. By contrast, Italian big urban LLS, such as 'Milano', 'Torino' or 'Roma', are more diversified across different sub-manufacturing activities. Standing as an exception, areas belonging to historical Italian Districts (like the LLS of 'Monte San Pietrangeli'), in which the concentration of manufacturing activities is very strong, are simultaneously strongly specialized, their manufacturing activity focusing mostly on a single sub-industry, such as footwear.

The relative specialization of an area in few industries as the degree to which an industry concentrates in few areas can be driven by two possible sources: differences in the number of manufacturing plants and differences in the size of these plants. Section 2.2 focuses more particularly on the issue of disentangling these sources.

2.2 Plant-based Indexes

The Gini and the Hoover-Balassa indexes measure specialization and concentration taking workers as the unit of analysis. By contrast, the literature on externalities and location is concerned with the interaction of plants while labor measures, like the thickness of the market in Marshall (1890) labor pooling idea, are considered as proxies for the externalities arising from the proximity of plants. Therefore, if we aim at measuring the strength of these spillover effects, we have to reconsider our unit of analysis in favor of plants.

In order to gain some insight, it is useful to think about the distribution of workers over space as the mixture of two distributions:

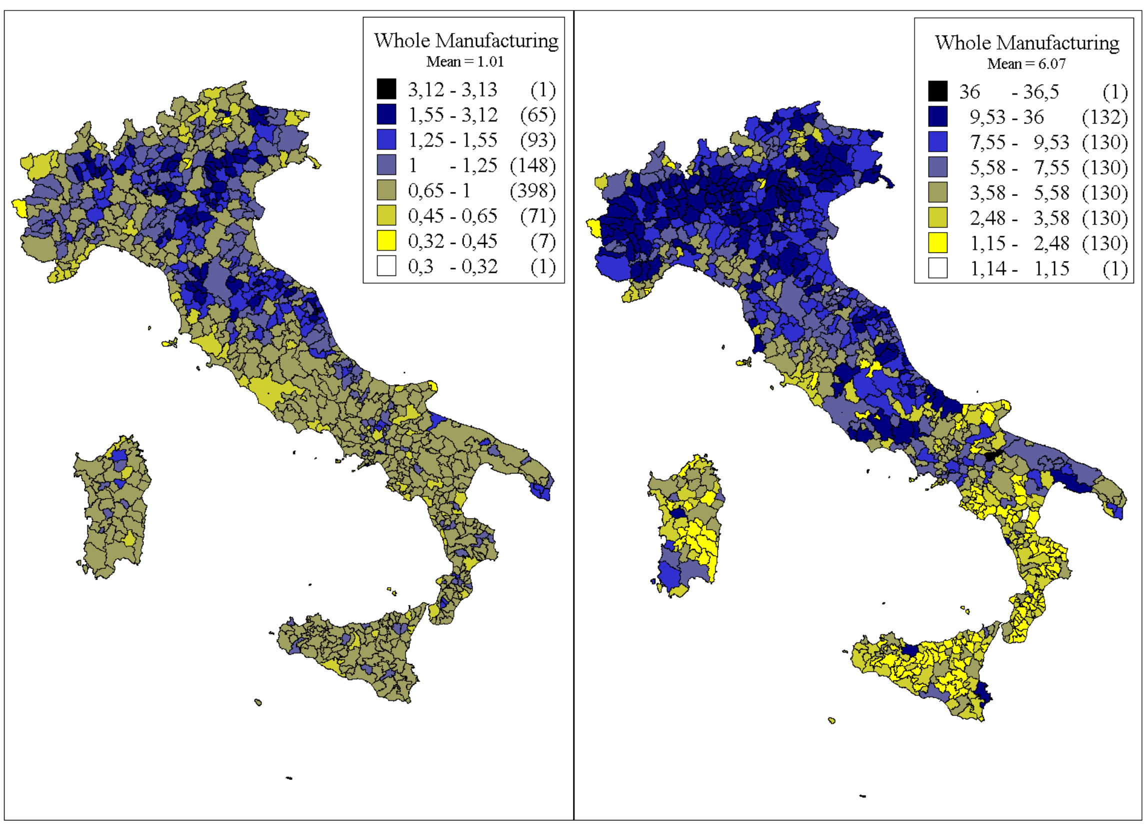
- The plant size distribution (i.e. the allocation of workers among plants).
- The plant location distribution (i.e. the outcome of the location choices of plants).

The concentration of workers may thus come from both sources, but spatial externalities matter only in the second distribution. Consider, for instance, the two following polar cases. In the first one, there is only one big plant in industry s, and it is located in region i. In the second case, there are many plants belonging to industry s, and they are all located in i. Clearly, the employment-based

Gini indexes would take the same high value in both cases, indicating a strong concentration pattern in industry s. However, the nature of this concentration is completely different in both cases. In the first one, concentration occurs at the establishment scale, resulting in a unique operating plant that hires all workers (*industrial concentration*). This could be reasonably associated to factors that are internal to a plant, like increasing returns to scale in production, but, no matter where this plant locates, we will observe a high value of Gini. The second case is, by contrast, characterized by a co-location of different plants in the same place (*spatial concentration*), and suggests that there are some "external" elements like natural resources, knowledge spillovers or local demand and input-output linkages, that drive this process.

A first simple way to disentangle the effects of *spatial concentration* and *industrial concentration* is thus to compute the LQ on the basis of the number of plants, to compare it with its employment-based alternative, and to examine whether differences in the spatial distribution of employment and plants are driven by the heterogeneity of plant sizes. Figure 2 depicts the spatial distribution of Italian manufacturing plants in 1996 (left) and relates it to the average size distribution of these plants (right).

Figure 2: Number of Plants-based LQ (left), Average size of plants (right)



Generally speaking, three striking features stem from the comparison of Figures 1 and 2.

- Patterns of concentration are strongest for employment than for plants, due to the strong uneven spatial distribution of the size of establishments in the manufacturing industry. Although the left-map of Figure 2 exhibits LQ now ranging from 0.30 (in the insular LLS of 'Porto Azzurro') to 3.13 (in the LLS of 'Montegranaro', which is contiguous to 'Monte San Pietrangeli', in the same leather district), the plant-based Gini, 0.187, is lower than for employment. This casts some doubt on the quantitative significance of a measure of concentration that would not correct for industrial concentration.

- The average size of manufacturing plants and their industrial concentration across the 784 Italian Labor Systems according to the number of plant-based LQ are strongly correlated (0.38).[8] This suggests that manufacturing plants located in areas where other manufacturing plants also concentrate would be larger, on average, than outside such locations. This feature is also reported by Kim (1995) and by Holmes and Stevens (2002), for the US Census Regions and counties. The correlation that may therefore arise between employment-based measures of concentration and the average size of plants may be partly due to randomness, however, and one has to purge from the endogenous concentration stemming from the location choice of big plants.

- Both plants and employment indexes display strong spatial auto-correlation patterns, suggesting that the choices of plants to locate inside an area might depend upon the number and the size of other plants *inside* this area as in the *neighboring* locations. Spillovers in the manufacturing industry therefore overlap the borders of Local Labor Systems and their propensity to spread might also depend upon the size of plants.

Section 3 is intended to explore both the issues of purging concentration measures from industrial concentration and studying the nature of its relation with the average plant size by means of the EG location model. The issues of spatial auto-correlation and of its causal relation with the size of plants are left for Section 4.

3 The Ellison and Glaeser model and its robustness to plant size effects

Inconsistencies arising from the comparison of raw employment-based and plant-based measures of concentration are huge and put forward the need to correct for sectoral differences in the size

[8]However, the correlation obtained for other 3-digit activities such as services is much more lower, around 0.1.

of establishments. In order to deal with the issue of spatial concentration, Ellison and Glaeser (1997) propose a more structured framework consisting in starting from the employment-based index $G_{EG} = \sum_{i=1}^{M}(s_i - x_i)^2$,[9] and then neutralizing the contribution of industrial concentration by means of an appropriate measure. Importantly, and this is one of their greatest contribution, they derive this measure from an explicit and rigorous probabilistic model of plant location decisions. In particular, Ellison and Glaeser (1997) spell out two sources of spatial concentration: natural advantages and spillover effects among neighbor plants producing the same kind of goods. However, the two parameters corresponding to these forces are observationally indistinguishable, and only a linear combination of them can be estimated.[10] Consequently, from now on, we will refer to this linear combination as the spillover parameter γ. Despite the impossibility to directly disentangle resources-driven from externalities-driven concentration, this model approach has several nice features.

3.1 Basic features of the EG model

Let us recall the basic properties fulfilled by the EG model and the associated index of concentration.[11] Let N denote the number of plants and $z_1, ..., z_j, ...z_N$, the shares of these plants in the total employment of an industry. The fraction of sectoral employment related to location i is therefore

$$s_i = \sum_{j=1}^{N} z_j \, u_{ji}, \tag{4}$$

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

$$Corr(u_{ji}, u_{ki}) = \gamma \quad for\ j \neq k, \tag{5}$$

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

$$p(i,i) = E[u_{ji}\, u_{ki}] = Cov(u_{ji}, u_{ki}) + E[u_{ji}]\, E[u_{ki}] = \gamma x_i(1 - x_i) + x_i^2, \tag{6}$$

[9] For notation convenience, we omit the sector superscript s.

[10] Regarding the issue of disentangling the contribution of natural resources from other determinants of concentration, see Ellison and Glaeser (1999).

[11] See Ellison and Glaeser (1997), and Maurel and Sedillot (1999) for further details.

and the probability P that the same pair of plants locates in *any* of the M locations is

$$P = \sum_{i=1}^{M} p(i,i) = \gamma \left(1 - \sum_{i=1}^{M} x_i^2\right) + \sum_{i=1}^{M} x_i^2, \tag{7}$$

that is a (linear) function of γ. From an empirical point of view, this simple relation means that using actual data on plants' location as an estimator of P, one can trace back the parameter γ.

Coming back to intuition, one of the most appealing way to interpret this model is, as suggested by Ellison and Glaeser themselves, to think about plants as darts to be thrown in space. 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 thrown randomly 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 must occur.

Ellison and Glaeser (1997), propose the following statistic as an estimator of γ:

$$\hat{\gamma}_{EG} = \frac{\frac{G_{EG}}{1-\sum_{i=1}^{M} x_i^2} - H}{1 - H}, \tag{8}$$

where $H = \sum_{j=1}^{N} z_j^2$ is the Herfindahl index of the industrial plant size distribution, and G_{EG} is defined as above. The authors show that $\hat{\gamma}_{EG}$ has useful properties:

- It is an unbiased estimator of the spillover parameter γ.
- As the variance of $\hat{\gamma}_{EG}$ can be computed under the null of independent location choices ($\gamma = 0$), one can make inference about the significance of estimates.
- This estimator gives a "pure" measure of spatial concentration that is comparable across industries in which the size distribution of plants differs. Indeed, as shown in expression (8), starting from a raw employment-based index of concentration (G_{EG}), it then controls for the overall distribution of activities ($1 - \sum_{i=1}^{M} x_i^2$), as well as for the degree of industrial specific concentration (H).
- The expected $\hat{\gamma}_{EG}$ does not depend on the way data are partitioned in space. So that, no matter what level of geographical disaggregation is chosen, if the scope of externalities does not change over space the expectation of the estimator will not change as well.

In section 3.2, we re-examine the underlying hypothesis of the EG model, focusing more particularly on the neutrality of plants size. In section 3.3, we then try to shed some light on the relation between size and concentration, and characterize it for different sectors.

3.2 Weighted *vs* Un-weighted estimators

One simple way to test for the robustness of the EG model, is to compare different estimators. As far as these estimators are equivalent under the null that the model is a good approximation of reality, but still different under the alternative, one can use them to indirectly test the EG model. Furthermore, if differences under the alternative can be related to the size of plants, one can also question the robustness of the EG model to size effects.

Based on the same location model, Maurel and Sedillot (1999) actually propose two different estimators for γ. The first one, that is very close to $\hat{\gamma}_{EG}$, is given by

$$\hat{\gamma}_{MS} = \frac{\frac{G_{MS}}{1-\sum_{i=1}^{M} x_i^2} - H}{1 - H}, \tag{9}$$

where the only difference is given by $G_{MS} = \sum_{i=1}^{M}(s_i^2 - x_i^2) \neq G_{EG} = \sum_{i=1}^{M}(s_i - x_i)^2$.

The other one, to which we will refer as the un-weighted estimator for a reason that will become evident, is instead given by

$$\hat{\gamma}_{UW} = \frac{\frac{G_{UW}}{1-\sum_{i=1}^{M} x_i^2} - \bar{H}}{1 - \bar{H}}, \tag{10}$$

where $G_{UW} = \sum_{i=1}^{M} \left(\frac{n_i}{N}\right)^2 - \sum_{i=1}^{M} x_i^2$ and $\frac{n_i}{N}$, the share of plants located in i, while $\bar{H} = 1/N$ is the corresponding Herfindahl index of plants concentration.

Maurel and Sedillot (1999) prove that, under the assumptions of the location model, these two estimators should have the same properties as $\hat{\gamma}_{EG}$ and, in particular, they should give an un-biased estimate of the spillover parameter γ_{EG}. However, there is a fundamental difference between $\hat{\gamma}_{UW}$ and the others. Indeed, the authors show that both $\hat{\gamma}_{EG}$ and $\hat{\gamma}_{MS}$ can be viewed as weighted frequency estimators of the co-location event, with weights being given by the size of co-locating plants.[12] This means that they both give a disproportional importance to the co-location of big plants, while small establishments have a smaller impact on the value of estimators. By contrast, $\hat{\gamma}_{UW}$ has an un-weighted nature, because it treats all observations in the same manner.[13]

[12]Formally, when a co-location between two establishments j and k occurs ($u_{ji} = u_{ki} = 1$), this information receives a weight that is proportional to the product $z_j * z_k$.

[13]Another way to interpret the difference between $\hat{\gamma}_{UW}$ and the other two estimators is in terms of the data they focus on. Both $\hat{\gamma}_{EG}$ and $\hat{\gamma}_{MS}$ start from a raw measure of geographical concentration based on *employment* data (respectively G_{EG} and G_{MS}), and then neutralize the industrial concentration component due to an unequal allocation of workers to plants by means of the Herfindahl index. On the other hand, $\hat{\gamma}_{UW}$ uses directly *plants* data (G_{UW}), and the distribution of workers is used only as a measure of a location attractiveness (x_i). As the model is concerned with establishment interactions over space, the second approach seems more natural.

Turning back to the model, the use of a weighted estimator could be justified if the information about co-location of big business units would be more reliable or, in other words, if there is some degree of heteroscedasticity linked to the establishment size. However, the EG underlying location model excludes such an heteroscedasticity because the co-location probability $p(i,i)$ is independent from the size of j and k. Therefore, $\hat{\gamma}_{UW}$ should, if the model is correct, be preferred to the others because it is more efficient.[14] Moreover, an eventual discrepancy in the results stemming from the data comparison of the un-weighted and weihgted estimators could also suggest that there is a more fundamental problem: the endogeneity of the plant size distribution with respect to their location choice. If we relax the assumption that the tendency to concentrate does not depend on the establishment scale, then the estimates given by a weighted and a un-weighted estimator should depart significantly from each other. In particular, the weighted one will give values that are representative of big plants spillovers while the other (the number of small plants being much higher), should be closer to the γ of small units.

Comparing the weighted and un-weighted estimators

Table 1 contains the results of a comparison between $\hat{\gamma}_{EG}$ and $\hat{\gamma}_{UW}$, as computed from the Italian Census of Economic Activities for the year 1996, disaggregated at both spatial (LLS and provinces) and industry (NACE, 2 and 3-digits) levels. This data source, on which a complete description can be found in Appendix A, gives detailed geographic (up to the 8192 Italian commons) and Industrial (up to 3-digit NACE revision 1 for 1981 and 1991, and 5-digits for 1996) information on location and employment of the *universe* of Italian plants. Contrary to the US Census for instance, there is therefore no problem of withheld data in our sample, the only limitation being that the size of plants has to be recovered from the size-range groups to which data are allocated. However, given the level of data disaggregation, in roughly 90% of the cases, the size of a plant can be directly identified.[15]

Table 1: Comparison of the two estimators ($\hat{\gamma}_{EG}$ and $\hat{\gamma}_{UW}$) on Italian 1996 data.

	3-digits				2-digits			
	LLS		Provinces		LLS		Provinces	
	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$
Average value	0.033	0.022	0.048	0.030	0.016	0.009	0.021	0.014
Average st. deviation	0.0115	0.0018	0.0177	0.0027	0.0010	0.0001	0.0015	0.0001
R^2	0.20		0.10		0.09		0.17	
R^2 ranks	0.54		0.42		0.51		0.58	
Number of industries	103		103		23		23	
Number of spatial units	784		95		784		95	
Number of manuf. plants	591,110		591,110		591,110		591,110	

[14] Like in the standard linear regressions framework, both a weighted and an un-weighted estimator are un-biased but, if there is no heteroscedasticity, the latter should have a smaller variance.

[15] See Appendix A for additional information on the procedure used to recover the Herfindahl index.

It is straightforward to notice that the correlation between both estimators is quite weak. The average means across sectors are significantly different, with the weighted estimator systematically giving values from 50% to 100% higher than the un-weighted one. Although a proper inference cannot be performed,[16] these discrepancies are very large, suggesting that the scope of spillover effects for big business units is significantly higher than for small ones. The difference of average standard errors is also important, with a magnitude so strong (up to 15 times) that it is unlikely to be the result of an unreliable co-location information for small plants. This is instead suggestive of the presence of a more deep source of un-accounted heterogeneity, and is consistent with a problem of sample heterogeneity due to a structural difference in the γ of small and big units. Finally, the correlation between the two estimators, both for values and ranks, is very weak, indicating that these two indicators measure forces that are not equally represented in all sectors. Small and big plants have indeed an heterogenous weight in different industries, with small plants playing in some cases (the Italian Industrial Districts for instance) a leading role.

Identifying the most concentrated industries

Most standard results stem from the comparison of $\widehat{\gamma}$ with respect to both the geographical and sectoral disaggregations. Independently on the level of industrial disaggregation considered, $\widehat{\gamma}$ increases with the geographical scope.[17] This suggests that the concentration process overlaps the borders of Local Labor Systems in Italy, meaning that other forces than localized spillovers shape the distribution of manufacturing activities. However, we leave the issue of measuring the geographical scope of spillovers and disentangling localized externalities from other agglomeration sources along with the size of plants for Section 4.

In the same spirit as Maurel and Sedillot (1999), we identify those industries for which there is a substantial departure from the randomness hypothesis ($\gamma = 0$), by using the variance of the estimator and performing a test based on a two sigma rule criterion.[18] Coherently with the French manufacturing analysis of Maurel and Sedillot (1999), we find that at least 90% of industries[19] appear to be significantly concentrated in Italy, this % increasing when passing from the 3- to the 2-digits classification.

As for comparability with other studies, we also consider the values of $\widehat{\gamma}$ as having their own

[16]Unfortunately, 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 properly test the differences between two positive values of the estimators. However, knowing the variances of both estimators allows to compute confidence intervals based on twice the standard deviations, which leads to significant discrepancies for 80% to 95% of industries depending on the level of disaggregation.

[17]The result that concentration rises when considering bigger spatial units is now well established. See Ellison and Glaeser (1997), Maurel and Sedillot (1999), Barrios, Bertinelli, Strobl and Teixeira (2003), and Pagnini (2003) among others.

[18]The difference between $\gamma_{\hat{U}W}$ and its expected value under the null (0) has to be larger than twice its standard deviation to accept the sector to have a significant degree of concentration.

[19]For 2-digits sectoral results, see Appendix B. Detailed 3-digits results are available upon request.

meaning. For example, Ellison and Glaeser (1997) regard an industry as very concentrated if its $\widehat{\gamma}_{EG}$ exceeds 0.05 (the mean value of $\hat{\gamma}_{EG}$ in their sample of 459 manufacturing industries), and as low concentrated if it goes below 0.02 (their median $\hat{\gamma}_{EG}$): 25% of industries are therefore found to be very concentrated against 50% as low concentrated. The Maurel and Sedillot (1999) and Devereux, Griffith and Simpson (2003) corresponding values are respectively, 27% (for France) and 16% (for the UK) of very concentrated industries ($\hat{\gamma}_{EG} > 0.05$), against respectively 50% (for France) and 65% (for the UK) of low concentrated ones ($\hat{\gamma}_{EG} < 0.02$). Within our sample of 103 3-digits manufacturing industries, we find, at the geographical disaggregation of provinces which is the most comparable with former studies, a median $\hat{\gamma}_{EG}$ of 0.019 while the mean $\hat{\gamma}_{EG}$ is 0.047. These values are in line with those obtained in previous findings and, using the same cut-off values as Ellison and Glaeser (1997), we find that 21% of manufacturing industries are very concentrated, against 52% as low concentrated. Moreover, we find larger $\hat{\gamma}$ at the finer 3-digits sectoral level, suggesting that spillovers are stronger between business units of the same sub-activity.[20]

Among the most significantly concentrated industries according to the un-weighted estimator, one can find 'Textile weaving' (172), 'Preparation and spinning of textile fibres' (171) and 'Tanning and dressing of leather' (191), with $\widehat{\gamma}_{UW}$ ($\widehat{\gamma}_{EG}$) of respectively, 0.217 (0.050), 0.215 (0.062) and 0.206 (0.208). Huge discrepancies between un-weighted and weighted estimators are therefore likely to occur at the 3-digits level of industrial classification, where differences in the size of firms are even more pronounced than in manufacturing considered as a whole. Next section is thus devoted to the issue of disentangling the role of big and small plants in driving the concentration of manufacturing activities.

3.3 Concentration and the size of plants

The idea that the size of a plant is not exogenous to its location decision has been put forward in other investigations. From a theoretical point of view, the nursery city location model developed by Duranton and Puga (2001b) is probably one of the most convincing explanation of this issue. The authors show that, in equilibrium, new products are developed in diversified cities, trying processes borrowed from different activities. On finding their ideal process, firms then switch to mass-production and relocate to specialized cities where production costs are lower.[21] This is consistent with big plants showing a larger propensity to co-locate in specialized areas (i.e. a larger spillover parameter). In a recent paper, Holmes and Stevens (forthcoming) develop a new theoretical background that, by

[20] See Maurel and Sedillot (1999) for comparing results, for instance.

[21] The authors also find a strong evidence of this predicted pattern for French employment areas for the period $1993 - 1996$.

assuming different elasticities for local and export sales, leads to an heterogenous size of firms depending on the population sizes and, more importantly, on the degree of industrial specialization across locations. Manufacturing activities, that bear lower transport costs than services or retail industries, are shown to concentrate in low population locations because of labor costs advantages that largely exceeds the benefits of markets proximity. Furthermore, in their empirical analysis, both Holmes and Stevens (2002), and Barrios, Bertinelli and Strobl (2003), find evidence that big plants are disproportionately located in specialized locations, and that this pattern is statistically significant. However, these studies do not address neither the issue of the direction of causality, nor its characterization by industries.[22]

Defining a criterion for partition

We therefore address here the issue of the endogeneity of plant size with respect to location, by partitioning the set of firms into two groups (big and small firms), as first suggested by Holmes and Stevens (2002). We then compare $\hat{\gamma}_{EG}$ and $\hat{\gamma}_{UW}$ in each sub-sample in order to understand if inconsistencies are a matter of sample heterogeneity linked to the establishment scale. To partition the universe of firms, however, one must draw a straight line between what is a big or a small establishment and this is certainly not an easy task. Studying this issue, which would reasonably involve a large number of individual and industrial features, certainly goes beyond the scope of this analysis, however. The simplest approach is to take the number of employees as a cut-off value, and we will follow this strategy using the same threshold value (20 employees, as in Holmes and Stevens, 2002) for all manufacturing industries. Table 2 contains summary statistics of the two sub-samples obtained.

Table 2: Summary Statistics.

	Sample of small establishments	Sample of big establishments
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	41,343
% of manuf. plants	93.01	6.99
% of manuf. employment	42.20	57.80

Another possibility would have been to define an industry specific critical level of employment calculated on the basis of market shares. According to the assumption that small (big) firms account for a share y $(1-y)$ of the market, one can trace back the only cut-off value of establishment size

[22]A noticeable exception to this lack of analysis, although in a different framework, is given by Duranton and Overman (2003).

that is consistent with this hypothesis.[23] Nevertheless, this approach seems at least as arbitrary as ours. Furthermore, the cut-off value of 20 workers for Italy has a valuable economic meaning, as both the fiscal and legal status of a firm remarkably change if it has more or less than 20 employees. For instance, many fiscal incentives exist for firms with less than 20 employees only ('piccole imprese'), while in order to have an employee board a firm needs to have at least 20 employees.

Comparing estimators along the partition of small vs big plants

Tables 3 and 4 refer respectively to calculations for the sub-sample of plants with less than and at least 20 employees.[24]

Table 3: Comparison of the estimators ($\hat{\gamma}_{EG}$ and $\hat{\gamma}_{UW}$) on small plants data.

	3-digits				2-digits			
	LLS		Provinces		LLS		Provinces	
	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$
Average value	0.024	0.022	0.034	0.030	0.009	0.010	0.014	0.017
Average st. deviation	0.0016	0.00010	0.0025	0.0015	0.0002	0.0001	0.0003	0.0002
R^2	0.90		0.89		0.83		0.87	
R^2 ranks	0.92		0.92		0.95		0.85	
Number of industries	103		103		23		23	
Number of spatial units	784		95		784		95	
Number of manuf. plants	549,747		549,747		549,747		549,747	

Table 4: Comparison of the estimators ($\hat{\gamma}_{EG}$ and $\hat{\gamma}_{UW}$) on big plants data.

	3-digits				2-digits			
	LLS		Provinces		LLS		Provinces	
	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$	$\hat{\gamma}_{EG}$	$\hat{\gamma}_{UW}$
Average value	0.036	0.033	0.049	0.046	0.019	0.012	0.026	0.019
Average st. deviation	0.0047	0.0020	0.0069	0.0030	0.0018	0.0004	0.0026	0.0006
R^2	0.81		0.82		0.38		0.48	
R^2 ranks	0.73		0.75		0.63		0.57	
Number of industries	103		103		23		23	
Number of spatial units	784		95		784		95	
Number of manuf. plants	41,343		41,343		41,343		41,343	

In both cases, the relation between the weighted and the un-weighted estimator is now quite good, suggesting that, once controlled for the relation among concentration and the establishment size, the EG location model is consistent with data. This is particularly true for the sub-sample of small establishments where the correlation among both measures is at 0.9 at least in all specifications.[25]

[23]This is actually the approach followed in a related work by Duranton and Overman (2003).

[24]See Tables in Appendix B for detailed 2-digits sectoral results. Detailed 3-digits results available upon request.

[25]Strong correlations for the sample of small plants is reasonably due to a larger homogeneity among small units. For the sample of big firms, the ratio between the smallest and the biggest plant is around one hundred (much higher than 20), as confirmed by the coefficients of variation in table 2.

Moreover, differences for the means reduce dramatically in most cases to values around 10%. The fact that a positive (although strongly reduced) gap still remains indicates that an approach that considers γ as a continuous function of size will probably be preferable to our simple partition scheme. Nevertheless, our approximation seems to be an acceptable compromise between the need to correct for size effects and computational complexity. The heterogeneity of concentration measures with respect to the size of plants is further supported by the strong decline of the discrepancies on standard errors when splitting the sample. While the un-weighted estimator has, coherently with the underlying model, a lower variance, the magnitude of the difference is now compatible with a simple efficiency problem rather than with a lower informational value of the co-location of small plants (heteroscedasticity).

Finally, as a further evidence that industrial spillovers depend positively on the size of plants, we present in Table 5 the result of an industry fixed-effects panel regression (for the three census years 81 91 96), using the difference $\hat{\gamma}_{EG} - \hat{\gamma}_{UW}$ as the independent variable to be regressed on Size=ln(average size of establishments) in each industry, with u^s being the industry fixed-effect.[26]

Table 5: Panel industry fixed-effects regression of $\hat{\gamma}_{EG} - \hat{\gamma}_{UW}$ on Size=ln(average size of plants)

	Coefficients or Tests Values			
	LLS		Provinces	
Size	0.0222*	(0.0035)	0.0181*	(0.0055)
Constant	-0.0276*	(0.0095)	-0.0176	(0.0151)
R^2	0.84		0.76	
Number of observations	309		309	
Number of years	3		3	
Year dummies	Yes		Yes	
Degree of freedom	203		203	
$Corr(Size, \hat{u}^s)$	0.32		0.24	
$F(3, 203)$ Test on coefficients	14.51*		3.92*	
$F(102, 203)$ Test on $u^s = 0$	9.85*		8.41*	

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

As expected, the coefficient of plant size is positive and highly significant. Furthermore, in unreported GMM estimations in which the model is estimated in first differences using both contemporaneous and lagged levels of Size as instruments,[27] the Sargan test strongly refuses the null of the strong exogeneity of Size, suggesting that there is no clear direction of causality between the size of plants and concentration. This is certainly a very preliminary result that would deserve further investigations that are, however, beyond the scope of this paper.[28]

[26] We use data on the universe of firms for the 3-digits industrial classification.

[27] See Arellano (2003) for further details on estimation and hypothesis testing in panel data model with endogenous and/or predetermined variables.

[28] In this simple panel regression framework, we have only one explanatory variable (Size), while unobserved variables are accounted for by means of the time-invariant industry effect u^s. The share of the variance of the independent

Once it is recognized that the partition into small and big units gives coherent results independently on the particular estimator used, it seems then reasonable to rely on these estimates to evaluate different concentration patterns. Comparing the corresponding values of Tables 3 and 4, one can see that the estimated correlation of co-locations among establishments is about 50% higher for big plants compared to small ones. Therefore, we can reasonably presume that, if splitting plants according to a larger cut-off than 20, this gap would be even larger. Holmes and Stevens (2002) find comparable values for the U.S., with a ratio of EG indexes of plants belonging to the fourth quartile (average size of 268 employees) over the first quartile (average size of 25 employees) of around 2.

Industrial concentration patterns of big vs small plants

Measures of concentration being different according to the size of plants, one can try to address the following interesting question: Is the observed concentration in an industry mainly driven by big or small establishments? Apart from its value for theory, this is likely to be an important feature in the design of industrial or firm-size oriented policies. Comparing the $\gamma's$ of big *vs* small plants for each industry is a way to tackle this issue. In Table 6, we exhibit the industries that display the largest and lowest ratios between big ($\hat{\gamma}_{UW}^{\geq 20}$) and small ($\hat{\gamma}_{UW}^{<20}$) plants spillovers, at the scale of LLS. In particular, the first column contains the 10 3-digits industries whose spatial concentration is mainly due to small units (low $\hat{\gamma}_{UW}^{\geq 20}/\hat{\gamma}_{UW}^{<20}$), while in the second one are listed the 10 sectors whose concentration is mainly driven by big plants.[29] Within the first group, one can identify industries belonging to major Italian Industrial Districts, as textile industries (171 and 172), 'Fur' (183) or 'Watches and Clocks' (335). By contrast, the second group encloses mass-production activities like 'Motor vehicles' (341), 'Transport equipment' (355), and 'Aircraft' (353), as well as industries belonging to the final stage of production like 'Wearing apparel' (182) or 'Made-up textile' (174). Let us now turn to the issue of spatial auto-correlation and its causal relation with the size of plants.

4 Spatial Agglomeration and the Size of Plants

Starting from the concept of spatial correlation, we first introduce the Moran's index, which is often used to measure it. Afterwards, we use this indicator in order to disentangle the different externalities driving the process of spatial agglomeration, as well as to provide some insights on the relation between size and the agglomeration scope of plants.

variable $\hat{\gamma}_{EG} - \hat{\gamma}_{UW}$ explained by those fixed effects alone being huge (78% for LLS and 70% for provinces), our results potentially suffer from an important bias due to time-varying omitted variables.

[29] We prefer to use data on LLS because Industrial Districts are more easily identified at this level of disaggregation. This leaves open the possibility of overlooking concentration phenomena that spill-over LLS boundaries. Nevertheless, the general picture coming out from provinces data is not dramatically different.

Table 6: The process of concentration: big vs small plants.

Concentration mainly driven by small plants			Concentration mainly driven by big plants		
NACE	3-digits industry	$\hat{\gamma}^{\geq 20}_{UW} / \hat{\gamma}^{<20}_{UW}$	NACE	3-digits industry	$\hat{\gamma}^{\geq 20}_{UW} / \hat{\gamma}^{<20}_{UW}$
372	Recycling of non-metal waste and scrap	0.203	355	Manuf. of other transport equipment n.e.c.	88.058
171	Preparation and spinning of textiles fibres	0.228	341	Manuf. of motor vehicles	20.953
172	Textile weaving	0.261	223	Reproduction of recorded media	12.880
160	Manuf. of tobacco products	0.299	267	Cutting, shaping and finishing of stone	6.061
183	Dressing and dyeing/Manuf. of fur	0.303	287	Manuf. of other fabricated metal products	5.890
272	Manuf. of tubes	0.361	331	Manuf. of medical and surg. equipment etc.	5.656
312	Manuf. of electricity distribution etc.	0.409	353	Manuf. of aircraft and spacecraft	4.570
283	Manuf. of steam generators etc.	0.581	174	Manuf. of made-up textile articles	4.527
316	Manuf. of electrical equipment n.e.c.	0.595	243	Manuf. of paints, varnishes etc.	4.517
335	Manuf. of watches and clocks	0.643	182	Manuf. of other wearing apparel etc.	4.510

4.1 Concentration *vs* Agglomeration

As mentioned in Section 3.2, the concentration process extends outside the limits of Local Labor Systems, which casts some doubt on the idea that the two mainly sources of concentration should restrict to natural advantages or localized externalities, as studied in Ellison and Glaeser (1999). However, one could claim these large-scale forces to be the result of an uneven factor endowments distribution. As endowments play a role at all spatial scales, if one assumes that their distribution is more heterogeneous for bigger geographical units rather than for small ones, then one should find evidence of the above described pattern. Although this assumption seems quite plausible, we argue that, by ignoring the spatial content of location choices, the standard approach of concentration misses an important feature, we illustrate by means of the following example. Let us distribute 12 plants over the 9 locations corresponding to the cells of a 3x3 grid. All the measures of concentration we have presented so far satisfy the property of being invariant to a permutation in the observations order, so that their value would be the same in both case a) and b) of Figure 3. However, one has to admit that the two distributions differ in a substantial way that cannot be captured by such measures.

Figure 3: Concentration vs Agglomeration

3	3	
3	3	

a)

	3	
3		3
	3	

b)

As pointed out by Arbia (2001), the uneven distribution of a spatial phenomenon is characterized by (at least) two different features:

1. The first feature is 'concentration': i.e. an a-spatial concept of variability that is invariant to permutation. Indicators like LQs, Locational Gini, $\widehat{\gamma}_{EG}$, or $\widehat{\gamma}_{UW}$ are actually measures of spatial concentration, because they give *a permutation invariant quantification of how much variable is a phenomenon with respect to some average.* However, they treat data without considering their actual position of is space (i.e. they do not consider distances among spatial units).

2. The second feature is 'agglomeration': i.e. the degree of spatial correlation among observations. As far as a phenomenon displays some variability, observations may be distributed in a two-dimensional space in order to form some specific distance-based pattern. More precisely, there is spatial correlation as long as knowing the value of the observation in location i is 'linearly' informative about neighboring spaces.[30]

Coming back to Figure 3, in case b) plants are evenly distributed among locations, like if there was a completely random process that associates our given 4 clusters of 3 plants to the 9 available locations. In such a situation, although there is some concentration, spatial agglomeration is clearly not the source of such variability. By contrast, in case a), plants are distributed in such a way that we can clearly identify an "agglomeration" of plants in the upper-left corner of our grid, that leaves an empty space of "abandoned" locations. Location is here highly informative because high (low) values are surrounded by high (low) values following a pattern that is known as *positive spatial auto-correlation*. Now, the interesting thing is that positive spatial correlation is a key feature of New Economic Geography (NEG) models, and in particular of the so-called *market potential functions* that can be derived from them.[31] Using different NEG models, Fujita, Krugman and Venables (1999), for instance, obtain several reduced-form equilibrium equations (the market potential functions), where a variable expressing the attractiveness of a location (as profits, or real wages) turns out to be a positive function of the level of economic activity in the surrounding regions (i.e. positive spatial correlation).

Moreover, as detailed in Fujita and Thisse (2002), there is a substantial difference in the geographical scope of traditional externalities and the so-called *pecuniary externalities* that are the engine of agglomeration in NEG models. Traditional spillovers such as Marshallian externalities,

[30] For further details about spatial processes and measures of association, see Cliff and Ord (1981), and Anselin (1995).

[31] For a comprehensive review of the NEG theoretical framework, see Fujita, Krugman and Venables (1999), Fujita and Thisse (2002), and Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud, (2003).

require some degree of physical interaction among agents within the same place. They are therefore bounded to relatively small geographical units that should fit the definition of local labor markets. In terms of Figure 3, traditional *localized* externalities may give the plants an incentive to form clusters of 3 units instead of remaining isolated, but not control for the mutual position of these clusters within space. By contrast, acting trough the price system, *pecuniary* externalities arising from the final markets proximity and/or the input-output linkages, can have a large-scale impact on the location of economic activities.

With respect to our analysis, pecuniary externalities can thus be part of those forces that boost concentration beyond the LLS disaggregation, with positive spatial correlation being their distinctive feature with respect to both traditional externalities or factor endowments. Even though factor endowments clearly matter at large-scale, it would be difficult to understand why nature should have put resources in space following some specific pattern and in particular that of positive spatial correlation.

4.2 The Moran's I index of spatial agglomeration

In order to measure the degree of spatial correlation of a centered variable y, one can use the Moran's I index. To do so, one must start by defining the $M \times M$ spatial weighting matrix W, as the matrix whose generic element w_{il} is the relative weight of location l for region i. w_{il} is inversely related to the distance d_{il} between i and l, which may actually take different analytic forms like $d_{il}^{-\tau}$, or $\exp^{-\tau d_{il}}$, or can be based on a contiguity criterion. For our investigations, we choose a first-order contiguity matrix, that gives weight one to those locations that are contiguous to a given unit i and zero otherwise. The Moran's I formula is then given by

$$I = \frac{(M/S_0) \sum\limits_{i=1}^{M} y_i \sum\limits_{l=1}^{M} w_{il} y_l}{\sum\limits_{i=1}^{M} y_i^2}, \tag{11}$$

where y_i is the variable value in location i, and $S_0 = \sum_{l=1}^{M} \sum_{l=1}^{M} w_{il}$.[32] Moreover, as suggested in Anselin (1988), we row-standardize[33] this matrix so that S_0 equals to M. The most intuitive interpretation of the Moran's I index is found in the regression context. If we actually regress the (spatially) weighted variable Wy on y (where y is the vector of y_i), then the slope coefficient of the regression is precisely given by I, as the Moran index is the ratio of $cov(W_i y, y_i)$ over $var(y_i)$, where $W_i = (w_{i1}, ..., w_{il}, ..., w_{iM})$ is the corresponding raw of the weight matrix W. To this respect, the

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

[33] Elements of each row are therefore divided by the sum of the row elements. For further details about spatial weighting matrices, see Anselin (1988).

Moran index can be interpreted as a measure of the linear correlation between y_i and its neighbors' observations.

This index can be interpreted closely to the one of Ellison and Glaeser (1997). Let the variable y, whose values are assumed to be exogenous, be distributed over space according to a random scheme. Each value has therefore the same probability to be observed in any location i, and data will not show any specific spatial pattern. Under this null, the Moran index has a mean $E[I] = -1/(M-1)$[34] while on the alternative of spatial correlation, it takes large positive or negative values according to the sign of the correlation. A test on the absence of spatial correlation can therefore be made by using the variance of I under the null. To be consistent with the EG model, and more particularly with the un-weighted specification of the γ estimator, we use the variable $y_i = \left(\frac{n_i}{N}\right) - x_i$ as the basis of our investigations. The reason of this choice is that $\widehat{\gamma}_{UW}$ can be rewritten as a "spatially corrected" measure of the variability of our $y_i's$.

Controlling for the overall manufacturing and industrial concentration

Starting from a "raw" measure of concentration (G_{UW}), the index γ_{UW} further controls as well for the overall distribution of activity $(1 - \sum_{i=1}^{M} x_i^2)$, as for industrial concentration (1/N).[35] For the two investigations to be comparable, the index I needs therefore to account for both sources of concentration. However, an overall spatial concentration control is not necessary here because, in the linear regression of Wy on y, multiplying y on both sides by a constant would not affect the slope coefficient (I). The issue of industrial concentration is, by contrast, more difficult to deal with. Since there is no proper formal way to do it, we resort to the same basic reasoning as the EG framework. The kind of situation that one wants to avoid is to say that a sector where only few firms operate is strongly agglomerated. In such a case indeed, spatial correlation simply comes from the fact that the number of plants is small compared to locations (industrial concentration). As an attempt to control for this problem, we then "re-size" the number of locations for computing the index I on the basis of the number of "active" plants. In particular, for each sector, we calculate I using only those $y_i's$ for which $n_i/N \neq 0$ (*i.e.* we do not consider those regions where no plant is located), and this actually amounts to redefine the spatial weighting matrix by eliminating those rows and columns corresponding to "abandoned" areas. Although quite demanding in terms of computations, this strategy makes sure that to a positive significant spatial correlation will also correspond a positive significant $\widehat{\gamma}_{UW}$.[36]

[34] According to the (conditional) randomization hypothesis developed by Cliff and Ord (1981).
[35] The industry control variable for γ_{EG} and γ_{MS} is instead the Herfindahl index H.
[36] This logical requirement comes from the relation among the a-spatial variability of the γ_{UW} and the Moran's I coefficient, we saw so far. A significant spatial correlation implies an important data variability, while the reverse is not true. In fact, as shown in Figures 3a and 3b, spatial correlation is one of the way data dispersion organizes itself in a two-dimensional space.

Controlling for "localized" clusters of production

In the real world, both "localized" and "pecuniary" externalities influence the location of plants and its degree of spatial correlation. In some cases, local clusters of production spread over more than one location unit, generating an observed spatial correlation that has a very local nature. Therefore, failing to take into account such phenomena might lead to an over-estimation of the degree of the spatial agglomeration related to "pecuniary" externalities (i.e. the correlation related to large distances patterns). As a control for the presence of such local clusters, we follow the procedure suggested by Anselin (1995), which consists in identifying those locations that correspond to local outliers, or "hot spots" and, subsequently, eliminating them in the same way as for empty locations.[37]

Let us illustrate the presence of "hot spots", by looking at the distribution of activities in the 'Motor vehicles trail and semi trail' industry (2-digits NACE number 34) and the 'Footwear' industry (3-digits NACE number 193). Figure 4 shows the respective 21 and 10 spatial outliers we find for these industries, with different colors corresponding to their significance level.[38] The five clusters one can trace on the left-map of Figure 4 are all centered around a major Italian city. Going from left to right, and then from top to bottom, these "hot" cities are 'Torino', 'Milano', Venezia', 'Firenze' and 'Roma'. The test thus suggests that something "special" occurs in big metropolitan areas, making them depart from the overall spatial distribution pattern of the motor vehicles industry. With respect to the footwear industry (right-map of Figure 4), one is not surprised to find among the outliers both 'Monte San Pietrangeli' and 'Montegranaro', depicted in Section 2.2 as the top specialized LLS in manufacturing according to the employment and the number of firms, respectively. These two locations form, together with other two contiguous LLS, the center-right cluster that corresponds, as the center-left cluster, to an historical shoes production site. The upper cluster is instead centered around the city of 'Milano'.

Further insights can be gained by looking at Figure 5, that depicts the "type" of local correlation observed in the data. While the motor vehicle cluster of 'Torino' is characterized by highly specialized locations surrounded by highly specialized neighbor units (High-High, as shown on the left-map of Figure 5), all other motor vehicle clusters exhibit a low level of specialization. 'Torino' is a spatial

[37] The identifying procedure suggested in Anselin (1995) uses information coming from different sources. In particular, for each location, we calculate the so-called *local* Moran statistic I_i that measures the correlation between a particular y_i and its specific neighbors. Using the conditional randomization approach, one can make a test of the local instability of $I_i's$. Furthermore, as the Moran index corresponds to the sample average of the $I_i's$, one can use the sample variance of the $I_i's$ and, assuming normality, identify the extreme values with a two-sigma rule. Finally, only those locations corresponding to extreme values for both tests are actually considered as spatial outliers. 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, the presence of outliers is more likely to occur than under a random scheme.

[38] The probabilities given refer to the second test, i.e. the one that uses normality and the sample variance of the $I_i's$.

Figure 4: Outliers for Motor Vehicles (left) and Footwear Industries (right)

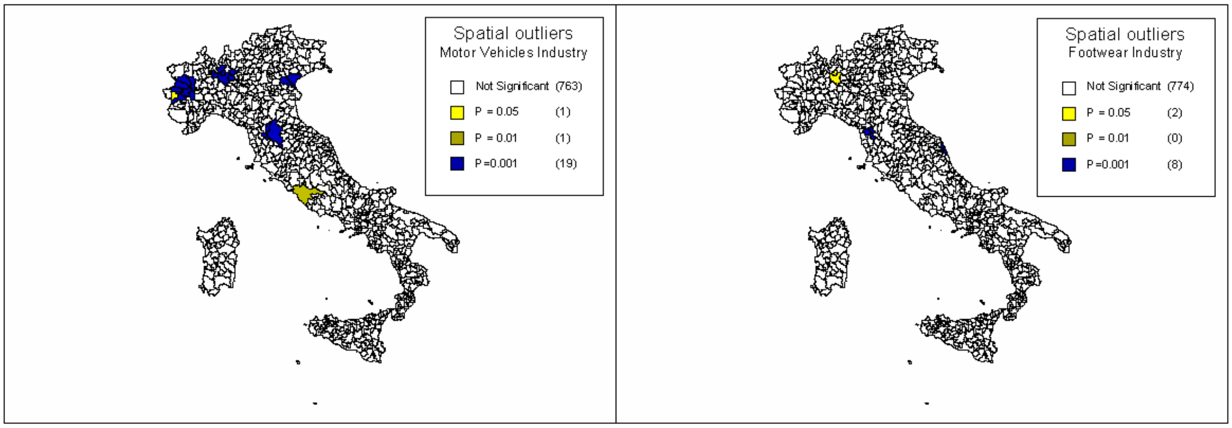
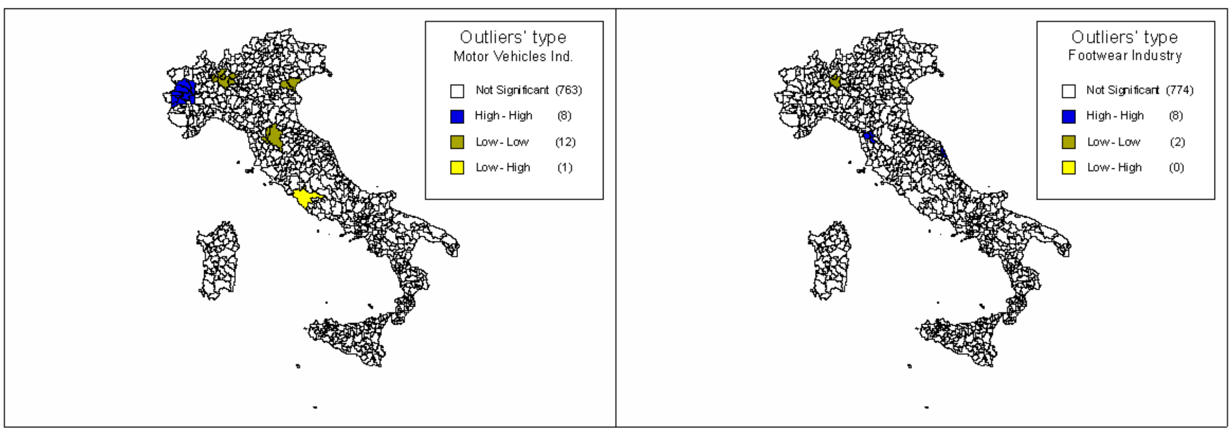


Figure 5: The Nature of Spatial Outliers for Motor Vehicles (left) and Footwear Industries (right)



outlier because of both historical reasons (FIAT was born there), as well as for Marshall (1890) stories of Labor Market size and acquired workers' skills. The same story applies to the eight high-high clusters of production in the footwear industry. By contrast, for the remaining outliers, input requirements and urbanization externalities are probably the best explanation. Building cars, trucks and their accessories requires space, and the land rent of metropolitan areas is probably too high for land-consuming activities to be profitable there.[39]

As one can notice, the analysis of spatial outliers has a value *per se*, because it allows to identify locations where localized externalities foster industrial concentration (clusters of productions), in areas where diversification forces dominate (extremely low specialized locations). However, the problem is that, when considering small distances, natural advantages, localized and pecuniary externalities can operate at the same time. Our main goal here is to emphasize the contribution of

[39]Both Holmes and Stevens (2002) and Duranton and Puga (2001a) found evidence of such a relocation pattern for mass-production activities.

pecuniary externalities to specialization patterns. We therefore prefer to take a rather demanding approach that considers local outliers as noisy observations to be removed in order not the over-evaluate the degree of spatial agglomeration.[40]

We show in Section 4.3 that distance-based externalities play an important role in many manufacturing industries, even once controlled for extreme agglomeration cases.

4.3 Evidence of Pecuniary Externalities

Table 7 presents summary statistics about the two spatial matrices (one for LLS and the other for provinces).[41] In particular, the last row contains the shorter distance needed for all units to have at least one neighbor (minimum allowable distance cut-off). Roughly speaking, our first-order contiguity matrices thus correspond to a distance-bands criterion that leads to cut-off values of 34 (LLS) and 84 (Provinces) km. As a robustness check, we have actually tried to use matrices based on these distance-bands, but, as results are almost equivalent, we present here the results arising from the contiguity criterion only.

Table 7: Summary Statistics on Spatial Weighting Matrices

	LLS	Provinces
Dimension	782	95
% of non-zero weights	0.67	4.66
Average number of neighbors	5.23	4.38
Average distance among units (km)	467	455
Minimum allowable distance cut-off (km)	34	84

Table 8 summarizes the analysis of spatial correlation for the manufacturing industries.[42] Interestingly, the mean Moran increases with the scale of spatial aggregation (the Moran's I for provinces exceeds that of LLS), and with the level of industry aggregation (the Moran's I for 2-digits industries exceeds that of 3-digits). The first pattern indicates that pecuniary externalities do matter at the scale of LLS and even more at the larger scale of provinces, despite larger distances. If not, we should have found evidence of a declining (or at least stable) degree of spatial correlation for provinces compared to LLS. The second trend, that is even stronger, is suggestive of important input-output linkages among plants belonging to the same broad industrial definition (2-digits). This

[40]Our approach is demanding with respect to the presence of pecuniary externalities in the sense that, with all our controls, we are excluding *a priori* as well their impact on a small geographical scale as their contribution to extreme agglomeration phenomena. Nevertheless, it is important to remark that results obtained without eliminating spatial outliers are not tremendously different, with industries displaying slightly stronger agglomeration patterns.

[41]In the case of LLS, two units (corresponding to two islands) have been actually dropped because of their lack of contiguous locations.

[42]See Appendix B for results on specific 2-digits industries. Detailed results for the 3-digits classification are availbale upon request.

linkages were already emphasized by Ellison and Glaeser (1997) and Maurel and Sedillot (1999), in order to explain their results on co-location indexes. Our analysis complements their findings by assessing that, coherently with the NEG framework, they have a spatial nature and a large spatial scope.

Table 8: Evidence of Agglomeration for all Plants: $Moran's\, I$

	3-digits		2-digits	
	LLS	Provinces	LLS	Provinces
Average value	0.112	0.126	0.156	0.182
Average st. deviation	0.1020	0.1302	0.0912	0.1285
Average number of outliers	12.82	4.63	13.35	4.70
Average number of units used	299.31	75.12	538.13	86.52
Average number of connected units used	284.90	73.81	532.65	86.17
Number of industries	103	103	23	23
Number of spatial units	782	95	782	95
Number of manuf. plants	591,110	591,110	591,110	591,110

Concerning statistic significance, using a two-sigma rule leads to the conclusion that pecuniary externalities do matter for 37% to 87% of industries, depending on the disaggregation chosen. This is a quite strong result that confirms how important is the role of distance-based externalities for the location of plants location. As for comparability with the study on concentration, the mean (respectively median) Moran being 0.126 (0.116) for provinces, 47.2% of industries can be classified as very agglomerated.

4.4 Agglomeration and the size of Plants

In Section 3.3 we saw that big plants were more concentrated than small ones. As we already mentioned, this result is coherent with previous empirical findings, and suggests that big business units are more sensitive to localized externalities, while small establishments are attracted by diversified areas where both new products and knowledge are created. An interesting related issue we want to address in Section 4.4 is the relation between the size of plants size and spatial agglomeration. Tables 9 and 10 contain summary results on spatial agglomeration for the sample of small and big establishments.[43]

As one can immediately notice, small units show a stronger pattern of spatial agglomeration than big plants. They seem to be more sensitive to pecuniary externalities both in terms of the mean Moran, and in the number of significantly agglomerated industries. The mean I is 2-3 times larger for small plants, and only few industries display a significant spatial correlation in the sample of big

[43] See Appendix B for detailed 2-digits results (3-digits results available upon request).

Table 9: Evidence of Agglomeration for Small Plants: $Moran's\, I$

	3-digits		2-digits	
	LLS	Provinces	LLS	Provinces
Average value	0.102	0.137	0.181	0.205
Average st. deviation	0.1050	0.1356	0.0941	0.1314
Average number of outliers	12.70	4.66	13.65	5.13
Average number of units used	299.36	75.07	537.70	86.09
Average number of connected units used	284.91	73.88	532.04	85.61
Number of industries	103	103	23	23
Number of spatial units	782	95	782	95
Number of manuf. plants	549,747	549,747	549,747	549,747

Table 10: Evidence of Agglomeration for Big Plants: $Moran's\, I$

	3-digits		2-digits	
	LLS	Provinces	LLS	Provinces
Average value	0.057	0.046	0.065	0.074
Average st. deviation	0.0564	0.07554	0.0490	0.0906
Average number of outliers	13.33	4.82	15.22	5.09
Average number of units used	298.64	74.94	536.26	86.17
Average number of connected units used	284.16	73.66	530.52	85.82
Number of industries	103	103	23	23
Number of spatial units	782	95	782	95
Number of manuf. plants	41,343	41,343	41,343	41,343

business units. Furthermore, looking at the number of connected geographical units,[44] reveals that this result is not due to a different sample size. Once eliminated both spatial outliers and locations with no industry employment, the number of remaining regions which are connected among each other is virtually the same for both samples.

This relation between size and agglomeration is further confirmed by an industry panel fixed-effects regression (for the three census years 81 91 96), where I is the independent variable to be regressed on Size=ln(average size of establishments) and u^s is an industry fixed-effect.[45] The coefficient of Size in Table 11 is negative and significant and, parallel to the case of concentration, unreported GMM estimations of the model in first-differences suggest that the size of plants and agglomeration simultaneously influence each other. Again, this result should be taken with care because the regression framework is potentially biased by omitted time-varying variables, and further investigations are needed.

As a comparative exercise with previous results on concentration patterns (see Table 6), we report in Table 12 the 10 3-digits industries whose agglomeration patterns are more driven by small (big)

[44] By ignoring those regions with no employment and spatial outliers, one may create 'islands' that are no longer connected with other locations.

[45] As before, we use data on the universe of firms for the 3-digits industries.

Table 11: Panel industry fixed-effects regression of Moran's I on Size=ln(average size of plants)

	Coefficients or Tests Values			
	LLS		Provinces	
Size	-0.0256**	(0.0148)	-0.03401*	(0.0161)
Constant	0.1871*	(0.0459)	0.2271	(0.0614)
R^2	0.76		0.78	
Number of observations	309		309	
Number of years	3		3	
Year dummies	Yes		Yes	
Missing industries	8		8	
Degree of freedom	195		195	
$Corr(Size, \hat{u}^s)$	-0.15		-0.09	
$F(3, 203)$ Test on coefficients	3.00**		8.77*	
$F(102, 203)$ Test on $u^s = 0$	5.78*		5.53*	

Note: Standard errors in brackets. *, ** denote significance at respectively the 1% and 5% levels. 'Missing' stand for those industries for which, after deletion of no-employment areas and outliers, the number of remaining connected locations is too low to compute the Moran's I.

plants, as identified by the ratio of the Moran's I in the two samples ($I^{\geq 20}$/ $I^{<20}$). Some interesting conclusions can be drawn from this comparison at the scale of LLS:[46]

1. First of all, manufacturing industries whose large concentration is mainly driven by small plants, do not display the same pattern for agglomeration. Among these industries, one can find some well-known Italian historical districts, like 'Textile weaving' (172), 'Wearing apparels' (182), or 'Watches and clocks' (335), that are likely to benefit from localized externalities.

2. Many mass-production activities whose concentration is essentially driven by big plants also lack a spatial pattern, like 'Motor vehicles' (341) and 'Other transportation equipment' (355). This is likely to be the outcome of strong internal returns to scale in these sectors in which only very few big plants operate.

3. By contrast, mass-production industries belonging to the final stage of production, like 'Wearing apparel' (182), 'Made-up textile' (174), Food industries (154, 156 and 158), 'Printing' (222) or 'Luggage and bags' (192), do exhibit positive spatial correlation. These consumers' demand sensitive industries are actually among those who are mostly agglomerated, suggesting that final-demand linkages are a key variable in understanding their location. Most examples of these downstream industries can be found in the sample of small plants.

4. Industries producing inputs for other sectors, like 'Manufacture of parts and accessories for vehicles' (343), 'Cutlery, tools, and gen. hardware' (286), 'Machine tools' (294), or 'Forging,

[46]Data on LLS are preferred here mainly for comparability of results with those contained in Table 6. Again, this leaves open the possibility of overlooking concentration phenomena that spill-over LLS boundaries. Nevertheless, the general picture coming out from provinces data is not dramatically different.

pressing, stamp of metals' (284), also display strong spatial effects. Most examples of these upstream industries can be found in the sample of big plants.

Table 12: The process of agglomeration: big *vs* small plants.

Agglomeration mainly driven by small plants			Agglomeration mainly driven by big plants		
NACE	3-digits industry	$I^{\geq 20}/\ I^{<20}$	NACE	3-digits industry	$I^{\geq 20}/\ I^{<20}$
158	Manuf. of other food products	0.004	363	Manuf. of music instruments	20.099
222	Printing and service activities related to printing	0.007	343	Manuf. of parts for motor vehicles, etc.	4.966
246	Manuf. of other chemical products	0.056	275	Casting of metals	2.936
201	Sawmilling and planing, impregnation of wood	0.103	286	Manuf. of cutlery, tools, etc.	2.039
154	Manuf. of vegetable, animal oils and fats	0.147	192	Manuf. of luggage, handbags, etc.	1.371
351	Building and repairing of ships and boats	0.147	292	Manuf. of other gen. purpose machinery	1.325
203	Manuf. uilders' carpentry and joinery	0.162	294	Manuf. of machine tools	1.238
281	Manuf. of structural and metal products	0.178	284	Forging, pressing, stamping metal, etc.	1.156
287	Manuf. of other fabricated metal products	0.182	362	Manuf. of jewellery and related articles	1.043
156	Manuf. of grain mill products, etc.	0.186	182	Manuf. of other wearing apparel, etc.	1.019

5 Conclusions

Following the model-based approach of Ellison and Glaeser (1997), we develop a coherent framework to test for the presence of both spatial concentration and agglomeration in Italian manufacturing industries. In particular, we focus on positive spatial correlation, that is a peculiar feature of NEG models, as a mean to disentangle pecuniary externalities from other forces shaping the distribution of manufacturing activities.

Furthermore, by means of this framework complemented with some panel estimations, we shed some light on the relation between the size of plants and the degree of concentration and agglomeration. Our findings suggest that, once controlled for size effects, the EG model is actually coherent with its *a priori*, giving a reliable measure of the concentration of plants. This result thus casts some doubt on the relevance of the fixed plants size DSK monopolistic framework to structurally account for the role of the so-called "pecuniary" externalities compared to more "localized" ones.

Almost all industries display a significant departure from randomness and, for the majority of them, big plants have a much higher propensity to concentrate in specialized areas. This result is coherent with the empirical and theoretical findings of Holmes and Stevens (2002, forthcoming), and suggests that big business units are more sensitive to traditional externalities, while small establishments are attracted by diversified areas where both new products and knowledge are created.

As for "pecuniary" externalities, many sectors (at least 40%) do show a significant positive spatial auto-correlation. Coherently with the NEG framework, upstream and final-consumers oriented industries are those which display more agglomeration patterns. By contrast, our analysis suggests that mass-production activities of big plants, as well as some historical Italian Districts, are mainly driven by traditional externalities rather than pecuniary ones. Furthermore, contrary to the case of concentration, small plants seem to be more sensitive to pecuniary externalities, and especially to final demand linkages. This is also confirmed by our panel regressions, which still suggests a simultaneous relationship among size and agglomeration.

There are several directions for further research. A first valuable contribution would be to extend the EG location model by introducing distance declining spill-over effects in order to account for spatial correlation. This would allow to treat localized and large-scope externalities within the same theoretical ground facilitating the separation task. Another interesting topic is related to the two-way causality we found between the size of plants and the spatial features of their location choice (in particular agglomeration). Further research on this field, especially from theorists, is certainly needed. As for the pervasiveness of pecuniary externalities, it would indeed be desirable to have further evidence from other countries. This will probably be the next step of our research. Finally, the topic of input-output linkages and their impact on location choices deserves a deeper analysis. Although it is possible to some extent to deal with this issue within the EG framework (by means of a co-agglomeration index), what one would really need is a characterization of both vertical and horizontal interrelations among sectors that would allow to go beyond the limit of the simple NACE classification.

References

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

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

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

Arellano, M., (2003). *Panel Data Econometrics. Oxford University Press.*

Baldwin, R., Forslid, R., Martin, P., Ottaviano, P., and Robert-Nicoud, F., (2003). "Economic Geography and Public Policy". *Princeton University Press.*

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., (2003). "Agglomeration economies and the location of industries: a comparison of three small European countries". *CORE Discussion Papers*, 67.

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

BURCHFIELD, M., OVERMAN, H.G., PUGA D., AND TURNER, M.A., (2003). "Sprawl:A Portrait from Space". *Mimeo.*

CLIFF, A.D., AND ORD, J. K., (1981). *Spatial Processes: Models and Applications. Pion*, London.

COMBES, P.-P., (2000). "Economic Structure and Local Growth: France, 1984-1993". *Journal of Urban Economics*, Vol. 47, 329-355.

COMBES, P-P. AND M. LAFOURCADE, (2001). "Transport Costs, Geography and Regional Inequalities: Evidence from France", *CEPR Discussion Paper*, 2894. Revised version: http://www.enpc.fr/ceras/lafourcade/artinf.pdf.

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

DIXIT, A. AND J. STIGLITZ, (1977). "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, Vol.67(3), 297-308.

DUMAIS, G., ELLISON, G., AND GLAESER, E.L., (2002). "Geographic concentration as a Dynamic Process". *The Review of Economics and Statistics*, Vol.84(2), 193-204.

DURANTON, G., AND OVERMAN, H.G., (2002). "Testing for Localisation Using Micro-Geographic Data". *CEPR Discussion Papers*, 3379.

DURANTON, G., AND PUGA, D., (2001a). "From sectoral to functional urban specialisation". *CEPR Discussion Papers*, 2971.

DURANTON, G., AND PUGA, D., (2001b). "Nursery cities: Urban diversity, process innovation, and the life cycle of products". *American Economic Review*, Vol.91(5), 1454-1477.

ELLISON, G., AND GLAESER, E., (1997). "Geographic concentration in US manufacturing industries: A dartboard approach". *Journal of Political Economy*, Vol.105, 889-927.

ELLISON, G., AND GLAESER, E., (1999). "The geographic concentration of industry: Does natural advantage explain agglomeration?". *American Economic Review*, Vol.89, 311-316.

FORSLID, R., J. HAALAND, AND K.-H MIDELFART-KNARVIK, (2002). "A U-Shaped Europe? A Simulation Study of Industrial Location". *Journal of International Economics*, 57: 273-297.

FUJITA, M., HENDERSON, J.V. AND MORI, T., (2004)."Spatial Distribution of Economic Activities in Japan and China". Forthcoming, *Handbook of Regional and Urban Economics*, Vol.4., J.V. Henderson and J.-F. Thisse (eds.), Elsevier-North Holland, Amsterdam.

FUJITA, M., KRUGMAN, P. AND VENABLES, A.J., (1999). *The spatial economy: Cities, regions and international trade. MIT Press.*

FUJITA, M., AND THISSE, J.F., (2002). *Economics of agglomeration. Cities, industrial location, and regional growth. Cambridge University Press.*

HANSON, G., (1998). "Market Potential, Increasing Returns, and Geographic Concentration", *NBER Working Paper*, 6429. http://www-irps.ucsd.edu/faculty/gohanson/Uscnty.pdf.

HENDERSON, J.V., KUNCORO A., AND TURNER M., (1995). "Industrial development in cities", *Journal of Political Economy*, 103, 1067-1090.

Holmes, T.J., and Stevens, J.J., (forthcoming). "Geographic concentration and Establishment Size: Analysis in an Alternative Economic Geography Model". *Journal of Economic Geography*.

Holmes, T.J., and Stevens, J.J., (2004). "Spatial Distribution of Economic Activities in North America". *Handbook of Regional and Urban Economics*, Vol.4, 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*, Vol.84, 682-690.

Jaffe, A.B., Trajtenberg M., and Henderson R., (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations". *Quarterly Journal of Economics*, 108, 577-598.

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*, Vol.110(4), 881-908.

Krugman, P., (1980). "Scale Economies, Product Differentiation, and the Pattern of Trade". *American Economic Review*, Vol.70, 950-959.

Marcon, E., and Puech F., (2003). "Evaluating the Geographic Concentration of Industries Using Distance-Based Methods". *Journal of Economic Geography*, Vol.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*, Vol.29, 575-604.

Melitz M., and Ottaviano G. I. P., (2004). "Market size, trade, and productivity". Mimeo.

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

Mion, G., (2004). "Spatial Externalities and Empirical Analysis: The Case of Italy". Forthcoming *Journal of Urban Economics*.

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

Redding S. and A. Venables, (2004). "Economic Geography and International Inequality", *Journal of International Economics*, 62(1):53-82.

Schmalensee R., (1977). "Using the H-Index of Concentration with Published Data." *Review of Economics and Statistics*, 59: 186193.

S. Usai, and Paci, R., (2002). "Externalities and Local Economic Growth in Manufacturing Industries". *Crenos Working Paper*, 01/13.

Appendix A: Data and the Variance of $\gamma's$ Estimators

Data on plants come from the Italian Census of Economic Activities for the years 1981, 1991 and 1996. These data-sets give detailed geographic (up to the 8192 Italian commons) and Industrial (up to 3-digit NACE revision 1 for 1981 and 1991, and 5-digits for 1996) information on the location and employment of the universe of Italian plants. Contrary to Ellison and Glaeser (1997), there is no problem of withheld data in this sample, the only limitation being that the size of plants has to be reconstructed from the size-range groups to which data are allocated. Consequently, the Herfindahl index $\sum_{j=1}^{N} z_j^2$, with z_j denoting the plant' size, is obtained using the procedure suggested in Schmalensee (1977), which is based on the assumption of a linear distribution within size-groups. Under the same assumption, we also calculate $\sum_{j=1}^{N} z_j^3$ and $\sum_{j=1}^{N} z_j^4$ in order to compute the variance of the $\gamma's$ estimators. It is important to notice that, given the high level of data disaggregation, in roughly 90% of the cases, the size of plants was directly identified. Moreover, such estimations are not necessary for $\hat{\gamma}_{UW}$, which is less demanding with respect to both $\hat{\gamma}_{EG}$ and $\hat{\gamma}_{MS}$.

Finally, as shown in Ellison and Glaeser (1997) and Maurel and Sedillot (1999), the variance of our estimators under the null of no spillovers ($\gamma = 0$) is given by:

$$var(\hat{\gamma}_{est}) = \frac{2(1-H)^2}{\left(1-\sum_{i=1}^{M} x_i^2\right)^2} \quad \Omega_{est},$$

where the quantity Ω_{est} depends on the particular estimator.

In the case of $\hat{\gamma}_{EG}$ we have:

$$\Omega_{EG} = H^2 \left(\sum_{i=1}^{M} x_i^2 - 2\sum_{i=1}^{M} x_i^3 + \left(\sum_{i=1}^{M} x_i^2\right)^2 \right) - \sum_{j=1}^{N} z_j^4 \left(\sum_{i=1}^{M} x_i^2 - 4\sum_{i=1}^{M} x_i^3 + 3\left(\sum_{i=1}^{M} x_i^2\right)^2 \right),$$

while for $\hat{\gamma}_{MS}$:

$$\Omega_{MS} = \Omega_{EG} + \left(4\sum_{j=1}^{N} z_j^3 - 2H^2\right) \left(\left(\sum_{i=1}^{M} x_i^2\right)^2 - \sum_{i=1}^{M} x_i^3 \right)$$

and finally Ω_{UW} is obtained by putting $H = 1/N$ and $\sum_{j=1}^{N} z_j^4 = 1/N^3$ in Ω_{EG}.

It can be derived from such formulas that, as long as there exists an upper bound to the absolute size of plants (locations), the variances go to zero as N (M) goes to infinity, i.e. these estimators converge (under the null) in square mean.

Appendix B: Detailed Industry Tables

The indexes G_{EG} and G_{UW} reported in the following Tables have been actually divided by $\left(1 - \sum_{i=1}^{M} x_i^2\right)$. The column 'Sign' takes the value of one if the statistic presented in the previous column is significant according to the two-sigma rule, and of zero otherwise. The column 'Outl' stands for the number of spatial outliers, while 'Units' indicates the number of locations remaining after both spatial outliers and zero industry-employment zones have been removed. Finally, 'Connect' indicates how many of these remaining units are connected with each-other by a first-order contiguity criterion. Whenever a statistic cannot be calculated (like I if 'Connect' is less or equal to one) we report n.a.

Table 13: LLS 2-digits all plants

NACE	Industry	G_{UW}	G_{EG}	$1/M$	H	$\hat{\gamma}_{UW}$	$Sign$	$\hat{\gamma}_{EG}$	$Sign$	I	$Sign$	Outl	Units
15	man of food prod and beverages	0.0049	0.0033	0.0000	0.0003	0.0049	1	0.0030	1	0.3163	1	9	773
16	manufacture of tobacco prod	0.0383	0.0414	0.0061	0.0336	0.0324	1	0.0081	0	-0.0239	0	20	56
17	manufacture of textiles	0.0378	0.0225	0.0000	0.0003	0.0378	1	0.0222	1	0.0841	1	3	675
18	man of wear appar; dress, dye of fur	0.0038	0.0051	0.0000	0.0002	0.0038	1	0.0049	1	0.1576	1	20	716
19	tan and dress of leath, man of lug	0.0226	0.0246	0.0000	0.0003	0.0226	1	0.0243	1	0.0831	1	13	538
20	man of wood and wood prod	0.0045	0.0050	0.0000	0.0001	0.0045	1	0.0049	1	0.2180	1	9	772
21	man of paper and paper prod	0.0032	0.0049	0.0002	0.0016	0.0030	1	0.0034	1	0.0463	0	13	436
22	publ, print and reprod of rec med	0.0125	0.0268	0.0000	0.0008	0.0125	1	0.0260	1	0.0330	1	8	697
23	man of coke, ref petrol prod, nucl	0.0060	0.0556	0.0012	0.0232	0.0048	1	0.0332	1	0.2037	1	19	258
24	man of chem and chem prod	0.0084	0.0345	0.0001	0.0019	0.0083	1	0.0327	1	0.2895	1	14	506
25	man. of rubber and plastic prod	0.0018	0.0028	0.0001	0.0008	0.0017	1	0.0020	1	0.0805	1	10	577
26	man of other non-met min prod	0.0055	0.0094	0.0000	0.0005	0.0055	1	0.0089	1	0.1975	1	10	761
27	man of basic metals	0.0039	0.0126	0.0002	0.0056	0.0036	1	0.0071	1	0.1574	1	8	389
28	man of fabric met prod, exc machin	0.0008	0.0014	0.0000	0.0001	0.0008	1	0.0014	1	0.2066	1	18	764
29	man of machin and equip n.e.c	0.0007	0.0021	0.0000	0.0005	0.0007	1	0.0016	1	0.1555	1	14	683
30	man of off, account, comput mach	0.0163	0.1015	0.0017	0.0605	0.0147	1	0.0436	1	0.0086	0	11	133
31	man of elect mach and appar n.e.c.	0.0029	0.0104	0.0001	0.0024	0.0028	1	0.0080	1	0.1327	1	10	585
32	man of radio, TV, commun equip	0.0036	0.0307	0.0001	0.0061	0.0035	1	0.0247	1	0.2595	1	22	581
33	man of med, prec and opt instr, watch	0.0034	0.0108	0.0000	0.0013	0.0034	1	0.0095	1	0.1750	1	25	682
34	man of mot veh, trail and semi-trail	0.0108	0.0837	0.0004	0.0161	0.0104	1	0.0687	1	0.3103	1	21	310
35	man of other transport equipment	0.0098	0.0235	0.0002	0.0086	0.0096	1	0.0149	1	0.2010	1	23	349
36	man of furniture, manufact n.e.c.	0.0070	0.0099	0.0000	0.0002	0.0070	1	0.0097	1	0.0775	1	4	740
37	recycling	0.0034	0.0040	0.0005	0.0022	0.0029	1	0.0019	1	0.2263	1	17	382

Table 14: Provinces 2-digits all plants

NACE	Industry	G_{UW}	G_{EG}	$1/M$	H	$\hat{\gamma}_{UW}$	$Sign$	$\hat{\gamma}_{EG}$	$Sign$	I	$Sign$	Outl	Units
15	man of food prod and beverages	0.0116	0.0069	0.0000	0.0003	0.0116	1	0.0066	1	0.4733	1	4	91
16	manufacture of tobacco prod	0.0802	0.0632	0.0061	0.0336	0.0746	1	0.0307	1	0.0417	0	6	38
17	manufacture of textiles	0.0381	0.0257	0.0000	0.0003	0.0380	1	0.0254	1	-0.0684	0	3	92
18	man of wear appar dress, dye of fur	0.0079	0.0094	0.0000	0.0002	0.0079	1	0.0092	1	0.1016	0	6	89
19	tan and dress of leath, man of lug	0.0398	0.0404	0.0000	0.0003	0.0398	1	0.0401	1	0.2130	1	4	91
20	man of wood and wood prod	0.0085	0.0085	0.0000	0.0001	0.0084	1	0.0084	1	0.3071	1	4	91
21	man of paper and paper prod	0.0043	0.0058	0.0002	0.0016	0.0041	1	0.0043	1	0.3294	1	4	91
22	publ, print and reprod of rec med	0.0131	0.0256	0.0000	0.0008	0.0130	1	0.0248	1	0.0781	0	4	91
23	man of coke, ref petrol prod, nucl	0.0129	0.0611	0.0012	0.0232	0.0117	1	0.0389	1	0.3733	1	7	88
24	man of chem and chem prod	0.0115	0.0403	0.0001	0.0019	0.0114	1	0.0385	1	0.2295	1	5	90
25	man. of rubber and plastic prod	0.0039	0.0052	0.0001	0.0008	0.0039	1	0.0044	1	-0.0247	0	3	92
26	man of other non-met min prod	0.0104	0.0130	0.0000	0.0005	0.0103	1	0.0125	1	0.2913	1	4	91
27	man of basic metals	0.0129	0.0211	0.0002	0.0056	0.0127	1	0.0156	1	0.2373	1	6	87
28	man of fabric met prod, exc machin	0.0014	0.0024	0.0000	0.0001	0.0014	1	0.0023	1	0.3313	1	3	92
29	man of machin and equip n.e.c	0.0012	0.0034	0.0000	0.0005	0.0012	1	0.0029	1	0.1030	0	7	88
30	man of off, account, comput mach	0.0182	0.0909	0.0017	0.0605	0.0165	1	0.0323	1	0.0447	0	2	69
31	man of elect mach and appar n.e.c.	0.0046	0.0146	0.0001	0.0024	0.0045	1	0.0123	1	0.0460	0	4	91
32	man of radio, TV, commun equip	0.0054	0.0336	0.0001	0.0061	0.0053	1	0.0276	1	0.1665	1	5	90
33	man of med, prec and opt instr, watch	0.0048	0.0184	0.0000	0.0013	0.0048	1	0.0171	1	0.1574	1	4	91
34	man of mot veh, trail and semi-trail	0.0219	0.1034	0.0004	0.0161	0.0215	1	0.0887	1	0.3215	1	4	87
35	man of other transport equipment	0.0174	0.0324	0.0002	0.0086	0.0172	1	0.0240	1	0.1677	1	6	83
36	man of furniture, manufact n.e.c.	0.0053	0.0096	0.0000	0.0002	0.0053	1	0.0094	1	0.1241	0	4	91
37	recycling	0.0062	0.0065	0.0005	0.0022	0.0057	1	0.0043	1	0.1355	0	9	86

Table 15: LLS 2-digits big plants

NACE	Industry	G_{UW}	G_{EG}	$1/M$	H	$\hat{\gamma}_{UW}$	$Sign$	$\hat{\gamma}_{EG}$	$Sign$	I	$Sign$	Outl	Units
15	man of food prod and beverages	0.0049	0.0061	0.0004	0.0016	0.0046	1	0.0045	1	0.1557	1	14	768
16	manufacture of tobacco prod	0.0289	0.0454	0.0115	0.0359	0.0176	1	0.0098	0	-0.2530	0	21	56
17	manufacture of textiles	0.0271	0.0230	0.0003	0.0008	0.0268	1	0.0223	1	0.0597	1	7	671
18	man of wear appar dress, dye of fur	0.0107	0.0090	0.0003	0.0009	0.0104	1	0.0081	1	0.2176	1	24	712
19	tan and dress of leath, man of lug	0.0302	0.0294	0.0004	0.0010	0.0299	1	0.0284	1	0.0464	0	18	533
20	man of wood and wood prod	0.0113	0.0129	0.0010	0.0019	0.0103	1	0.0110	1	0.1134	1	16	765
21	man of paper and paper prod	0.0049	0.0081	0.0011	0.0029	0.0038	1	0.0053	1	0.0687	1	7	442
22	publ, print and reprod of rec med	0.0247	0.0490	0.0007	0.0037	0.0240	1	0.0454	1	0.0004	0	13	692
23	man of coke, ref petrol prod, nucl	0.0124	0.0765	0.0056	0.0329	0.0068	1	0.0451	1	0.0649	0	23	254
24	man of chem and chem prod	0.0226	0.0374	0.0006	0.0025	0.0219	1	0.0350	1	0.1845	1	11	509
25	man. of rubber and plastic prod	0.0031	0.0033	0.0004	0.0018	0.0027	1	0.0015	1	0.0355	0	13	574
26	man of other non-met min prod	0.0125	0.0202	0.0004	0.0014	0.0121	1	0.0188	1	0.0650	1	12	759
27	man of basic metals	0.0061	0.0160	0.0009	0.0073	0.0051	1	0.0088	1	0.1242	1	8	389
28	man of fabric met prod, exc machin	0.0028	0.0036	0.0002	0.0005	0.0026	1	0.0032	1	0.0441	1	10	772
29	man of machin and equip n.e.c	0.0026	0.0029	0.0002	0.0010	0.0024	1	0.0019	1	0.1924	1	16	681
30	man of off, account, comput mach	0.0330	0.1294	0.0093	0.0908	0.0239	1	0.0424	1	-0.0409	0	13	130
31	man of elect mach and appar n.e.c.	0.0038	0.0127	0.0006	0.0051	0.0032	1	0.0076	1	0.0687	1	11	584
32	man of radio, TV, commun equip	0.0112	0.0411	0.0016	0.0108	0.0096	1	0.0306	1	0.0151	0	16	587
33	man of med, prec and opt instr, watch	0.0129	0.0204	0.0011	0.0052	0.0118	1	0.0153	1	0.0288	0	11	696
34	man of mot veh, trail and semi-trail	0.0188	0.0840	0.0013	0.0178	0.0176	1	0.0674	1	0.1399	1	15	316
35	man of other transport equipment	0.0146	0.0303	0.0018	0.0119	0.0128	1	0.0186	1	0.0302	0	23	349
36	man of furniture, manufact n.e.c.	0.0164	0.0152	0.0004	0.0009	0.0160	1	0.0143	1	0.1145	1	8	736
37	recycling	0.0178	0.0262	0.0172	0.0228	0.0005	0	0.0036	0	0.0235	0	44	355

Table 16: Provinces 2-digits big plants

NACE	Industry	G_{UW}	G_{EG}	$1/M$	H	$\hat{\gamma}_{UW}$	$Sign$	$\hat{\gamma}_{EG}$	$Sign$	I	$Sign$	Outl	Units
15	man of food prod and beverages	0.0093	0.0095	0.0004	0.0016	0.0090	1	0.0079	1	0.1515	1	5	90
16	manufacture of tobacco prod	0.0556	0.0694	0.0115	0.0359	0.0447	1	0.0347	1	0.0952	0	6	38
17	manufacture of textiles	0.0325	0.0307	0.0003	0.0008	0.0322	1	0.0299	1	0.0552	0	6	89
18	man of wear appar dress, dye of fur	0.0201	0.0159	0.0003	0.0009	0.0198	1	0.0150	1	0.2812	1	5	90
19	tan and dress of leath, man of lug	0.0480	0.0477	0.0004	0.0010	0.0477	1	0.0467	1	0.1909	1	7	88
20	man of wood and wood prod	0.0182	0.0222	0.0010	0.0019	0.0172	1	0.0204	1	0.0231	0	4	91
21	man of paper and paper prod	0.0064	0.0092	0.0011	0.0029	0.0053	1	0.0063	1	0.0517	0	4	91
22	publ, print and reprod of rec med	0.0261	0.0469	0.0007	0.0037	0.0254	1	0.0434	1	0.0505	0	5	90
23	man of coke, ref petrol prod, nucl	0.0174	0.0828	0.0056	0.0329	0.0118	1	0.0516	1	0.1321	0	4	91
24	man of chem and chem prod	0.0328	0.0436	0.0006	0.0025	0.0322	1	0.0412	1	0.2149	1	5	90
25	man. of rubber and plastic prod	0.0042	0.0050	0.0004	0.0018	0.0037	1	0.0032	1	-0.0912	0	5	90
26	man of other non-met min prod	0.0190	0.0280	0.0004	0.0014	0.0186	1	0.0266	1	0.0214	0	4	91
27	man of basic metals	0.0206	0.0246	0.0009	0.0073	0.0197	1	0.0175	1	0.0522	0	4	89
28	man of fabric met prod, exc machin	0.0036	0.0033	0.0002	0.0005	0.0034	1	0.0028	1	-0.0063	0	4	91
29	man of machin and equip n.e.c	0.0048	0.0037	0.0002	0.0010	0.0046	1	0.0028	1	0.1276	0	6	89
30	man of off, account, comput mach	0.0287	0.1074	0.0093	0.0908	0.0196	1	0.0182	0	0.0381	0	4	68
31	man of elect mach and appar n.e.c.	0.0052	0.0173	0.0006	0.0051	0.0046	1	0.0122	1	0.0356	0	5	90
32	man of radio, TV, commun equip	0.0143	0.0449	0.0016	0.0108	0.0127	1	0.0345	1	-0.0289	0	5	90
33	man of med, prec and opt instr, watch	0.0264	0.0364	0.0011	0.0052	0.0253	1	0.0314	1	0.0407	0	5	90
34	man of mot veh, trail and semi-trail	0.0369	0.1022	0.0013	0.0178	0.0356	1	0.0859	1	0.0240	0	7	84
35	man of other transport equipment	0.0210	0.0416	0.0018	0.0119	0.0192	1	0.0300	1	0.1188	0	7	82
36	man of furniture, manufact n.e.c.	0.0192	0.0182	0.0004	0.0009	0.0188	1	0.0174	1	0.1237	0	4	91
37	recycling	0.0203	0.0319	0.0172	0.0228	0.0031	0	0.0093	0	-0.0094	0	6	89

Table 17: LLS 2-digits small plants

NACE	Industry	G_{UW}	G_{EG}	$1/M$	H	$\hat{\gamma}_{UW}$	$Sign$	$\hat{\gamma}_{EG}$	$Sign$	I	$Sign$	Outl	Units
15	man of food prod and beverages	0.0032	0.0028	0.0000	0.0000	0.0032	1	0.0028	1	0.4285	1	12	770
16	manufacture of tobacco prod	0.0707	0.0526	0.0128	0.0240	0.0586	1	0.0293	1	0.0815	0	9	61
17	manufacture of textiles	0.0358	0.0331	0.0000	0.0001	0.0357	1	0.0330	1	0.1253	1	2	676
18	man of wear appar dress, dye of fur	0.0021	0.0031	0.0000	0.0000	0.0021	1	0.0031	1	0.1938	1	17	719
19	tan and dress of leath, man of lug	0.0193	0.0234	0.0000	0.0001	0.0193	1	0.0233	1	0.1694	1	11	540
20	man of wood and wood prod	0.0025	0.0025	0.0000	0.0000	0.0025	1	0.0024	1	0.3220	1	14	767
21	man of paper and paper prod	0.0049	0.0055	0.0002	0.0004	0.0046	1	0.0050	1	0.0984	1	18	431
22	publ, print and reprod of rec med	0.0148	0.0151	0.0000	0.0001	0.0147	1	0.0150	1	0.1858	1	7	698
23	man of coke, ref petrol prod, nucl	0.0051	0.0069	0.0015	0.0027	0.0036	1	0.0042	1	0.2044	1	28	253
24	man of chem and chem prod	0.0082	0.0090	0.0002	0.0003	0.0081	1	0.0087	1	0.2185	1	15	505
25	man. of rubber and plastic prod	0.0033	0.0036	0.0001	0.0001	0.0032	1	0.0034	1	0.1551	1	9	578
26	man of other non-met min prod	0.0035	0.0035	0.0000	0.0001	0.0035	1	0.0034	1	0.2459	1	14	757
27	man of basic metals	0.0054	0.0060	0.0003	0.0006	0.0050	1	0.0054	1	0.1853	1	12	385
28	man of fabric met prod, exc machin	0.0009	0.0015	0.0000	0.0000	0.0009	1	0.0014	1	0.2280	1	14	768
29	man of machin and equip n.e.c	0.0014	0.0021	0.0000	0.0001	0.0014	1	0.0021	1	0.1357	1	20	677
30	man of off, account, comput mach	0.0223	0.0201	0.0021	0.0030	0.0203	1	0.0172	1	0.0213	0	8	135
31	man of elect mach and appar n.e.c.	0.0051	0.0056	0.0001	0.0001	0.0050	1	0.0055	1	0.1799	1	12	583
32	man of radio, TV, commun equip	0.0046	0.0075	0.0001	0.0003	0.0044	1	0.0073	1	0.2358	1	20	583
33	man of med, prec and opt instr, watch	0.0048	0.0065	0.0000	0.0001	0.0048	1	0.0064	1	0.1205	1	16	691
34	man of mot veh, trail and semi-trail	0.0108	0.0121	0.0007	0.0012	0.0102	1	0.0109	1	0.1392	1	13	318
35	man of other transport equipment	0.0086	0.0109	0.0003	0.0005	0.0083	1	0.0104	1	0.1866	1	18	354
36	man of furniture, manufact n.e.c.	0.0053	0.0077	0.0000	0.0000	0.0053	1	0.0077	1	0.0889	1	12	732
37	recycling	0.0033	0.0031	0.0005	0.0011	0.0028	1	0.0021	1	0.2114	1	18	381

Table 18: Provinces 2-digits small plants

NACE	Industry	G_{UW}	G_{EG}	$1/M$	H	$\hat{\gamma}_{UW}$	$Sign$	$\hat{\gamma}_{EG}$	$Sign$	I	$Sign$	Outl	Units
15	man of food prod and beverages	0.0089	0.0070	0.0000	0.0000	0.0089	1	0.0070	1	0.5296	1	4	91
16	manufacture of tobacco prod	0.1490	0.0881	0.0128	0.0240	0.1379	1	0.0657	1	0.1237	0	8	36
17	manufacture of textiles	0.0336	0.0317	0.0000	0.0001	0.0336	1	0.0316	1	-0.0250	0	2	93
18	man of wear appar dress, dye of fur	0.0046	0.0062	0.0000	0.0000	0.0046	1	0.0061	1	0.2012	1	8	87
19	tan and dress of leath, man of lug	0.0331	0.0372	0.0000	0.0001	0.0330	1	0.0372	1	0.1856	1	4	91
20	man of wood and wood prod	0.0058	0.0052	0.0000	0.0000	0.0058	1	0.0051	1	0.4156	1	3	92
21	man of paper and paper prod	0.0058	0.0064	0.0002	0.0004	0.0056	1	0.0060	1	0.0960	0	5	90
22	publ, print and reprod of rec med	0.0146	0.0143	0.0000	0.0001	0.0145	1	0.0143	1	0.0914	1	4	91
23	man of coke, ref petrol prod, nucl	0.0116	0.0134	0.0015	0.0027	0.0101	1	0.0107	1	0.4591	1	5	90
24	man of chem and chem prod	0.0103	0.0137	0.0002	0.0003	0.0102	1	0.0133	1	0.2762	1	6	89
25	man. of rubber and plastic prod	0.0071	0.0081	0.0001	0.0001	0.0070	1	0.0080	1	0.0707	0	3	92
26	man of other non-met min prod	0.0071	0.0056	0.0000	0.0001	0.0071	1	0.0055	1	0.3362	1	4	91
27	man of basic metals	0.0134	0.0188	0.0003	0.0006	0.0131	1	0.0182	1	0.2492	1	4	89
28	man of fabric met prod, exc machin	0.0021	0.0036	0.0000	0.0000	0.0021	1	0.0036	1	0.2401	1	6	89
29	man of machin and equip n.e.c	0.0023	0.0036	0.0000	0.0001	0.0022	1	0.0035	1	0.2111	1	6	89
30	man of off, account, comput mach	0.0243	0.0242	0.0021	0.0030	0.0222	1	0.0213	1	0.0839	0	2	69
31	man of elect mach and appar n.e.c.	0.0078	0.0087	0.0001	0.0001	0.0077	1	0.0086	1	0.1559	1	3	92
32	man of radio, TV, commun equip	0.0063	0.0090	0.0001	0.0003	0.0062	1	0.0087	1	0.2209	1	9	86
33	man of med, prec and opt instr, watch	0.0060	0.0088	0.0000	0.0001	0.0059	1	0.0087	1	0.2049	1	9	86
34	man of mot veh, trail and semi-trail	0.0207	0.0218	0.0007	0.0012	0.0200	1	0.0206	1	0.2409	1	5	86
35	man of other transport equipment	0.0152	0.0152	0.0003	0.0005	0.0149	1	0.0146	1	0.1953	1	4	85
36	man of furniture, manufact n.e.c.	0.0042	0.0072	0.0000	0.0000	0.0042	1	0.0071	1	0.0782	0	7	88
37	recycling	0.0060	0.0049	0.0005	0.0011	0.0055	1	0.0038	1	0.0844	0	7	88