Geographic Concentration of Economic Activity: An aggregate index *

Olga Alonso-Villar# and Coral del Río

Universidade de Vigo

Abstract

This paper first proposes a measure, the mutual information index derived from the information theory, to quantify overall concentration from an axiomatic perspective. The analysis reveals that this overall concentration measure can be written as the weighed sum of the Theil index for each sector of the economy (partial concentration). Next, the generalized entropy family of concentration indexes is characterized in terms of basic axioms borrowed from the literature on income distribution and occupational segregation. Finally, these measures are used to analyze the spatial patterns of manufacturing industries in Spain along the last three decades, paying special attention to their technological intensity.

JEL Classification: R12; D63

Keywords: Geographic concentration; Inequality measures; Segregation measures

---

* Financial support from the *Xunta de Galicia* (INCITE08PXIB300005PR) and from FEDER is gratefully acknowledged.

Correspondence address: Universidade de Vigo; Facultade de CC. Económicas; Departamento de Economía Aplicada; Campus Lagoas-Marcosende s/n; 36310 Vigo; Spain. Tel.: +34 986812507; fax: +34 986812401; e-mail: ovillar@uvigo.es

1. Introduction

In recent years, the study of production location patterns has received increasing interest in the field, both empirically and theoretically. This flourishing interest is in part motivated by the general concern with the effects of economic integration processes on industrial localization, especially in Europe where the creation of the Single Market has stimulated the debate (Amiti, 1999; Haaland et al., 1999; Brülhart, 2001; Aiginger and Pfaffermayr, 2004; and Resmini, 2007, inter alia).[1]

Among the spatial concentration measures existing in the literature, those borrowed from the literature on income inequality are some of the most widely used.[2] In this regard, the Gini index has been traditionally used for analyzing the spatial location patterns of manufacturing industries (Krugman, 1991; Amiti, 1999; Brülhart, 2001; Suedekum, 2006, inter alia). More recently, the generalized entropy family of indexes has been used as well because of its advantages in terms of decomposability (Brülhart and Traeger, 2005; Brakman et al., 2005; Pérez-Ximénez and Sanz-Gracia, 2007; Cutrini, 2009a).

In quantifying the spatial concentration of a given sector, most measures used in empirical analysis have followed a *relative* notion, so that the spatial distribution of the sector is compared with that of the whole set of sectors (Amiti, 1999; Brülhart, 2001; Brülhart and Traeger, 2005).[3] If the economic activity is measured in terms of employment, as is traditionally done, and the focus is on manufacturing industries, the distribution of overall manufacturing employment is usually considered the distribution of reference against which to compare that of any single sector. Thus, no concentration exists in a given sector so long as its employment distribution among locations coincides with that of overall manufacturing employment.

[1] From a theoretical perspective, the literature of the new economic geography has contributed extensively to this debate. A review of this literature can be seen in Fujita et al. (2000), Neary (2001), and Ottaviano and Thisse (2004), among others.

[2] Other concentration measures proposed in the literature are formally derived from location models (Ellison and Glaeser, 1997; Maurel and Sédillot, 1999; and Guimarães et al., 2007). There are also distance-based measures related to the literature on spatial statistics (Marcon and Puech, 2003; and Duranton and Overman, 2005).

[3] For other perspectives, such as *topographic* and *absolute* concentration, see Aiginger and Davis (2004); Brülhart and Traeger (2005), Brakman et al. (2005), and Mori et al. (2005).

By following this approach, many studies have calculated the concentration level of each manufacturing sector in different economies. However, there has been no discussion on how to aggregate this information in order to calculate concentration for the whole manufacturing industry. In other words, no relationship is formally established in the literature between the concentration level of each sector (which can be labeled as partial concentration) and overall concentration.

Certainly, the *relative* version of one of the members of the generalized entropy family of concentration indexes, Ψ_2, can be additively decomposed by subsectors, so that it is possible to determine the contribution of each subsector to the concentration of the sector when considering the overall manufacturing industry as the distribution of reference. However, by using this approach it would not be possible to determine the contribution of each manufacturing sector to the overall manufacturing concentration since this aggregate level cannot be determined according to the same criterion. The reason is that the distribution of reference would be the same as the one to be analyzed, so that the *relative* concentration of the manufacturing industry would be necessarily equal to zero. For this reason, in order to determine the concentration level of the overall manufacturing industry some studies compare the employment distribution of this industry across locations with the employment distribution of the whole economy (Brülhart and Traeger, 2005). In other words, in measuring the concentration of the manufacturing industry, the benchmark considered is outside that industry. Alternatively, this paper proposes to quantify the spatial concentration of the whole manufacturing industry by using an aggregate measure that is compatible with the concentration measurement of single manufacturing sectors. In other words, in order to measure the spatial concentration of the overall manufacturing industry, it is not necessary to use an external benchmark.

Therefore, two different perspectives can be used to measure the overall concentration of the manufacturing industry. One measurement emphasizes the relationship that exists between the manufacturing distribution and the distribution of total economic activity by quantifying the discrepancies between them. The other measurement calls attention, instead, to what happens in the internal distribution of the manufacturing industry by taking into account the spatial disparities among manufacturing sectors. Both perspectives seem reasonable and complementary, and this paper focuses on the latter.

As a consequence of all the above, this approach permits not only a calculation of the concentration of the manufacturing industry, but also that of the whole economy, which cannot be obtained within the traditional approach. In particular, it brings the possibility of determining the contribution of each large sector of the economy (industry, services, construction, and agriculture-fishing) to the concentration of the whole economic activity.

The first aim of this paper is to propose, and analyze axiomatically, a measure derived from the information theory to quantify the overall concentration of the manufacturing industry without using a benchmark outside that industry. In doing so, we use the mutual information index, which has been recently proposed to quantify overall segregation in a framework of multiple population subgroups. Frankel and Volij (2008) and Mora and Ruiz-Castillo (2008) use this measure to quantify school and residential segregation by race in the US, respectively, because of its good axiomatic properties, in a particular because of its decomposability (Frankel and Volij, 2007; Mora and Ruiz-Castillo, 2009).[4] As discussed later in this paper, this index seems also suitable to be used in our context given the parallelism that exists between the measurement of segregation across organizational units (schools, cities, occupations, etc.) and the measurement of spatial concentration.

This overall concentration index can be expressed as the weighed average of the concentration level of each sector (labeled here as partial concentration) measured according to one of the members of the generalized entropy family of concentration indexes: Ψ_1, which allows one to determine how much each sector contributes to overall manufacturing concentration by using only the information of the manufacturing industry. Even though the properties of the generalized entropy family of inequality indexes are well known (Shorrocks 1984; Foster, 1985; and Cowell, 2000), as far as we know, the corresponding family of concentration indexes has not yet been axiomatically explored in a location context. For this reason, this paper also characterizes axiomatically the *relative* version of the partial concentration measures derived from the generalized entropy family (denoted by Ψ_α, where α is a sensitivity parameter). Some of these axioms are adapted from the income inequality measurement while others are

[4] In fact, Frankel and Volij (2008) can be considered as the first attempt to characterize axiomatically an overall segregation measure in a multigroup context; since, so far, the literature on segregation has only done that in a binary context.

borrowed, instead, from the literature on occupational segregation (Hutchens, 2004; Alonso-Villar and Del Río, 2007). As a consequence of all the above, this paper brings theoretical support to empirical studies that use weighted averages of entropy indexes in order to measure overall concentration (Aiginger and Davies, 2004; Cutrini, 2009a, b).[5]

The second aim of this paper is to use these measures to analyze the spatial patterns of manufacturing industries in Spain during the last three decades, paying special attention to differences among sectors according to their technological intensity. In order to check the robustness of our results, other measures borrowed from the literature of segregation are adapted to measure overall and partial concentration (Reardon and Firebaugh, 2002; Silber, 1992; and Alonso-Villar and Del Río, 2007).

The paper is structured as follows. Section 2 characterizes, in terms of basic axioms, the generalized entropy family of indexes used in the literature to measure the geographic concentration of a sector (partial concentration). It also proposes additional partial concentration measures that are derived from the segregation field. Section 3 introduces several aggregate concentration indexes related with the above partial measures. In particular, the mutual information index is presented and its basic properties shown. These partial and aggregate measures are used in Section 4 to analyze the manufacturing industry in Spain along its whole democratic period.[6] Finally, Section 5 presents the main conclusions.

2. The spatial concentration of an industry

Consider an economy with $L>1$ locations among which the economic activity, denoted by *T*, is distributed according to distribution $t \equiv (t_1, t_2, ..., t_L)$, where $T = \sum_l t_l$. For the sake of simplicity and without loss of generality, in what follows, we assume that the economic activity is measured in terms of employment. Thus, $t_j > 0$ represents the number of workers in location l $(l = 1, ..., L)$. Let us denote by $x^s \equiv (x_1^s, x_2^s, ..., x_L^s)$ the distribution of sector s across locations $(s = 1, ..., S)$, where x_l^s represents the number

[5] Aiginger and Davies (2004) use weighted averages of absolute entropy indexes, and unweighted averages of relative entropy indexes in order to analyze concentration and specialization in the European Union. Our analysis suggests that in the relative case a weighted average should be used instead.

[6] The death of the dictator F. Franco took placed at the end of 1975, and the Spanish Constitution was signed in 1978.

of workers of sector s in location l. Therefore, the total number of workers in location l is $t_l = \sum_s x_l^s$, while the total number of workers in sector s is $X^s = \sum_l x_l^s$.

In this paper, an index of partial geographic concentration is a function $I_c : D \to \mathbb{R}$, where $D = \bigcup_{L>1} \left\{ \left(x^s ; t \right) \in \mathbb{R}^L{}_+ \times \mathbb{R}^L{}_{++} : x_l^s \le t_l \forall l \right\}$, such that $I_c(x^s;t)$ represents the concentration level of sector s, which is distributed across locations according to x^s, when comparing it with the distribution of reference t.

2.1 Partial measures: The generalized entropy family of indexes

In order to measure the spatial concentration level of an industry according to a *relative* notion, the generalized entropy family of indexes can be written as:

$$\Psi_\alpha(x^s;t) = \begin{cases} \dfrac{1}{\alpha(\alpha-1)} \displaystyle\sum_l \frac{t_l}{T} \left[\left(\frac{x_l^s / X^s}{t_l / T} \right)^\alpha - 1 \right] & \text{if } \alpha \neq 0,1 \\ \displaystyle\sum_l \frac{x_l^s}{X^s} \ln \left(\frac{x_l^s / X^s}{t_l / T} \right) & \text{if } \alpha = 1 \end{cases},$$

where α is a sensitivity parameter.[7] If sector s is distributed across locations in the same way as aggregate employment, i.e., if $x_l^s / X^s = t_l / T \quad \forall l$, any index of this class is equal to zero. An advantage of these concentration measures is that they are additively decomposable, which is very helpful for empirical analysis (Brülhart and Traeger, 2005).

Even though -- in the literature on income distribution -- this family of indexes has been characterized in terms of basic axioms, to our knowledge such a characterization does not exist in the field of spatial concentration. In order to better understand the axioms

---

[7] Note that according to the information theory, index $\Psi_1(x^s;t)$ can be interpreted as the relative entropy or *Kullback Leibler distance* between distributions $\left(\frac{x_1^s}{X^s}, ..., \frac{x_L^s}{X^s} \right)$ and $\left(\frac{t_1}{T}, ..., \frac{t_L}{T} \right)$. A *topographic* version of this index has been used by Mori et al. (2005) to analyze industrial localization in Japan. According to their *topographic* perspective, the distribution of reference is not that of aggregate employment, but that represented by the economic area of each location unit. Another difference with respect to their index is that in this paper, Ψ_1 measures the spatial concentration of employment, while the aforementioned paper focuses on the concentration of establishments.

listed below, first of all, we formally establish the relationship between the measurement of spatial concentration of economic activity and the measurement of income inequality. For that purpose, a hypothetical "income" distribution derived from vector $\left(x^s;t\right)$ is obtained. In doing so, in each location, the variable of study (employment in the sector of study) is equally split among all individuals (both those working in the sector of study and those in the remaining sectors). This per capita employment level, $\frac{x_l^s}{t_l}$, represents the employment in the sector of study that corresponds to each individual in location l, and it plays the role of individual "income". Namely, the fictitious "income" distribution is constructed as follows: there are t_1 persons with an individual "income" of $\frac{x_1^s}{t_1}$, t_2 persons with an individual "income" of $\frac{x_2^s}{t_2}$, and so on. Therefore, we have built "income" distribution $y \equiv (\underbrace{\frac{x_1^s}{t_1},\ldots,\frac{x_1^s}{t_1}}_{t_1 \text{ individuals}},\ldots,\underbrace{\frac{x_L^s}{t_L},\ldots,\frac{x_L^s}{t_L}}_{t_L \text{ individuals}})$ in a world of $T=\sum_l t_l$ individuals where total "income" is $X=\sum_l t_l \frac{x_l^s}{t_l}$.

Suppose, for example, that we want to measure the geographic concentration of the chemical sector by comparing its employment distribution across regions with that of manufacturing employment. Consider that the economy has three locations and that the employment distribution of the chemical industry among them is $\left(3,2,5\right)$, while the distribution of manufacturing workers is $\left(30,10,30\right)$. In other words, $\left(x^s;t\right)=\left(3,2,5;30,10,30\right)$. Therefore, our fictitious "income" distribution would be one with 70 people having a total income of 10 units: there are 30 people with an individual "income" of 0.1, 10 people with an individual "income" of 0.2, and 30 people with an individual "income" of 0.6 i.e., the "income" distribution is equal to

$$y \equiv \left(\underbrace{\frac{3}{30},\ldots,\frac{3}{30}}_{30},\underbrace{\frac{2}{10},\ldots,\frac{2}{10}}_{10},\underbrace{\frac{5}{30},\ldots,\frac{5}{30}}_{30}\right).$$

The parallelism between employment distribution $\left(x^s;t\right)$ and hypothetical "income" distribution y will be helpful for understanding the axiomatic framework presented in what follows -- where some basic axioms, borrowed from the literature on income distribution and occupational segregation -- are adapted to analyze spatial concentration measures: [8]

1) *Symmetry in locations* (the (partial) concentration index is unaffected by the order in which locations are enumerated);

2) *Movement between locations* (when a region with a lower employment level in the sector of study than another, but with the same aggregate employment, loses employment in the sector in favor of the other location, the concentration of the sector must increase);

3) *Scale invariance* (the concentration index should not change when the employment level of the aggregate distribution and/or that of the sector under consideration vary, so long as the weight that each location represents in distributions t and x^s remains unaltered);

4) *Insensitivity to proportional divisions of locations* (subdividing a location into several units of equal size, both in terms of aggregate employment and in terms of employment in the sector of study, does not affect the concentration level of the sector);

5) *Aggregation* (when partitioning locations into two mutually exclusive classes, the concentration level of the sector of study can be written as a function of the concentration level of the sector in each class of locations, the employment level in each class, and the employment share of the sector in each class of locations).

Axioms 1-3 are related to the three basic axioms that are usually required in the literature on income distribution (symmetry, the Pigou-Dalton principle of transfers, and scale invariance), while axiom 5 is a very helpful additional property that is related with the decomposition of inequality indexes by population subgroups. Axiom 4 is borrowed, instead, from the literature on occupational segregation (Hutchens, 2004). These five

[8] For a more technical definition of these axioms, see Appendix A.

axioms, altogether, completely characterize the generalized entropy family of indexes employed in concentration analyses from a *relative* perspective, as shown in the next proposition.

Proposition. *Let* I_c *be a continuous concentration index that takes a zero value when the distribution of the sector of study among locations coincides with the distribution of reference* (i.e., when $\frac{x_l^s}{X^s}=\frac{t_l}{T}$). *Then,* I_c *is a concentration index satisfying axioms 1-5 if and only if it can be written as an increasing monotonic transformation of index*

$$\Psi_\alpha(x^s;t)=\begin{cases}\dfrac{1}{\alpha(\alpha-1)}\sum_l \dfrac{t_l}{T}\left[\left(\dfrac{x_l^s/X^s}{t_l/T}\right)^\alpha-1\right] & \text{if } \alpha\neq 0,1\\ \sum_l \dfrac{x_l^s}{X^s}\ln\left(\dfrac{x_l^s/X^s}{t_l/T}\right) & \text{if } \alpha=1\end{cases},$$

where α *is a sensitivity parameter.*[9]

Proof: See Appendix.

2.2 Other partial measures

One of the most popular concentration indexes derived from the literature on income distribution is the locational Gini coefficient, which satisfies axioms 1-4. For a given sector s, this index can be written as the sum of the differences between the employment shares of the sector in each pair of locations weighted by their demographic weights, and divided by twice the employment share of the sector in the whole economy:

$$G^s=\frac{\sum_{l,l'}\dfrac{t_l}{T}\dfrac{t_{l'}}{T}\left|\dfrac{x_l^s}{t_l}-\dfrac{x_{l'}^s}{t_{l'}}\right|}{2\dfrac{X^s}{T}}.$$

[9] If we had considered concentration indexes defined on the space of employment distributions $(x^s;t)$ where all components of vector x^s were strictly positive, rather than positive, then another index would have appeared: $\Psi_\alpha(x^s;t)=\sum_l \frac{t_l}{T}\ln\left(\frac{t_l/T}{x_l^s/X^s}\right)$ if $\alpha=0$.

There are several interpretations of this measure. On one hand, the locational Gini coefficient is equal to twice the area between the corresponding employment Lorenz curve and the 45º line. On the other hand, following the segregation approach of Flückiger and Silber (1999), we can also think of this measurement as the degree of conformity between “a priori” and “a posteriori” employment shares. Thus, no concentration exists in the sector so long as the employment share in each location, $\frac{x_l^s}{X^s}$ (the “a posteriori” share), coincides with the employment share that the respective location represents, $\frac{t_j}{T}$ (the “a priori” share).

Given the parallelism that exists between the spatial concentration of a sector across locations and the segregation of a population group across organizational units, the popular index of dissimilarity proposed by Duncan and Duncan (1955) can also be conveniently adapted to measure the *relative* concentration of a sector, so that it can be written as follows:[10]

$$D^s = \frac{1}{2}\sum_l \left| \frac{x_l^s}{X^s} - \frac{t_l}{T} \right|.$$

This index is also related to the employment Lorenz curve of the sector since it equals the maximum vertical distance between the curve and the 45º line. It is easy to see that this index satisfies the aforementioned axioms 1, 3 and 4, but not axiom 2.

3. Overall concentration measures

In this section, we first propose an overall concentration index derived from the literature on information theory, the mutual information index, and present some of its properties. This index is later shown to be the weighted sum of index Ψ_1 for each of the mutual exclusive sectors in which the economy can be partitioned. Next, we propose two more aggregate concentration measures derived from the segregation literature in order to use them in the empirical section to analyze the robustness of our results. One is the weighted average of the locational Gini coefficient G^s for each sector, and the other is the weighted average of index D^s.

[10] In an occupational segregation context a similar index has been proposed by Moir and Selby Smith (1979).

3.1 The mutual information index

The mutual information index is "a measure of the amount of information that one random variable contains about another random variable" since it quantifies "the reduction in the uncertainty of one random variable due to the knowledge of the other" (Cover and Thomas, 1991, p. 18). Given two random distributions, V and Z, the mutual information index, $M(V;Z)$, can be written as:[11]

$$M(V;Z) = \underbrace{H(V)}_{entropy} - \underbrace{H(V|Z)}_{conditional\ entropy} ,$$

where $H(V) = -\sum_{v} p(v) \ln p(v)$, $H(V|Z) = -\sum_{z} p(z) \sum_{v} p(v|z) \ln p(v|z)$, $p(v)$ and $p(z)$ denote the probability mass functions of V and Z, respectively, and $p(v|z)$ is the probability distribution of conditional distribution $V|Z = z$.[12]

The mutual information index has been recently proposed, and characterized, by Frankel and Volij (2007) to analyze overall school segregation in a multiracial context. According to an evenness perception of segregation, which is the most popular dimension of this phenomenon, overall segregation exists so long as the population subgroups in which the economy can be partitioned (blacks/whites/Hispanics, for example) are not similarly distributed among organizational units. The parallelism between the measurement of segregation across organizational units and the measurement of geographic concentration becomes evident. For example, in the case of school segregation by race, the former involves comparisons among the distributions of racial groups across schools, while the latter requires comparing distributions of industries across locations.

Given the good properties of this index in a segregation context (Frankel and Volij, 2007, 2008; Mora and Ruiz-Castillo, 2009), and the parallelism that exists between segregation and concentration measurements, this measure seems also reasonable to quantify overall concentration.

[11] Note that the mutual information index is a symmetric concept.

[12] In principle, the logarithm could be in any base, but for convenience we use natural logarithms.

In our case, $v \equiv l$ represents location units and $z \equiv s$ sectors, so that the probability distribution of variable V is $\left(\frac{t_1}{T},...,\frac{t_L}{T}\right)$, the probability distribution of Z is $\left(\frac{X^1}{T},...,\frac{X^S}{T}\right)$, and the conditional distribution $V|Z=z$ is $\left(\frac{x_1^s}{X^s},...,\frac{x_L^s}{X^s}\right)$. Therefore, in our case, the mutual information index can be written as[13]

$$M(V;Z) = -\sum_l \frac{t_l}{T}\ln\left(\frac{t_l}{T}\right) + \sum_s \frac{X^s}{T}\sum_l \frac{x_l^s}{X^s}\ln\left(\frac{x_l^s}{X^s}\right).$$

In what follows, we enumerate the basic axioms satisfied by the mutual information index (shown by Frankel and Volij, 2007) when adapting them to our context. First, it satisfies *continuity* and it is invariant to:

a) Any reordering of the sectors and locations (*symmetry*);

b) Proportional changes in all sectors and locations (*weak scale invariance*);

c) Splitting one location into two if both have the same sectoral structure (*location division property*);[14] and

d) Splitting a sector into two subsectors if both have the same distribution across locations (*group division property*).

In addition,

e) If a location, with its own sectoral structure, is adjoined to two economies having the same total employment and sectoral structures, the ranking between both economies according to the index does not change (*type I independence*).

f) Moreover, the ranking between two economies with the same employment level is the same whether another economy is joined to them by merging all the locations of that economy into a single one or not (*type II independence*).[15]

Finally, note that the mutual information index can also be expressed as

[13] The mutual information index can also be written as $M(Z;V) = -\sum_s \frac{X^s}{T}\ln\left(\frac{X^s}{T}\right) + \sum_l \frac{t_l}{T}\sum_s \frac{x_l^s}{t_l}\ln\left(\frac{x_l^s}{t_l}\right).$

[14] In Frankel and Volij (2008) this property is actually labeled "school division property".

[15] This axiom allows decomposability among locations in a simple way and its motivation concerns *within-group* and *between-group* component differences.

$$M\left(V;Z\right)=\sum_{s}\frac{X^{s}}{T}\Psi_{1}(x^{s};t),$$

since $\sum_{l}\frac{t_l}{T}\ln\left(\frac{t_l}{T}\right)=\sum_{l}\frac{\sum_{s}x_l^s}{T}\ln\left(\frac{t_l}{T}\right)=\sum_{s}\frac{X^s}{T}\sum_{l}\frac{x_l^s}{X^s}\ln\left(\frac{t_l}{T}\right)$. Therefore, the mutual information index can be written as the weighted average, according to demographic weights, of partial concentration index Ψ_1. In other words, it can be thought of as an average concentration measure, and, therefore, by using the above expression it is easy to determine the contribution of each sector to the overall concentration level. Consequently, this paper brings theoretical support to those works that use the weighted average of the Theil index.

An advantage of the mutual information index is that it can be additively decomposed by subgroups. Thus, if the manufacturing industry is partitioned into several mutual exclusive subgroups, it is possible to find out whether aggregate manufacturing concentration is mainly due to the *between-group* component (i.e., to differences among the concentration levels of subgroups) or to the *within-group* component (i.e., to internal differences in each subgroup). Without loss of generality, let us assume that the manufacturing industry is partitioned into two subgroups of industries: G and H. Then,

$$M=M^{B}+M^{W},$$

where:

$$M^{B}=\frac{X^{G}}{T}\sum_{l}\frac{x_l^G}{X^G}\ln\left(\frac{\frac{x_l^G}{X^G}}{\frac{t_l}{T}}\right)+\frac{X^{H}}{T}\sum_{l}\frac{x_l^H}{X^H}\ln\left(\frac{\frac{x_l^H}{X^H}}{\frac{t_l}{T}}\right), \text{ and}$$

$$M^{W}=\frac{X^{G}}{T}\sum_{s\in G}\frac{X^s}{X^G}\sum_{l}\frac{x_l^s}{X^s}\ln\left(\frac{\frac{x_l^s}{X^s}}{\frac{x_l^G}{X^G}}\right)+\frac{X^{H}}{T}\sum_{s\in H}\frac{X^s}{X^H}\sum_{l}\frac{x_l^s}{X^s}\ln\left(\frac{\frac{x_l^s}{X^s}}{\frac{x_l^H}{X^H}}\right).$$

On the other hand, the mutual information index can be also used to compare overall concentration in two different years. Thus, an intertemporal change between years (2) and (1) can be decomposed in two terms, one showing the gap due to changes in the spatial concentration of sectors (concentration factor) and another due to changes in the sectoral structure of the economy (sectoral structure factor):

$$M^{(2)} - M^{(1)} = \underbrace{\sum_s \frac{X^{s(2)}}{T^{(2)}} \left(\Psi_1 \left(x^{s(2)}; t^{(2)} \right) - \Psi_1 \left(x^{s(1)}; t^{(1)} \right) \right)}_{\text{concentration factor}} - \underbrace{\sum_s \Psi_1 \left(x^{s(1)}; t^{(1)} \right) \left(\frac{X^{s(2)}}{T^{(2)}} - \frac{X^{s(1)}}{T^{(1)}} \right)}_{\text{sectoral structure factor}}$$

3.2 Other overall concentration indexes

The unbounded Gini index, G, proposed by Reardon and Firebaugh (2002) to measure overall segregation in a context of multiple population subgroups can also be used to quantify aggregate concentration since it can be expressed as the weighted mean of index G^s for each sector:

$$G = \sum_s \frac{C^s}{T} G^s .$$

In a multigroup context, Silber (1992) also offers an overall segregation index that extends the popular index of dissimilarity proposed by Duncan and Duncan (1955). This index can be adapted in order to measure overall concentration as follows:

$$IS = \frac{1}{2} \sum_s \sum_l \left| \frac{x_l^s}{T} - \frac{X^s}{T} \frac{t_l}{T} \right| .$$

It is easy to prove that this modified version can also be written as the weighted sum of index D^s for each sector in which the manufacturing industry can be partitioned:

$$IS = \sum_s \frac{C^s}{T} D^s .$$

These indexes can be also decomposed as index M in order to undertake intertemporal comparisons. In the empirical section, these overall concentration measures will be compared with the mutual information index in order to analyze the evolution of the geographic concentration of the manufacturing industry in Spain along the last three decades.

4. Concentration of manufacturing industries in Spain

The data used in this paper comes from the Labor Force Survey (*EPA*) conducted by the Spanish Institute of Statistics (*INE*) by following EUROSTAT's guidelines. Our data corresponds to the second quarter of the year from 1977-2008. Manufacturing industries are considered at a two-digit level of the National Classification of Economic Activities (*CNAE*), and the territorial scale is that of provinces (nuts III).[16]

4.1 Concentration in 2008

In this subsection we analyze the spatial patterns of Spanish manufacturing industries in 2008. Table 1 shows the concentration level of each manufacturing industry, s, at two-digit level according to three of the indexes defined in Section 2: $\Psi_1\left(x^s;t\right)$, G^s, and D^s. We find that these indexes coincide in classifying the *tobacco industry* (**16**), the *leather industry* (**19**), the *refinement of petroleum* (**23**), and *office* and *IT equipment* (**30**) among the most concentrated sectors, while *electronics* (**32**), *other transport material* (**35),** and *recycling* (**37)** are the next in the ranking.

Manufacturing sectors at two-digit level 2008	$\Psi_1\left(x^s;t\right)$	G^s	D^s	**Employment share** (%)
15 *Manufacture of food products and beverages*	0.11	0.26	0.20	15.9
16 *Tobacco industry*	2.67	0.93	0.80	0.2
17 *Textile industry*	0.42	0.48	0.36	2.9
18 *Clothing and fur industry*	0.32	0.39	0.26	2.9
19 *Preparation, tanning and dressing of leather; manufacture of leather goods and luggage articles*	1.63	0.83	0.71	1.6
20 *Wood and cork industry, except furniture; basket making and wickerwork*	0.21	0.35	0.27	3.7
21 *Paper industry*	0.28	0.37	0.24	1.5
22 *Publishing, graphic arts, and reproduction of recorded supports*	0.21	0.36	0.27	5.9
23 *Manufacture of coke, refinement of petroleum and treatment of nuclear fuels*	1.20	0.76	0.60	0.6
24 *Chemical industry*	0.22	0.37	0.28	6.7
25 *Manufacture of rubber and plastic products*	0.23	0.35	0.25	3.4
26 *Manufacture of other non-metallic ore products (Glass, ceramic products, bricks, tiles, cement, etc.)*	0.25	0.33	0.23	6.9

[16] From 1977 to 1992 the *EPA* gathers information at two-digit level according to the *CNAE*-1974 classification, while from 1993 to 2008 the classification used is *CNAE*-93. From 1988 onwards the list of locations includes the 50 provinces and also Ceuta and Melilla. On the other hand, the *EPA* brings information about the location of the establishment where the individual work only from 1999 onwards. Up to that date we have used worker location.

27 *Metallurgy*	0.46	0.49	0.36	3.8
28 *Manufacture of metal products, except machinery and equipment*	0.03	0.13	0.09	12.5
29 *Machinery and mechanical equipment construction industry*	0.11	0.24	0.17	8.2
30 *Manufacture of office machines and IT equipment*	1.22	0.76	0.60	0.3
31 *Manufacture of electrical machinery and material*	0.20	0.34	0.25	3.0
32 *Manufacture of electronic material; manufacture of radio, television and communications apparatus*	0.51	0.52	0.38	1.1
33 *Manufacture of medical-surgical, precision and optical equipment and instruments, and clocks and watches*	0.44	0.49	0.37	1.2
34 *Manufacture of motor vehicles, trailers, and semi-trailers*	0.39	0.47	0.35	7.5
35 *Manufacture of other transport material (Ships, railway material, aircraft, bicycles, motorcycles, etc.)*	0.68	0.59	0.43	2.7
36 *Manufacture of furniture; other manufacturing industries (Jewelry, musical instruments, sport articles, toys, etc.)*	0.09	0.23	0.17	7.5
37 *Recycling*	0.70	0.59	0.45	0.4

Table 1: Concentration indexes and employment share of each manufacturing industry in 2008.

Since we find concentration either in low and high-tech industries at a two-digit level, we may wonder now if there are substantial differences among groups of industries depending on their technological intensity. In order to answer this, we have grouped manufacturing industries by following the OECD and *INE* classifications. Four groups of sectors have been considered (see Table A1 in Appendix B).

This partition of industries seems relevant since differences between these groups of industries (the *between-group* component) explain around 28.5% of the overall manufacturing concentration according to index M. In fact, as shown in Table 2, in Spain the geographic concentration increases with the technological intensity. Thus, the generalized entropy index Ψ_1 indicates that concentration in the high-tech group (jointly considered) triples that of the medium-high-tech group, while that of the latter doubles the value of the groups with lower technological intensity.

Technological groups 2008	$\Psi_1(x^s;t)$	G^s	D^s	**Employment share** (%)
High-tech	0.336	0.438	0.330	5.1
Medium-high-tech	0.101	0.250	0.190	25.3
Medium-low-tech	0.044	0.160	0.119	27.2
Low-tech	0.037	0.151	0.109	42.3

Table 2: Concentration indexes and employment share of each technological group in 2008

Next, we calculate the contribution of each group of sectors to the concentration of the whole manufacturing industry at a two-digit level (see Table 3). As mentioned in

Section 3, three overall measures are used in order to obtain the concentration of the whole manufacturing industry: the mutual information index (M), the aggregate Gini index (G), and the index derived from the segregation literature (IS).[17]

Technological groups 2008	**Contribution to** M **(%)**	**Contribution to** G **(%)**	**Contribution to** IS **(%)**
High-tech	12.8	8.7	8.6
Medium-high-tech	23.7	26.8	26.8
Medium-low-tech	21.3	22.1	21.4
Low-tech	42.2	42.4	43.2

Table 3: Contribution of each group of sectors to overall manufacturing concentration in 2008

We find that, with any of these measures, the contribution of the low-tech group to overall concentration is the highest (with values over 42%). Note, however, that this contribution is similar to the weight that this group represents in terms of manufacturing employment (see Table 2). Moreover, in the case of the medium-low-tech industry, the contribution of this group to the industrial aggregate is remarkable lower than expected (21-22% against 27%). It follows then that the concentration of the industrial aggregate does not rest on industries with low technological intensity, which is in line with the finding mentioned above. On the contrary, when using the mutual information index, the contribution of the high-tech industry to overall concentration more than doubles its demographic weigh (12.8% against 5.1%). According to the other two overall indexes, the contribution of this group to aggregate concentration is smaller (around 8.6%), but it exceeds in any case its demographic weight. Regarding the medium-high-tech industry, the results do depend on the index being used. Thus, according to the mutual information index its contribution is lower than expected, while according to the other indexes the opposite holds.

From all of the above, it seems that high-technology industries play an important role in explaining the spatial concentration of manufacturing employment in Spain. This finding should not surprise us since knowledge spillovers can be an important source of agglomeration externalities in this kind of sector. This result is in line with that obtained by Alonso-Villar et al. (2004) by using instead the index proposed by Maurel and Sédillot (1999) (M-S), which is obtained from a location model, and a different Spanish

[17] The values of these indexes for 2008 are shown in Figure 1, which includes not only these figures but also those of the whole period.

dataset for 1999. Therefore, the result for the Spanish economy seems rather robust. We should note, however, that this spatial pattern is not found in all countries, since Devereux et al. (2004) did not find evidence of this behavior in the UK by using the M-S index.

4.2 Evolution of manufacturing concentration in the last three decades

Figure 1 shows the evolution of the overall concentration level of the Spanish manufacturing industry from 1977-2008. Our results show that the values of index M and IS are rather similar, while G always takes higher values. In any case, according to these indexes we can establish three different periods of change in the evolution of spatial concentration. First, we observe a slight increase until 1981, a noteworthy decrease from 1985 to 1990, and finally, a remarkable decline from 2001 onwards.[18]

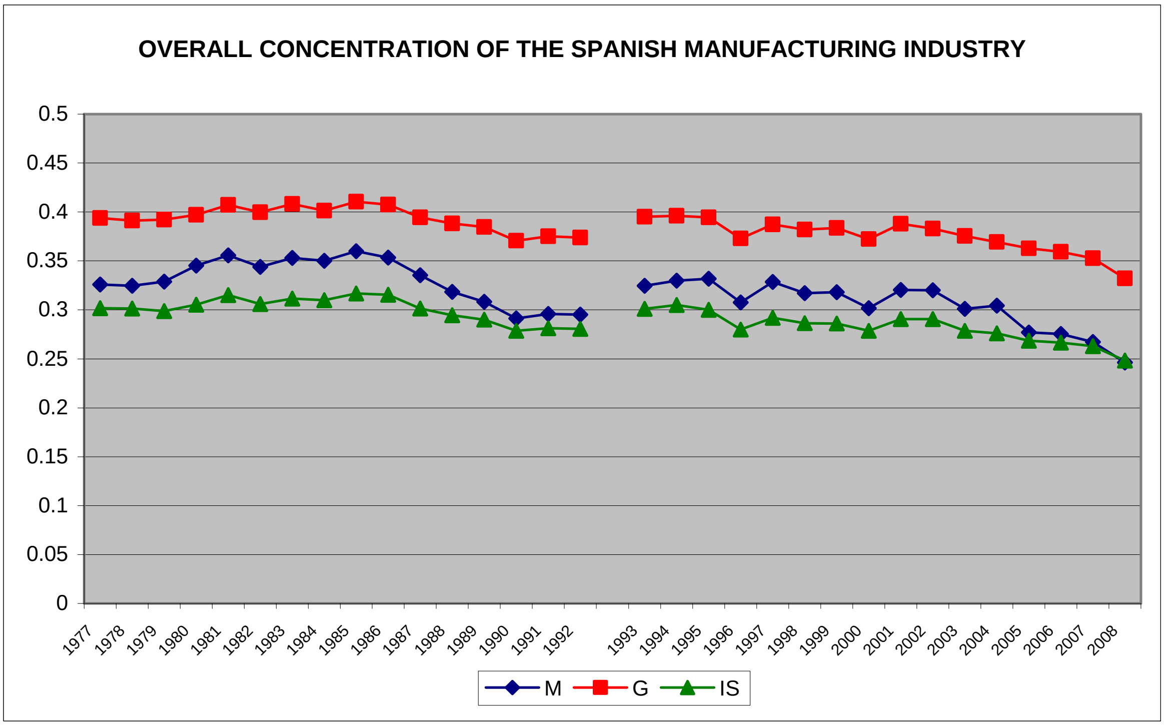


Figure 1. Overall concentration of the manufacturing industry over the period 1977-2008.

Table 4 shows the decomposition of the three overall measures in the three periods. We find that the overall concentration rise in the first years of democracy is entirely due to

[18] Note that there is also a change from 1992 to 1993 due to the use of another classification of manufacturing industries, as mentioned above.

the concentration increase of the manufacturing industry (concentration factor) and no to changes in the distribution of employment across sectors (sectoral structure factor). The decreases in the other two periods are also mainly due to changes in the concentration levels of the manufacturing industries, since the concentration factor accounts for at least 60% of total change.

Therefore, the process of economic integration (Spain joins the EU in 1986) and the improvements in transport infrastructures have not fueled the spatial concentration of the Spanish manufacturing industry.[19] On the contrary, these processes seem to be accompanied of a decreasing, even though intermittent, trend in the concentration level that reaches our days.

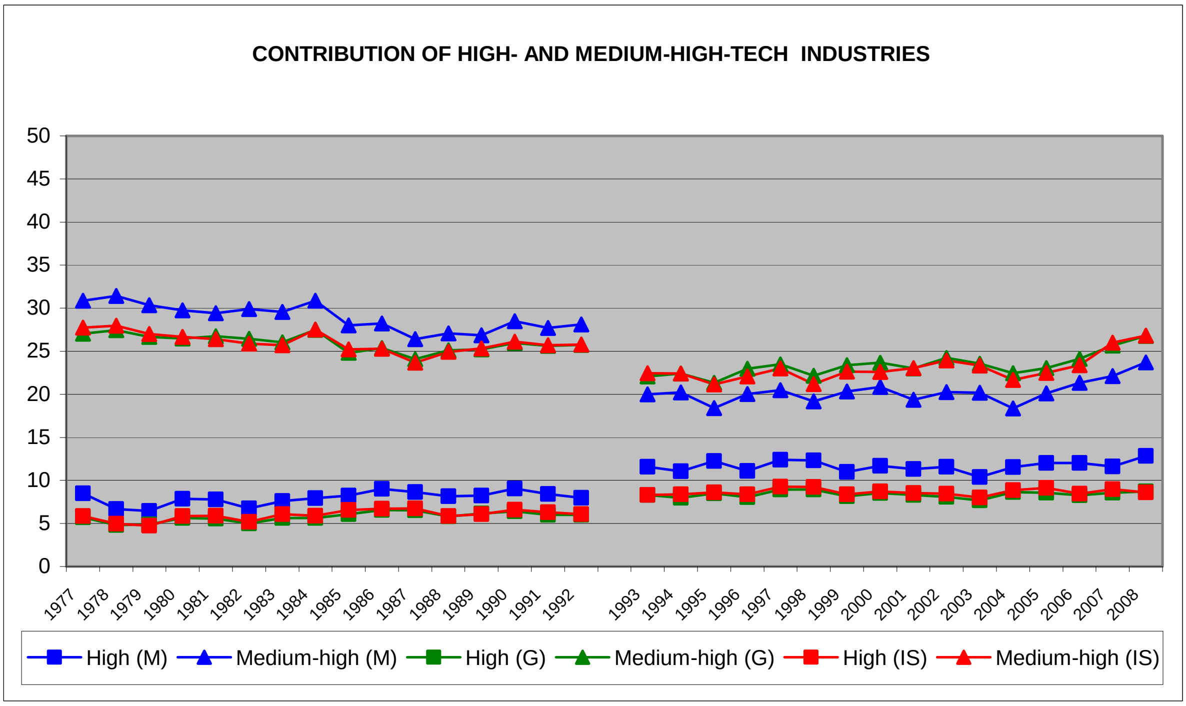
	Concentration/structure factor decomposition *M*		**Concentration/structure factor decomposition** *G*		**Concentration/structure factor decomposition** *IS*	
1977-1981	109.98%	-9.98%	109.17%	-9.17%	109.16%	-9.16%
1985-1990	64.21%	35.79%	72.73%	27.27%	73.89%	26.11%
2001-2008	59.86%	40.14%	75.82%	24.18%	75.34%	24.66%

Table 4: Intertemporal decompositions of overall concentration indexes.

Figure 2 shows the contribution of each technological group to the overall indexes along the whole period.[20] We observe a decreasing tendency in the contribution of low-tech industries to overall concentration from 1999 onwards, together with an increase in the contribution of medium-high-tech industries from 2004 onwards. We should note, however, that while the decrease in the low-tech group could be explained by the evolution of the employment in this group along the period, the evolution of the medium-high-tech group can be only partially explain by that factor (see Figure A1 in Appendix B).

[19] The Spanish openness to international economy had started some years before its accession to the UE. In particular, Spain signed the Preferential Agreement with the European Community in 1970, and the Agreement with the European Free Trade Association in 1979 (Alonso, 1995).

[20] Since the classification of the manufacturing industry at two-digit level between 1977 and 1992 is different from the one used between 1993 and 2008, we have used another technological grouping. In any case, a similar broad criterion to that used in 2008 has been used for the whole period. See Table A2 in Appendix B.



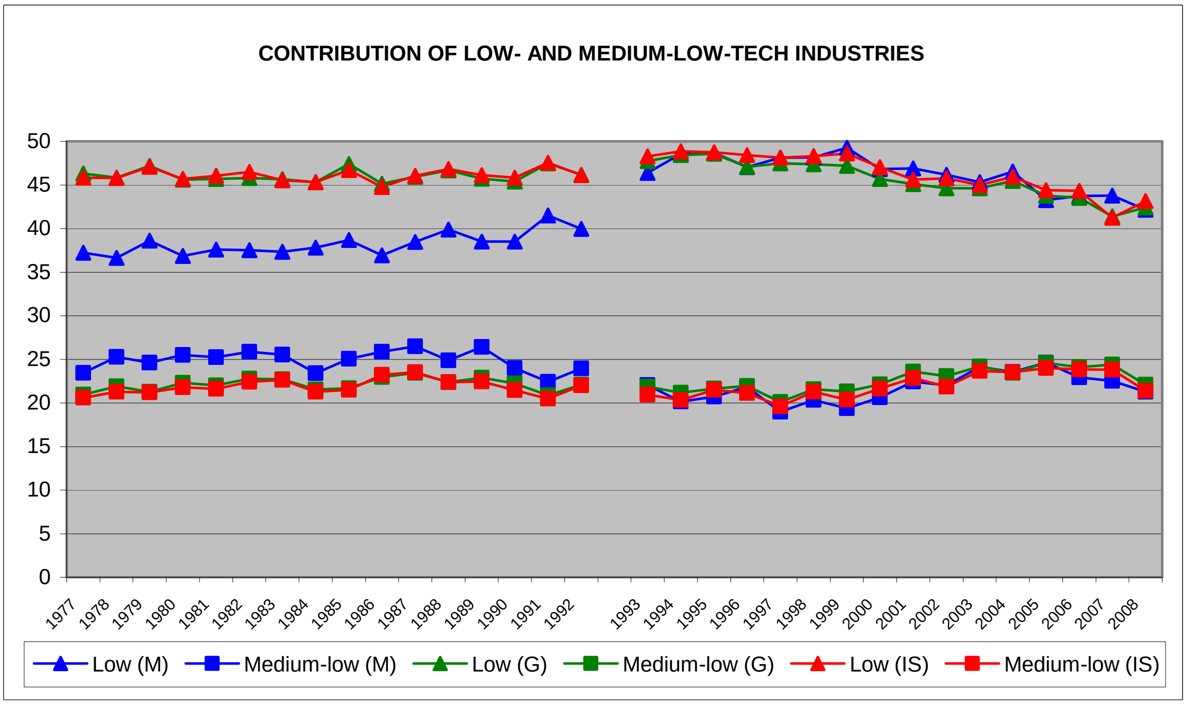


Figure 2. Contribution of each technological group to the three overall concentration indexes.

5. Final comments

The measurement of population segregation across organizational units (occupations, schools, neighborhoods, etc.) and the measurement of spatial concentration share much in common with each other. Thus, while the former focuses on the distribution of racial groups across schools, for example, the latter addresses the distribution of economic sectors across location units. However, each field has faced measurement from a different approach. On one hand, while the segregation literature has mainly tackled this matter from an axiomatic perspective, the literature on spatial concentration has not. On the other hand, the former has mostly focused on the measurement of overall segregation (exceptions are Moir and Selby Smith, 1979; and Alonso-Villar and Del Río, 2007), whereas the latter has dealt with the concentration of any single sector (which has been labeled here as partial concentration) rather than with overall concentration (exceptions are Aiginger and Davies, 2004; Cutrini, 2009a, b).

Given the parallelism between both phenomena, this paper has proposed an overall concentration measure that parallels that axiomatically characterized by Frankel and Volij (2007) in a context of school segregation. Based on this characterization, we have shown the properties that this index, the mutual information index, has in our context. Our analysis reveals that this overall concentration index can be written as the weighed sum of the Theil index for each sector in which the economic activity can be partitioned (partial concentration). We have also characterized this partial concentration index, together with the remaining members of the generalized entropy family of concentration indexes, in terms of basic axioms borrowed from the literature on income distribution and occupational segregation and adapted to our case. Consequently, this paper brings theoretical support to those works that use the weighed average of Theil index in order to quantify overall concentration (Aiginger and Davies, 2004; and Cutrini, 2009a, b).

Finally, the mutual information index together with other measures borrowed from the segregation literature has been used to measure overall concentration of Spanish manufacturing industry along the last three decades. We found that the concentration of the manufacturing industry experienced a slight increase during the first years of the democratic period, while it tended to diminish after the EU's accession.[21] This result is in line with the predictions of the new economic geography, according to which economic integration processes tend first to favor agglomeration and later dispersion between locations (Venables, 1996; Puga, 1999; and Ottaviano et al., 2002, inter alia).

[21] This finding corroborates that obtained by Paluzie et al. (2004), who use alternative indexes and datasets for the period 1955-1995.

The decomposition of the mutual information index in the *between-group, within-group* components demonstrates that the technological intensity of an industry is a relevant variable to explain geographic concentration. In addition, our analysis shows that the high-tech industry contributes to overall concentration at a larger extent than expected according to its employment share, which suggests the importance of knowledge spillovers for this kind of industries (Baptista and Mendonça, 2009; García Muñiz et al., 2009).

References

Aiginger, K. and Davies, S.W. (2004): "Industrial specialization and geographic concentration: two sides of the same coin? Not for the European Union", Journal of Applied Economics, VII(2), 231-248.

Aiginger, K. and Pfaffermayr, M. (2004): "The single market and geographic concentration in Europe", *Review of International Economics* 12(1), 1-11.

Alonso, J.A. (1995): "Comercio exterior", in J.L. García Delgado, R. Myro, and J.A. Martínez Serrano (eds.), *Lecciones de Economía Española*, Editorial Civitas, Madrid.

Alonso-Villar, O., Chamorro-Rivas, J.M., and González-Cerdeira, X. (2004): "Agglomeration economies in manufacturing industries: the case of Spain", *Applied Economics*, 36(18), 2103- 2116.

Alonso-Villar, O. and Del Río, C. (2007): "An alternative proposal for measuring occupational segregation", ECINEQ Working Paper 2007-82.

Amiti, M. (1999): "Specialization patterns in Europe", *Weltwirtschaftliches Archiv* 135(4), 573-593.

Baptista, R. and Mendonça, J. (2009): "Proximity to knowledge sources and the location of knowledge-based start-ups", *Annals of Regional Science*, forthcoming.

Brakman, S., Garretsen, H., Gorter, J., van der Horst, A. and Schramm, M. (2005): "New economic geography, empirics, and regional policy", CPB (Netherlands Bureau for Economic Policy Analysis) Special Publication 56.

Brülhart, M. (2001): "Evolving geographical concentration of European manufacturing industries", *Weltwirtschaftliches Archiv* 137(2), 215-243.

Brülhart, M. and Traeger, R. (2005): "An account of geographic concentration patterns in Europe", *Regional Science and Urban Economics* 35(6), 597-624.

Cover, T.M. and Thomas, Y.A. (1991): *Elements of Information Theory*, Wiley Interscience, New York.

Cowell, F.A. (2000): "Measurement of Inequality", in A.B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution vol. 1,* Elsevier-North Holland.

Cutrini, E. (2009a): "Using entropy measures to disentangle regional from national localization patterns", *Regional Science and Urban Economics*, 39, 243-250.

Cutrini, E. (2009b): “Specialization and concentration from a twofold geographical perspective: Evidence from Europe”, *Regional Studies,* forthcoming.

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

Duncan, O.D. and Duncan, B. (1955): A methodological analysis of segregation indexes, *American Sociological Review* 20(2), 210-217.

Duranton, G. and Overman, H. (2005): “Testing for localization using micro-geographic data”, *Review of Economic Studies* 72, 1077-1106.

Ellison, G. and Glaeser, E. (1997): “Geographic concentration in U.S. manufacturing industries: a dartboard approach”, *Journal of Political Economy* 105, 889-927.

Flückiger, Y. and Silber, J. (1999): *The Measurement of Segregation in the Labor Force*, Physica-Verlag, Heidelberg, Germany.

Foster, J.E. (1985): Inequality measurement, in H. Peyton Jound (ed.), *Fair Allocation, American Mathematical Society. Proceedings of Simposia in Applied Mathematics* 33, 31-68.

Frankel, D.M. and Volij, O. (2007): “Measuring segregation”, Working Paper 07009, Department of Economics, Iowa State University.

Frankel. D.M. and Volij, O. (2008): “Measuring School Segregation”. Mimeo. Iowa State University.

Fujita, M., Krugman, P. and Venables, A. (2000): *The Spatial Economy. Cities. Regions and International Trade*, MIT Press, second printing.

García Muñiz, A.S., Morillas Raya, A., and Ramos Carvajal, C. (2009): “Spanish and European innovation difusión: a structural hole approach in the input-output field”, *Annals of Regional Science*, forthcoming.

Guimarães, P., Figueiredo, O. and Woodward, D. (2007): “Measuring the localization of economic activity: A parametric approach”, *Journal of Regional Science* 47(4), 753-774.

Haaland, J., Kind, H., Midelfart Knarvik, K., and Torstensson, J. (1999): “What determines the economic geography of Europe?”, CEPR Discussion Paper 2072.

Hutchens, R.M. (2004): “One measure of segregation”, *International Economic Review* 45(2), 555-578.

Krugman, P. (1991): *Geography and Trade*, MIT Press.

Marcon, E. and Puech, F. (2003): “Evaluating the geographic concentration of industries using distance-base methods”, *Journal of Economic Geography* 3, 409-428.

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

Moir, H. and Selby Smith, J. (1979): “Industrial segregation in the Australian labour market”, *Journal of Industrial Relations* 21, 281-291.

Mora, R. and Ruiz-Castillo, J. (2008): “Multigroup and multilevel residential segregation: The U.S. case, 1989-2005”, Working Paper 08-61, Economic Series 28, Universidad Carlos III de Madrid.

Mora, R. and J. Ruiz-Castillo (2009): “The invariance properties of the mutual information index of multigroup segregation”, in Y. Flückiger, S. Reardon, and J. Silber (eds.), *Research on Economic Inequality. New Frontiers in the Field of Segregation Measurement and Analysis*, Jay Press. UK, forthcoming.

Mori, T., Nishikimi, K., and Smith, T. (2005): “A divergence statistic for industrial localization”, *The Review of Economics and Statistics* 87(4), 635-651.

Neary, P. (2001): “Of hype and hyperbolas: introducing the new economic geography”, *Journal of Economic Literature* 39 (2), 536-561.

OECD (2007): *OECD Science. Technology and Industry. Scoreboard 2007*.

Ottaviano, G., Tabuchi, T. and Thisse, J.-F. (2002): “Agglomeration and trade revisited”, *International Economic Review* 43, 409-436.

Ottaviano. G. and Thisse J.-F. (2004): “Agglomeration and economic geography”, in: J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Urban and Regional Economics*, vol. 4. Elservier-North Holland.

Paluzie, E., Pons, J., Tirado, D. (2004): “The geographical concentration of industry across Spanish regions, 1856-1995”, *Review of Regional Research*, 24(2), 143-160.

Pérez-Ximénez, D. and Sanz-Gracia, F. (2007): “Geographical concentration of service activities across U.S: states and counties. 1969-2000”. Paper presented at the XVII International RESER (European Research Network on Services and Space) Conference in Tempere (Finland).

Puga, D. (1999): “The rise and fall of regional inequalities”, *European Economic Review*, 43(2), 303-334.

Reardon, S.F. and Firebaugh, G. (2002): “Measures of multigroup segregation”, *Sociological Methodology* 32, 33-76.

Resmini, L. (2007): “Regional patterns of industry location in transition countries: Does economic integration with the European Union matter?”. *Regional Studies* 41(6), 747-764.

Silber, J. (1992): “Occupational segregation indices in the multidimensional case: A note”, *The Economic Record* 68, 276-277.

Shorrocks, A.F. (1984): “Inequality decompositions by population subgroups”, *Econometrica* 48, 1369-1386.

Suedekum, J. (2006): “Concentration and specialization trends in Germany since re-unification”, *Regional Studies* 40(8), 861-873.

Venables, A.J. (1996): “Equilibrium locations of vertically linked industries”, *International Economic Review*, 37, 341-359.

Appendix A

List of axioms

Axiom 1: *Symmetry in locations*. If $(\Pi(1),...,\Pi(L))$ represents a permutation of locations, then $I_c\left(x^s\Pi;t\Pi\right)=I_c\left(x^s;t\right)$, where $x^s\Pi=\left(x^s{}_{\Pi(1)},...,x^s{}_{\Pi(L)}\right)$ and $t\Pi=\left(t_{\Pi(1)},...,t_{\Pi(L)}\right)$.

Axiom 2: *Movement between locations*. If $\left(x^s\,';t'\right)\in D$ is obtained from $\left(x^s;t\right)\in D$ in such a way that:

(i) Location *i* loses employment in the sector of study, while the opposite happens to location *h*, i.e., $x_i^s\,'=x_i^s-d$, $x_h^s\,'=x_h^s+d$ $(0<d\le x_i^s)$, where *i* and *h* are two locations with the same aggregate employment level, $t_i=t_h$, but with different shares in the sector of study since $x_i^s<x_h^s$;

(ii) The employment level of the sector of study does not change in the remaining locations, i.e., $x_l^s\,'=x_l^s$ $\forall l\neq i,h$;

then $I_c\left(x^s\,';t'\right)>I_c\left(x^s;t\right)$.

In other words, if location *i* has initially the same manufacturing employment level as location *h*, but a lower employment level in the sector of study, then a movement of employment in that sector from location *i* to location *h* would be considered a disequalizing movement fostering the concentration of the sector. [22]

Axiom 3: *Scale Invariance*. If the distribution of the sector of study, x^s, is multiplied by a positive scalar, a, and the distribution of reference, t, is multiplied by another positive scalar, b, in such a way that $ax_l^s\le bt_l$, then $I_c\left(ax^s;bt\right)=I_c\left(x^s;t\right)$.

Therefore, this axiom means that in measuring spatial concentration it is only employment shares that matter, not employment levels.[23]

[22] In the literature on income distribution, this property corresponds to the Pigou-Dalton principle of transfers. This axiom has also been adapted to measure occupational segregation, where it is called "movement between groups" (Hutchens, 2004; Alonso-Villar and Del Río, 2007).

[23] In a context of occupational segregation, see Alonso-Villar and Del Río (2007).

The next axiom requires that subdividing a location into several units of equal size, both in terms of aggregate employment and in terms of employment in the sector of study, does not affect the concentration level of the sector. Without loss of generality, in the next axiom the subdivision is undertaken for the last location in order to make notation easier.

Axiom 4: *Insensitivity to proportional subdivisions of locations.* If $\left(x^{s}';t'\right) \in D$ is obtained from $\left(x^{s};t\right) \in D$ in such a way that:

(i) All locations except the last one remain unaltered both in terms of aggregate employment and employment in the sector of study, i.e., $t'_{l} = t_{l}$ and $x_{l}^{s}' = x_{l}^{s}$ for any $l = 1,...,L-1$;

(ii) The last location is subdivided in M location units without introducing any difference among them in terms of employment shares, i.e., $x_{l}^{s}' = x_{L}^{s}/M$, $t'_{l} = t_{L}/M$ for any $j = L,...,L+M-1$,

then, $I_{c}\left(x^{s}';t'\right) = I_{c}\left(x^{s};t\right)$.

Our next axiom, *aggregation*, is very helpful for empirical analyses since it has to do with the decomposition of indexes by subgroups. Like axiom 4, *aggregation* is related to spatial scale, but as opposed to it, location units are now aggregated rather than subdivided.

Axiom 5: *Aggregation.* Let us assume that locations can be partitioned into two mutually exclusive groups so that $(x^{s};t) = (x^{s1}, x^{s2}; t^{1}, t^{2})$, where the aggregate employment level in locations included in group 1 (2) is denoted by T_{1} (T_{2}), while X_{1}^{s} (X_{2}^{s}) represents the employment level of the sector of study in the corresponding group of locations. Concentration index I_{c} is defined as aggregative if there exists a continuous aggregator function A such that

$I_c\left(x^s;t\right)=A\left(I_c(x^{s1};t^1),\frac{X_1^s}{T_1},T_1,I_c(x^{s2};t^2),\frac{X_2^s}{T_2},T_2\right)$, where A is strictly increasing in the first and fourth argument.[24]

Therefore, the overall concentration level of the sector of study is a function of: (a) the concentration level of the sector in each group of locations (denoted by $I_c(x^{s1};t^1)$ in group 1); (b) the employment level in each group of locations (denoted by T_1 in group 1); and (c) the employment share of the sector in each group of locations (denoted by $\frac{X_1^s}{T_1}$ in group 1).

Proof of Proposition 1.

First step: If the concentration index I_c satisfies axioms 1-4, then the inequality index I evaluated at the fictitious income distribution y as $I(y):=I_c\left(x^s;t\right)$, where $y\equiv\left(\underbrace{\frac{x_1^s}{t_1},...,\frac{x_1^s}{t_1}}_{t_1},...,\underbrace{\frac{x_L^s}{t_L},...,\frac{x_L^s}{t_L}}_{t_L}\right)$, works as an inequality index satisfying scale invariance, symmetry, the Pigou-Dalton principle, and replication invariance.

a) I is well defined. Note that several vectors $\left(x^s;t\right)$ can be reached after grouping individuals in the fictitious "income distribution" who belong to the same location depending on how many locations are considered. However, by axiom 4, all these vectors have the same spatial concentration level, since they can be obtained from each other by proportional subdivisions.

b) Scale invariance. This property is certainly satisfied by index I since $I(\theta\frac{x_1^s}{t_1},...,\theta\frac{x_1^s}{t_1},...,\theta\frac{x_L^s}{t_L},...,\theta\frac{x_L^s}{t_L})=I_c(\theta x^s;t)$, which is equal to $I_c(x^s;t)$ because I_c satisfies axiom 3 (case where $a>0,\ b=1$).

c) Symmetry. This property requires that "individuals" play symmetric roles in the inequality index. This is satisfied by I since I_c satisfies axioms 1 and 4.

[24] The formulation used here is analogous to that put forward by Hutchens (2004) to measure occupational segregation.

d) The Pigou-Dalton transfer principle. From axiom 4, any regressive transfer in this fictitious economy can be expressed as a sequence of disequalizing employment movements in an economy constructed from the original one by proportional subdivisions of locations so that the distribution of reference becomes $\left(\underbrace{1,...,1}_{T}\right)$. Since I_c satisfies axiom 2, the second situation leads to a higher concentration index and, therefore, to a higher value of I.

e) Replication invariance. This axiom means that when replicating the economy k-times, so that for every individual in the previous economy there are now k identical individuals, income inequality is not altered. This axiom is satisfied here since a k-replication of the fictitious distribution leads to a k-replication of vector $\left(x^s;t\right)$, and I_c satisfies axiom 3 (case where $a=b$). □

Second step: *Any concentration index* I_c *satisfying axioms 1-5 can be written as a strictly increasing monotonic transformation of* Ψ_α.

Following Shorrocks (1984) and Foster (1985), any continuous inequality measure I taking a zero value at the egalitarian distribution and satisfying scale invariance, replication invariance, the Pigou-Dalton transfer principle, symmetry, and aggregation can be written as $I(y)=F^{-1}(I_\alpha(y))$ for some parameter α, where F is a strictly increasing function such that $F:[0,\infty)\to\mathbb{R}$, with $F(0)=0$ and I_α is the well-known generalized entropy family of inequality indexes:

$$I_\alpha(y)=\begin{cases}\dfrac{1}{n\alpha(\alpha-1)}\displaystyle\sum_i\left[\left(\dfrac{y_i}{\frac{1}{n}\sum_k y_k}\right)^\alpha-1\right] & \text{if } \alpha\neq 0,1\\[2ex] \dfrac{1}{\mathrm{n}}\displaystyle\sum_i\left[\dfrac{y_i}{\frac{1}{n}\sum_k y_k}\ln\left(\dfrac{y_i}{\frac{1}{n}\sum_k y_k}\right)\right] & \text{if } \alpha=1\\[2ex] \dfrac{1}{\mathrm{n}}\displaystyle\sum_i\ln\left(\dfrac{\frac{1}{n}\sum_k y_k}{y_i}\right) & \text{if } \alpha=0\end{cases}$$

In Step 1 we proved that any concentration index I_c satisfying axioms 1-4 can be regarded as an inequality index I satisfying scale invariance, symmetry, the Pigou-Dalton transfer principle and replication invariance. It is easy to see that if I_c is a continuous function, so too is I. If we additionally show that I is aggregative and also that it is equal to zero at the egalitarian distribution, we can use Shorrocks's result in order to characterize inequality index I.

An inequality index I is defined as aggregative if $I(y) = A(I(y^1), \mu(y^1), n(y^1), I(y^2), \mu(y^2), n(y^2))$, where A is a continuous function that is strictly increasing in the first and fourth arguments, y^i represents the income distribution corresponding to individuals' group i, $\mu(.)$ is the average of the corresponding distribution, and $n(.)$ is the number of individuals in the corresponding group. In our case, the "income" distribution is $y \equiv (\underbrace{\frac{x_1^s}{t_1},...,\frac{x_1^s}{t_1}}_{t_1},...,\underbrace{\frac{x_L^s}{t_L},...,\frac{x_L^s}{t_L}}_{t_L})$, and the average of that distribution is equal to $\frac{X^s}{T}$. In what follows, we show that our I is an aggregative inequality index. For the sake of simplicity, assume that class 1 includes locations $l = 1,...,i$, while class 2 is the complementary. By definition

$$I\left(\underbrace{\frac{x_1^s}{t_1},...,\frac{x_1^s}{t_1},...,\frac{x_i^s}{t_i},...,\frac{x_i^s}{t_i}}_{\text{class 1}}\underbrace{\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_L^s}{t_L},...,\frac{x_L^s}{t_L}}_{\text{class 2}}\right) = I_c(x^s;t)\cdot$$

On the other hand, since by axiom 5 I_c is an aggregative concentration index:

$$I_c(x^s;t) = I_c(x^{s1}, x^{s2}; t^1, t^2) = A\left(I_c(x^{s1};t^1), \frac{X_1^s}{T_1}, T_1, I_c(x^{s2};t^2), \frac{X_2^s}{T_2}, T_2\right).$$

Note that $I_c(x^{s1};t^1) = I(\frac{x_1^s}{t_1},...,\frac{x_1^s}{t_1},...,\frac{x_i^s}{t_i},...,\frac{x_i^s}{t_i})$, and $I_c(x^{s2};t^2) = I(\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_L^s}{t_L},...,\frac{x_L^s}{t_L})$.

Therefore,

$$I\left(\underbrace{\frac{x_1^s}{t_1},...,\frac{x_1^s}{t_1},...,\frac{x_i^s}{t_i},...,\frac{x_i^s}{t_i}}_{\text{class 1}}\underbrace{\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_L^s}{t_L},...,\frac{x_L^s}{t_L}}_{\text{class 2}}\right) = A\left(I(\frac{x_1^s}{t_1},...,\frac{x_1^s}{t_1},...,\frac{x_i^s}{t_i},...,\frac{x_i^s}{t_i}), \frac{X_1^s}{T_1}, T_1, I(\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_{i+1}^s}{t_{i+1}},...,\frac{x_L^s}{t_L},...,\frac{x_L^s}{t_L}), \frac{X_2^s}{T_2}, T_2\right),$$

where $\frac{X_1^s}{T_1}$ (respectively, $\frac{X_2^s}{T_2}$) represents the average "income" of "individuals" in class

1 (respectively, 2), while T_1 (respectively, T_2) is the number of “individuals” in that class. Therefore, the inequality index I is aggregative.

Finally, note that I is equal to zero when all “individuals” have the same “income,” i.e., when all locations have the same employment shares in the sector under consideration ($\frac{x_l^s}{t_l} = \frac{X^s}{T} \ \forall l$).

Therefore, by using Shorrocks’s result, it follows that $I(y) = F^{-1}(I_\alpha(y))$ for $\alpha \neq 0,1$ or $\alpha = 1$.[25] On the other hand, $I_c(x^s;t) = I(y)$ and $F^{-1}(I_\alpha(y)) = F^{-1}(\Psi_\alpha(x^s;t))$, which completes the proof of step two.

Third step: $F^{-1}(\Psi_\alpha)$ *is a concentration index satisfying symmetry in locations, movement between locations, scale invariance, insensitivity to proportional subdivisions of locations, and aggregation.*

In order to prove this, it suffices to show that Ψ_α satisfies the above properties, which is done in what follows. It is easy to prove that Ψ_α verifies *scale invariance*, *symmetry in locations*, and *insensitivity to proportional subdivisions*. To demonstrate that Ψ_α satisfies the axiom of *movement between locations*, note that any disequalizing movement from location *i* to *h*, where $t_i = t_h$ and $x_i^s < x_h^s$, implies moving from “income” distribution $y = \left(\frac{x_1^s}{t_1}, \ldots, \frac{x_1^s}{t_1}, \ldots, \frac{x_i^s}{t_i}, \ldots, \frac{x_i^s}{t_i}, \ldots, \frac{x_h^s}{t_h}, \ldots, \frac{x_h^s}{t_h}, \ldots, \frac{x_L^s}{t_L}, \ldots, \frac{x_L^s}{t_L} \right)$ to “income” distribution $y' = \left(\frac{x_1^s}{t_1}, \ldots, \frac{x_1^s}{t_1}, \ldots, \frac{x_i^s - d}{t_i}, \ldots, \frac{x_i^s - d}{t_i}, \ldots \frac{x_h^s + d}{t_h}, \ldots, \frac{x_h^s + d}{t_h}, \ldots, \frac{x_L^s}{t_L}, \ldots, \frac{x_L^s}{t_L} \right)$. On the other hand, $I_\alpha(y) = \Psi_\alpha(x^s;t)$ and $I_\alpha(y') = \Psi_\alpha(x^{s}{}';t')$. Since I_α is an inequality measure satisfying the Pigou-Dalton transfer principle and *y’* can be obtained from *y* by a finite sequence of regressive transfers it follows that $\Psi_\alpha(x^{s}{}';t') > \Psi_\alpha(x^s;t)$.

---

[25] The case where $\alpha = 0$ is discarded, because when the sector of study has no employment in location *l* (i.e., when $x_l^s = 0$) and $\alpha = 0$, the index value would be infinite and, therefore, it has no sense. The case where $\alpha = 1$ does not have the same problem since $\lim_{x_l \to 0} \frac{x_l^s / X^s}{t_l / T} \ln\left(\frac{x_l^s / X^s}{t_l / T} \right) = 0$.

Next, we prove that Ψ_α is aggregative. By simple calculations Ψ_α can be written as

$$\Psi_\alpha(x^{s1}, x^{s2}; t^1, t^2) = \begin{cases} -\dfrac{1}{\alpha(\alpha-1)} + \left[\begin{array}{l} \left(\dfrac{T_1}{T}\right)^{1-\alpha} \left(\dfrac{X_1^s}{X^s}\right)^{\alpha} \left(\Psi_\alpha\left(x^{s1}; t^1\right) + \dfrac{1}{\alpha(\alpha-1)}\right) + \\ + \left(\dfrac{T_2}{T}\right)^{1-\alpha} \left(\dfrac{X_2^s}{X^s}\right)^{\alpha} \left(\Psi_\alpha\left(x^{s2}; t^2\right) + \dfrac{1}{\alpha(\alpha-1)}\right) \end{array} \right] & \text{for } \alpha \neq 0,1 \\ \dfrac{X^{s1}}{X^s}\left[\Psi_\alpha(x^{s1}; t^1) + \ln\left(\dfrac{X_1^s}{T_1}\dfrac{T}{X^s}\right)\right] + \dfrac{X_2^s}{X^s}\left[\Psi_\alpha(x^{s2}; t^2) + \ln\left(\dfrac{X_2^s}{T_2}\dfrac{T}{X^s}\right)\right] & \text{for } \alpha = 1 \end{cases}$$

On the other hand, $T = T_1 + T_2$ and $X^s = X_1^s + X_2^s$. Therefore, Ψ_α can be written as $\Psi_\alpha(x^{s1}, x^{s2}; t^1, t^2) = A\left(\Psi_\alpha\left(x^{s1}; t^1\right), \dfrac{X_1^s}{T_1}, T_1, \Psi_\alpha\left(x^{s2}; t^2\right), \dfrac{X_2^s}{T_2}, T_2\right)$, which completes the proof. □

Appendix B

High-technology group	*Office, accounting, and computing machinery* (**30**) *Radio, TV, and communications equipments* (**32**) *Medical, precision and optical instruments* (**33**) *Manufacture of other transport material* (including *aircraft*) (**35**)
Medium-high-technology group	*Chemicals* (**24**) *Machinery and equipment n.e.c.* (**29**) *Electrical machinery and apparatus n.e.c.* (**31**) *Motor vehicles, trailers and semi-trailers* (**34**)
Medium-low-technology group	*Coke, refined petroleum products and nuclear fuel* (**23**) *Rubber and plastic products* (**25**) *Other non-metallic mineral products* (**26**) *Metallurgy* (**27**) *Fabricated metal products (machinery and equipment excluded)* (**28**)
Low-technology group	*Food products and beverages* (**15**) *Tobacco* (**16**) *Textile industry* (**17**) *Clothing and fur industry* (**18**) *Leather and footwear* (**19**) *Wood and cork industry, except furniture* (**20**) *Paper industry* (**21**) *Publishing, graphic arts, and reproduction of recorded supports* (**22**) *Manufacture of furniture and other manufacturing industries n.e.c.* (**36**) *Recycling* (**37**)

Table A1. Classification of two-digit industries by technological intensity: 1993-2008. [26]

[26] Since in this study industries are considered at a two-digit level, and the OECD (2007) and *INE* classifications consider both two- and three-digit industries, we have introduced some changes with respect to them. In particular, the *INE* includes one of the subsectors of sector **24** in the high-tech group and the remaining subsectors in the medium-high-tech group. Here, we have decided to include the whole

High-technology group	*Office and computing machinery* (**33**) *Electronics* (**35**) *Medical, precision and optical instruments* (**39**) *Manufacture of other transport material* (including *aircraft*) (**38**)
Medium-high-technology group	*Chemicals* (**25**) *Machinery and equipment* (**32**) *Electrical machinery and apparatus* (**34**) *Motor vehicle and trailers* (**36**) *Shipping building* (**37**)
Medium-low technology group	*Refined petroleum products* (**13**) *Rubber and plastic products* (**48**) *Non-metallic mineral products* (**24**) *Fabricated metal products(except machinery and transport material)* (**31**) *Production and first transformation of metals* (**22**)
Low-technology group	*Food products, beverages,* and tobacco (**41**, **42**) *Textile industry* (**43**) *Leather industry* (**44**) *Footwear and clothing* (**45**) *Wood, cork industry and furniture* (**46**) *Paper industry, publishing, and graphic arts* (**47**) *Other manufacturing industries n.e.c.* (**49**)

Table A2. Classification of two-digit industries by technological intensity: 1977-1992.

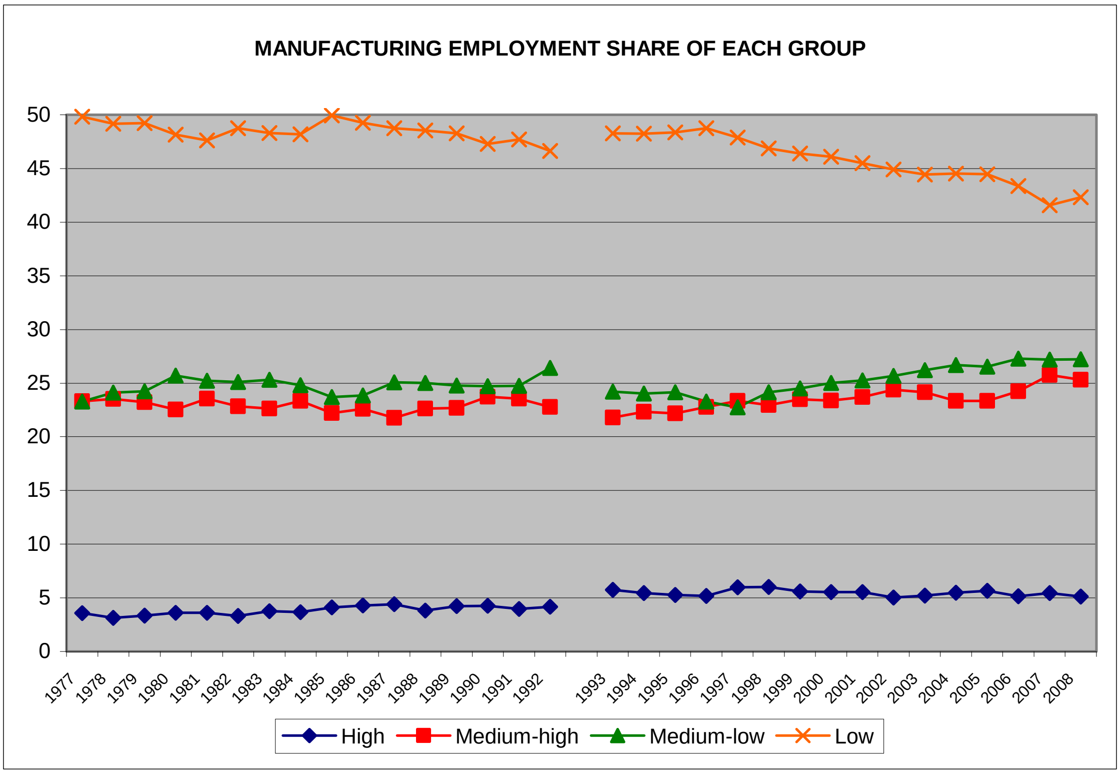


Figure A1. Manufacturing employment share along the period 1977-2008.

---

sector in the latter group. On the other hand, the *INE* classifies part of sector **35** (i.e. aircraft) in the former group and part in the latter. We have decided to include the whole sector in the former.