Location and Localization of Portuguese Manufacturing Industries

Paulo Guimarães
Universidade do Minho and CEMPRE

Octávio Figueiredo
Universidade do Porto and CEMPRE

March 8, 2004
Abstract
This paper offers an empirical investigation into the level of localization of Portuguese manufacturing industries. Using the conceptual framework of Ellison & Glaeser (1997), we develop an alternative method to measure localization. We find that most Portuguese industries exhibit modest levels of localization and that the industries with the highest tendency to cluster are more traditional sectors. By regions, levels of localization are particularly important in the Porto-Lisbon coastal corridor and around Serra da Estrela. We also find that accessibility to the Porto-Lisbon coastal corridor (and transportation costs) is the variable with the highest potential to reshape the Portuguese industrial landscape.
JEL classification: R12; R32
1 Introduction

In the last two decades, the recognition of the importance of agglomeration economies as a source of increasing returns for individual firms has led researchers in international trade, growth and industrial organization to join urban and regional economists in investigating why and to what extent they explain observed spatial concentration of economic activity. There is a wealth of examples from many countries showing that many individual industries tend to be localized (i.e. to concentrate over and above overall economic activity). There are the classic examples of Alfred Marshall for the clustering of cutlery industry in Sheffield and the jewelry industry in Birmingham as well as the more contemporaneous and prominent examples of the high-tech industry in Santa Clara County, California (the Silicon Valley), Boston’s Route 128 in Massachusetts, and Austin, Texas. But are these examples the rule or the exception? And how general and how strong is the tendency of industry to agglomerate?

The debate reignited by the advent of the “new economic geography”, with its emphasis on the importance of external economies, has again brought these questions to the forefront of many scientists’ research agendas. But clear answers to these questions have been marred by the lack of an adequate approach to the measurement of the localization level of an industry. More recently, Ellison & Glaeser (1997) tackled this problem. Based on a Random Utility (Profit) Maximization model of industrial location, they proposed an index that avoids several of the pitfalls of past measures. In a short period of time, their work spawned a significant number of studies and rapidly emerged as the standard approach for measuring the degree of localization of industry. Nevertheless, we contend that the index proposed in Ellison & Glaeser (1997) (henceforth the EG index) has some serious drawbacks. Thus, we elaborate on the work of these two authors and propose an alternative index which is more faithful to the theoretical construct that was exposed in their original work [Ellison & Glaeser (1997)].

In this paper we abstract from the dynamic aspects that lead to changes in the landscape of economic activity overtime. Instead, we try to characterize the level of localization at a given point in time. Using the most recently available data we provide a thorough characterization of the level of localization of the Portuguese industry and query into the mechanisms that underlie the location decisions of firms.

The rest of the paper is comprised of 4 sections. The following section reviews the measures of spatial concentration and attendant problems. In section 3 we explore the methodological questions and develop the alternative method for measuring localization. Section 4 explains the data and presents our main results on localization of Portuguese manufacturing industries. Finally, our last section (section 5) offers some concluding comments and points to policy implications.

2 The Measurement of Spatial Concentration

Past economists had no shortage of tools for measuring geographical concentration. Most prominent are Hoover’s location quotient and the Gini coefficient as applied by Krugman (1991). These measures quantify the discrepancy between the distribution of regional employment in a particular industry against the distribution of overall
manufacturing employment. But are these measures able to capture the concept of agglomeration? A first obvious problem is that they are sensitive to the levels of concentration within the industry. Take as an example two industries which have identical measures for the Gini index. The first industry is composed of many independent firms, all equally sized and located in a single region, while the second industry is composed of just one firm operating a large establishment. The first case agrees more with the notion of agglomeration, which may explain the clustering of all firms in that industry. But for the second industry, it is obvious that the concept of agglomeration does not apply. In the first case, external economies associated with firms’ clustering are most likely the source of geographic concentration, while in the second case the spatial concentration is entirely explained by industrial concentration, and then by returns to scale.

Another obvious problem is that these measures do not account for the inherent randomness of the underlying location decisions. Firms may exhibit some level of spatial concentration by chance. This idea can be easily explained by resorting to the well-known balls and urns example often used in statistics. If one has, say, 10 urns (regions) and 10 balls (firms) and drops the balls at random into these urns then, even though all urns are equally probable, it is very unlikely that we will observe exactly one ball in each urn. Some clustering will necessarily occur and that is perfectly compatible with the idea that the balls were thrown at random (the firms’ decisions were random). The above indexes are not able to control for this type of clustering.

From the above discussion it should be obvious that these indices do not adequately measure an industry’s degree of agglomeration. The recent index proposed in Ellison & Glaeser (1997) to measure localization overcomes the limitations that were just discussed. Like the Gini coefficient it attempts to measure the tendency of one industry to agglomerate in relation to the general tendency of all industries to agglomerate. But, unlike its predecessors, it accounts for the inherent discreteness (lumpiness) that will be observed if location decisions are driven by chance alone and it expurgates the effect of industrial concentration. Thus, it allows for a standardized measure of the industry agglomerative forces which can be readily used for temporal or inter-sectorial comparisons. Most notably, the EG index is rooted in the location choice model of Carlton (1983), which in turn is based on McFadden’s Random Utility Maximization framework (henceforth RUM framework) and has been the workhorse for the empirical literature on industrial location [e.g. Bartik (1985), Luger & Shetty (1985), Hansen (1987), Schmenner, Huber & Cook (1987), Coughlin, Terza & Arromdee (1991), Woodward (1992), Friedman, Gerlowski & Silberman (1992), Head, Ries & Swenson (1995), Guimarães, Figueiredo & Woodward (2000), and Figueiredo, Guimarães & Woodward (2002)]. Other authors have proposed indices which are very closely related to the EG index. Based on a different set of theoretical arguments, Maurel & Sedillot (1999) constructed an index which is similar to the EG index. By comparing the two formulas, they show that the difference between the indices has an expected value of zero. Also noteworthy is the work of Devereux, Griffith & Simpson (2003). They showed that the index of EG can be conveniently approximated by the difference between an index that measures geographic concentration and another that measures industrial concentration.

More recently, Duranton & Overman (2002) have proposed a different approach
to the measurement of spatial concentration. Their approach draws directly from
methods well-known to spatial statisticians to measure concentration of spatial phe-
nomena. They treat space as continuous and compute their measurements based
on the cartesian distances between each pair of plants. Treating space as contin-
uous has an inherent appeal but their approach lacks a theoretical underpinning.
Moreover, it is an essentially descriptive procedure that requires precise information
on the exact location of each business unit.

The new wave of literature initiated with Ellison & Glaeser (1997) has already
generated a substantial amount of applied work. Besides the U.S. [ Ellison & Glaeser
(1997), Dumais, Ellison & Glaeser (2002) and Holmes & Stevens (2002)], studies
characterizing the localization levels of industries have been produced for France
[ Maurel & Sedillot (1999) and Houdebine (1999)], Belgium [ Bertinelli & Decrop
(2002)], UK. [ Devereux et al. (2003)] and Spain [ Callejón (1997)]. Common to all
studies is the finding that the majority of industries are localized.

In the following, we offer an approach to the measurement of industrial localiza-
tion that is grounded more solidly on the RUM framework of industrial location, yet it
borrows from the conceptual approach of Ellison & Glaeser (1997). As will become
obvious in the next section, the link between the RUM location literature and the EG
index is feeble. We show how the two can be better integrated. Also, contrary to the
general trend in the literature, we argue that employment figures are a confounding
factor in the measurement of localization [as proposed by Ellison & Glaeser (1997)]
and advocate the use of counts of plants instead.

3 Methodologic Issues

Industrial location models based on the RUM framework provide an explanation for
the agglomeration of industry - idiosyncratic factors aside, firms choose locations
that yield the highest profits. In practice, most empirical models of industrial location
have been applied to new firm births and accounted for the importance of localization
economies in an indirect way - by introducing variables that measure the presence
of that same industry. If we abstract from the dynamic questions we can use the
theoretical framework of the RUM to justify the agglomeration of an industry. This
was the approach of Ellison & Glaeser (1997). Yet, as argued here, the integration
between the RUM and the derivation of the EG index can be strengthened and thus
it should provide a better way to measure localization of an industry.

3.1 The EG Index

To motivate our approach, we take a more in-depth look at the derivation of the EG
index. Let us assume at the outset that the economy is divided into \( J \) geographical
units (regions). Also, we take as our reference \( a \) given industry which has exactly \( n_j \)
plants located in each region \( j \). Thus, \( n = \sum_{j=1}^{J} n_j \) represents the total number of
existing plants in our reference industry. Next, we briefly sketch how the EG index is
obtained taking as a reference their model of "natural advantages". If firm \( i \) chooses
to locate in region \( j \) then its profits will consist of

\[
\ln \pi_{ij} = \ln \pi_j + \varepsilon_{ij} \tag{1}
\]
where $\pi_j$ is a non-negative random variable reflecting the profitability of locating in area $j$ for a typical firm in the industry. In this formulation of the model, Nature introduces the randomness in $\pi_j$ by selecting for each region the characteristics that make it unique (their natural advantages). $\varepsilon_{ij}$ is a random disturbance. If we assume that $\varepsilon_{ij}$ is an identically and independently distributed random term with an Extreme Value Type I distribution\footnote{In the past this distribution has been referred to by other names such as Weibull, Gumbel and double-exponential [Louviere, Hensher & Swait (2000)].} then, conditional on a realization of $\pi_j$, we can apply McFadden’s (1974) result to obtain,

$$p_{j/n} = \frac{\exp(\ln \pi_j)}{\sum_{j=1}^{J} \exp(\ln \pi_j)} = \frac{\pi_j}{\sum_{j=1}^{J} \pi_j},$$

which denotes the probability of a firm locating in region $j$. Thus, $p_j$ is obtained from the Random (Profit) Utility Maximization framework of Carlton (1983) which, as mentioned earlier, gives support to most recent studies of industrial location. To derive their index, Ellison & Glaeser (1997) introduced two parametric restrictions regarding the expected value and variance of $p_j$. Thus, they assume that the distribution of $\pi_j$ is such that:

$$E(p_j) = x_j,$$

and that,

$$V(p_j) = \gamma x_j (1 - x_j),$$

where $x_j$ may be thought of as the probability of a firm locating in region $j$ in the absence of any region specific advantages for that industry. Thus, the larger the discrepancy between $x_j$ and $p_j$, the larger the influence that these region specific effects (say, natural advantages) play in the location decisions of firms in that industry. That difference is captured by the parameter $\gamma$ (which we will refer to as the EG parameter) which belongs to the unit interval. It is easy to see that if $\gamma = 0$ then the industry will tend to replicate the pattern observed for the $x_j$ (what Ellison and Glaeser call the dartboard model) and we can conclude that there is no spatial concentration in excess of what we would expect to occur. If, however, $\gamma > 0$, then the actual location probabilities of the industry will differ from $x_j$ and in the limit when $\gamma = 1$, each $p_j$ has the largest variance and becomes a Bernoulli random variable. Thus, in the limit, all the investments for that industry would be located in a single region.

Ellison & Glaeser (1997) also show that the $\gamma$ parameter may be derived from an alternative model that emphasizes industrial spillovers as the force leading to "excessive concentration". Whichever theoretical motivation one uses is irrelevant because they are observationally equivalent and lead to the same functional form for the index, the practical implication being that we can not readily distinguish the two sources of geographic concentration (natural advantages and industrial spillovers).

To estimate $\gamma$ for a particular industry they let $x_j$ denote area $j$’s share of total manufacturing employment. Here, the idea is that the model should on average reproduce the overall distribution of manufacturing activity. In a next step they con-
considered the following "raw concentration index" of employment:

\[ G_E = \sum_{j=1}^{J} (s_j - x_j)^2 \]  \hspace{1cm} (5)

where, \( s_j \) denotes area \( j \)'s share of employment in that industry and the \( x_j \)'s are as described above. Now, taking the expected value of \( G_E \) they obtain a function of \( \gamma \) and the authors use that relation to propose an estimator for \( \gamma \). Their proposed estimator for \( \gamma \) (the EG index) is then

\[ \hat{\gamma}_{EG} = \frac{G_E - \left(1 - \sum_{j=1}^{J} x_j^2\right) H_E}{(1 - \sum_{j=1}^{J} x_j^2)(1 - H_E)}, \] \hspace{1cm} (6)

where \( H_E \) is the employment Herfindhal index for the industry and the expected value of \( G_E \) is replaced by its actual value. Note that the computation of the Ellison and Glaeser measure of concentration only requires employment information. In that sense this measure is remarkable because it provides a framework to extract information about the spatial concentration of industry based exclusively on regional employment information.

3.2 A EG Index Based on Counts of Plants

One could as well derive an alternative estimator based on counts of plants. To see this, define

\[ G_F = \sum_{j=1}^{J} \left(\frac{n_j}{n} - x_j\right)^2 \] \hspace{1cm} (7)

and proceeding in a fashion similar to EG (see Appendix A) we derive the following alternative estimator for \( \gamma \):

\[ \hat{\gamma}_A = \frac{nG_F - \left(1 - \sum_{j=1}^{J} x_j^2\right)}{(n - 1) \left(1 - \sum_{j=1}^{J} x_j^2\right)} \] \hspace{1cm} (8)

The above expression is very similar to that of the Ellison-Glaeser index. It replaces the Herfindhal index by \( 1/n \) and the "raw concentration index" is expressed in terms of number of plants instead of employment. Like the estimator proposed by Ellison and Glaeser this estimator for \( \gamma \) is also, by construction, unbiased. Most notably it has a much smaller variance. To see this note that:

\[ \frac{V(\hat{\gamma}_{EG})}{V(\hat{\gamma}_A)} = \left(\frac{n - 1}{n(1 - H_E)}\right)^2 \frac{V(G_E)}{V(G_F)}. \] \hspace{1cm} (9)

An heuristic argument suffices to justify the better efficiency of this estimator. If all plants had the same dimension, the indexes would be identical \( (H_E \) would be \( 1/n \). As the Herfindhal index increases, the first term of the product in the RHS of (9) increases. One would also expect the second term (the ratio of the variances)
to be larger with increases in the Herfindhal index. Thus, we argue that a more
precise estimate for $\gamma$ is obtained if we ignore the confounding influence of plant size
(employment) and work directly with counts of plants. From another perspective,
Holmes & Stevens (2002) provide additional evidence against the use of an index
based on employment plant size. These authors found evidence that plants located
in areas where an industry concentrates (as measured by the EG index) are larger,
on average, than plants in the same industry outside the same area, thus suggesting
that the EG index will tend to overstate the extent of localization economies.

A clear disadvantage of the EG index is that it does not provide an indication of
statistical significance. In Appendix B we show how one can construct and implement
a test for the null hypothesis that $\gamma = 0$ for the $\tilde{\gamma}_A$ statistic.

3.3 An Alternative Method for Measuring Localization

An implicit assumption in the work of Ellison-Glaeser is that in the absence of nat-
aral advantages (or spillover effects) all individual industries would be faced with
the same location probabilities, $p_j (= x_j)$. If these $p_j$s are obtained from the RUM
framework, as is claimed, then this amounts to the underlying assumption that all
industries would have identical profit functions. But, what drives the location of a
chemical plant may be very different from what drives the location of a textile or food
processing plant. In other words, we claim that if natural advantages (or spillovers)
were inexistent then one would still expect to find different patterns of location across
industries, simply because industries value regional characteristics differently. For
example, wages may be an important component of the profit function for the ap-
parel industry but may not be a determinant factor in the locational decision of an
automotive plant. To incorporate this dimension into the framework laid out by Elli-
son and Glaeser, we take a different route - we explicitly model the location decision
process of firms and measure concentration in excess of that which would result if all
industries were influenced by the same set of (observed) locational factors. That is,
instead of approximating the "attractiveness" of a region by its share of manufacturing
employment\footnote{At this point it should be noted that Ellison and Glaeser report the use of other alternatives to manu-
facturing employment such as the area and the population.}, we let each industry have a different valuation for the "attractiveness"
of a region based on the particular combination of factors that are relevant for that
industry.

Hence, we admit that the profit function faced by firm $i$ in our reference industry,
if it decides to locate in region $j$, may be written as,

$$\log \pi_{ij} = \theta^t y_j + \eta_j + \varepsilon_{ij},$$

(10)

where, the $y_j$ are those regional characteristics that affect the location decisions of
firms in all industries (e.g. wages, land costs, market accessibility and transportation
costs), $\theta$ is a vector of parameters, and $\eta_j$ is a (regional) random effect that picks
the unobservable locational advantages of that region for a particular industry. The
other random term, $\varepsilon_{ij}$, is as defined earlier. Now, conditional on the $\eta_j$s and again
drawing on McFadden’s (1974) result we can write,

\[ p_{j/\eta} = \frac{\exp(\theta' y_j + \eta_j)}{\sum_{j=1}^{J} \exp(\theta' y_j + \eta_j)} = \frac{\exp(\eta_j)\lambda_j}{\sum_{j=1}^{J} \exp(\eta_j)\lambda_j}. \]  

(11)

The likelihood function (conditional on the \( \eta_j \)s) implied by the above expression is that of a conditional logit model:

\[ L(n_1, n_2, ..., n_J/n, \eta) = \prod_{j=1}^{J} p_{j/\eta}^{n_j}. \]  

(12)

which in turn is the kernel of a multinomial distribution with parameters \((p_{1/\eta}, p_{2/\eta}, ..., p_{J/\eta}, n)\),

\[ L(n_1, n_2, ..., n_J/n, \eta) \propto n! \prod_{j=1}^{J} \frac{n_j^{n_j}}{n_j!}. \]  

(13)

Now, if we assume that the \( \exp(\eta_j) \)s are iid. gamma distributed with parameters \((\delta^{-1}, \delta^{-1})\) then \( \exp(\eta_j)\lambda_j \) follows a gamma distribution with parameters \((\delta^{-1}\lambda_j, \delta^{-1})\). It follows from Mosimann (1962) that the \((p_1, p_2, ..., p_J)\) are Dirichlet distributed with parameters \((\delta^{-1}\lambda_1, \delta^{-1}\lambda_2, ..., \delta^{-1}\lambda_J)\). Therefore the unconditional likelihood function may be written as,

\[ L(n_1, n_2, ..., n_J/n) = n! \int \prod_{j=1}^{J} \frac{n_j^{n_j}}{n_j!} g(p_1, p_2, ..., p_J) dp_1 dp_2, ..., dp_J. \]  

(14)

The above integral has a closed form, whose solution is known as the Dirichlet-Multinomial distribution (Mosimann 1962):

\[ L(n_1, n_2, ..., n_J/n) = n! \Gamma(\delta^{-1}\lambda_\bullet) \prod_{j=1}^{J} \Gamma(\delta^{-1}\lambda_j n_j) / \Gamma(\delta^{-1}\lambda_\bullet + n) \prod_{j=1}^{J} \Gamma(\delta^{-1}\lambda_j n_j + 1), \]  

(15)

where \( \lambda_\bullet = \sum_{j=1}^{J} \lambda_j \). The resulting likelihood function offers no particular challenge and can be easily implemented. But the interesting feature of this approach is that now,

\[ E(p_j) = \frac{\lambda_j}{\lambda_\bullet}, \]  

(16)

and

\[ V(p_j) = \frac{\lambda_j}{\lambda_\bullet} \left(1 - \frac{\lambda_j}{\lambda_\bullet}\right) \left(\frac{1}{\delta^{-1}\lambda_\bullet + 1}\right). \]  

(17)

and by analogy with (4) and the approach of Ellison and Glaeser we can interpret

\[ \gamma = \frac{1}{\delta^{-1}\lambda_\bullet + 1} = \frac{\delta}{(\lambda_\bullet + \delta)} \]  

(18)

as an index of excessive spatial concentration for that industry, that is, an alternative estimator for the EG parameter. As \( \delta \) (the variance of the region specific random
error) increases, so does $\gamma$ and in the limit when $\delta$ tends to infinity, $\gamma$ will tend to 1. On the other hand, $\gamma$ will approach zero as $\delta$ tends to zero.\footnote{Because we are testing a value which is in the boundary of the set of admissible values for $\delta$, we follow the suggestion in Cameron & Trivedi (1998) and adjust the level of significance of the chi-square statistic accordingly. Also, we should note that to apply the likelihood ratio test, we need to rescale the likelihood function of the Conditional Logit model as in (13).} Because in this latter situation, the Dirichlet-Multinomial distribution collapses to a standard multinomial distribution we can use a likelihood ratio test to test the hypothesis that the industry is more concentrated than what we would expect ($\delta = 0$).\footnote{Unlike the EG index which often produces negative estimates, our estimator will always generate estimates that belong to the unit interval.} To implement this model, we wrote the likelihood function in Stata (version 7) using that package’s standard numerical maximization routine (a modified Newton-Raphson algorithm). To obtain starting values, we first estimated a Poisson regression—which in this context produces the same estimates for the variable coefficients as the conditional logit model \cite{Guimaraes2003}. Convergence was fast with a very small number of iterations (less than 10 for most cases).\footnote{There were however, a few cases (see below) when the model was unable to converge. We took it as evidence that the data were not overdispersed enough, although this procedure may be questionable. For these industries, we assumed that $\gamma$ was zero.}

4 Localization of Portuguese Manufacturing Industries

4.1 Data and Variables

Our main source of data was the "Quadros do Pessoal" database for 1999, the most recent available year. The "Quadros do Pessoal" are a yearly survey collected by the Ministry of Employment for all the existing companies operating in Portugal (except family businesses without wage earning employees) and cover 45,350 plants for the year of 1999.\footnote{For a thorough description of this database see, for example, Mata, Portugal & Guimarães (1995). Unless otherwise noted the "Quadros do Pessoal" was the source for all the information used in this paper.} Using this source, we tallied the number of plants as well as employment for each "concelho" in continental Portugal.\footnote{The concelho is an administrative region in Portugal. In recent years some new concelhos have been created by the incorporation of parts of existing "concelhos". To maintain data compatibility, we use the spatial breakdown of 275 "concelhos" that was still valid in 1997. These have an average area of 322.5 Km2.} We rely on the 3-digit (103 industries) classification of the Portuguese Standard Industrial Classification system (CAE).\footnote{Revision 2 of the CAE.} Using the 275 Portuguese "concelhos" as the spatial choice set, we estimated a location regression for each industry (the Dirichlet-Multinomial model), as well as the corresponding measure of excessive concentration (localization) given by (18).

The choice of regressors for our location model was dictated by location theory. Location theory distinguishes three different sets of factors driving the firm’s location decision problem: external economies, costs of production factors, and accessibility (transportation costs) to input and final demand markets. External economies can arise from two different sources. Localization economies are those external...
economies that result from the spatial concentration of firms of a particular industry in a given region and that are internalized by firms of that particular industry. This is what we are indirectly measuring through \( \gamma \) (along with natural advantages of the regions). The other externality, urbanization economies, accrues from the clustering of general economic activity in a given area and benefits all plants locating in that particular area. Urbanization economies are proxied in our model by the "concelho" density of service and manufacturing establishments per square kilometer in 1999.

To control for the impact of factor prices, we obtained information on the cost of labor and land. Labor costs are measured by an index of the "concelho"s average manufacturing base wage rate in 1999. Since industrial and residential users compete for land, one may argue that when modeling with small areas and controlling for urbanization, as in our case, land costs can be proxied by population density. Consequently, following the suggestion of Bartik (1985), we use population density to approximate land costs. We did not consider the cost of capital because it is practically invariant across alternatives. Interest rates do not differ regionally, and despite some minor differences in municipal taxes, the overall tax burden on manufacturing activity comes mostly from taxes set at the national level.

To account for market accessibility at a given location (and transportation costs) we enter two variables in the model. The drive time distance from each "concelho" to the Porto-Lisbon corridor (the more urbanized coastal side of the country) measures large-scale accessibility, i.e., access to the largest markets. Small-scale accessibility, i.e., access to regional markets, is proxied by the distance in time by road from each "concelho" to the administrative center (the capital) of the related "distrito".

### 4.2 Results

We computed the localization index \( \gamma \) for each of the 3-digit SIC industries at the "concelho" level. As indicated before, for a small number of industries (8) the model did not converge and we assumed that \( \gamma \) was zero in these cases. Additionally, for 17 industries, the \( \gamma \) index was not statistically different from zero at the 95 per cent level of significance. For the remaining 75 industries (75 per cent of the 100 industries analyzed) we find evidence of "excess of concentration" \( \gamma > 0 \). Therefore, a high percentage of Portuguese manufacturing industries appear to be localized, a result that corroborates similar analyses based on the EG index but performed for others.
As previously observed for other countries as well, the degree of localization varies significantly across Portuguese industries. In Figure 1 we show a histogram of $\hat{\gamma}$ at the “concelho” level for the 100 3-digit SIC industries. As can be seen, the localization index displays a very skewed distribution, the majority of industries showing slight levels of localization. Indeed, of the 75 industries whose $\hat{\gamma}$ is significantly above 0 we find that 64 per cent display a degree of localization below the mean value of 0.013.

[insert Figure 1 about here]

In Table 1 we list the 27 3-digit SIC industries that have a $\hat{\gamma}$ index above the industry average. Among them we find a large number of traditional industries for which localization is determined by the historical specialization of particular regions. This is the case of the motorcycles and bicycles industry (SIC 354), the tannery industry (SIC 362), the textile industries (SICs 172, 173, 171, 176 and 177), the footwear industry (SIC 193), the ceramic industries (SICs 264 and 262), and the cork industry (SIC 205). Again, this pattern coincides with evidence for other countries that suggests that typically traditional industries are highly localized.14

Table 1 also shows that several more technologically advanced industries (such as fabrication of material for radio and television apparatus, the automobile industry, several industries that produce machinery and equipments, some primary metal industries, and the pharmaceutical industry), exhibit higher than average levels of localization.

As could be expected, shipbuilding and several industries that process sea products are also among the most localized industries. Access to a natural resource, the sea, is crucial to explain firms’ location decisions in this latter type of industries.

As these three different groups of localized industries suggest, high levels of localization can correspond to different location strategies. Past static externalities dominate in the case of the industrial clustering of firms in the traditional sectors. Current dynamic knowledge spillovers can explain the location strategy of the more technologically advanced industries. Natural advantages of the regions prevail in industries such as shipbuilding.

[insert Tables 1 and 2 about here]

Table 2 displays the group of 25 3-digit SIC sectors that are non localized industries according to our results ($\hat{\gamma} = 0$). As can be seen, the majority of these industries are capital-intensive industries for which returns to scale are a crucial factor. For this last group it is important to distinguish our measure of localization from a simple measure of geographic concentration. While some of these sectors (such as tobacco, petroleum refining or aircraft and space vehicles fabrication) can be highly

13Ellison & Glaeser (1997) found that 446 out of 459 4-digit SIC industries in the United States were localized. Maurel & Sedillot (1999) found that 77% of the 273 4-digit French industries display “excess of concentration”. Similar results were found for the UK by Devereux et al. (2003).

14See Ellison & Glaeser (1997), Table 4, Maurel & Sedillot (1999), Tables 1 and 2, Devereux et al. (2003), Tables 4 and 6, and Krugman (1991), Appendix D.
Location and Localization of Portuguese Manufacturing Industries

concentrated in space, this concentration is almost entirely explained by industrial concentration, and thus by returns to scale rather than natural advantages or external economies associated with firms’ clustering.

4.3 Comparison with results based on the EG Index

We now compare our estimates of localization with those provided by the EG index ($\hat{\gamma}_{EG}$) and the alternative EG index based on counts of plants ($\hat{\gamma}_{A}$). If we look at the extent of localization across 3-digits sectors, we find very similar results for the three measures. 68, 75, and 80 per cent of the industries exhibit “excess of concentration” according to the modified EG index, our index, and the EG index, respectively. In Figure 2 we display the box-whisker plots computed across the 100 3-digit SIC industries for the three measures. To increase readability the graph omits a few extreme (high) values for each one of the distributions. As can be seen, all distributions show the same pattern of skewness with increasing interquartile ranges. Nevertheless, as we anticipated, our proposed measure of localization (labeled as DM index in the figure) exhibits much smaller values when compared with the EG index ($\hat{\gamma}_{EG}$) and the modified EG index ($\hat{\gamma}_{A}$). We take these results as confirmatory evidence that the original EG index tends to overstate the degree of localization of industry.

The three measures also encompass substantial differences when we look at the hierarchy of individual industries. The Spearman rank correlation coefficients between $\hat{\gamma}$ and $\hat{\gamma}_{EG}$ (0.35) and between $\hat{\gamma}$ and $\hat{\gamma}_{A}$ (0.49) are small, in spite of the fact that we reject independence between any of three statistics. Furthermore, as a quick inspection of Table 1 will reveal, among the top 27 most localized industries according to $\hat{\gamma}$ we only find 14 (18) of the industries for a similar ranking based on $\hat{\gamma}_{EG}$ ($\hat{\gamma}_{A}$).

4.4 Extensions

To compute our localization index we had to explicitly model the location decision process of firms. This, in turn, allows for a much richer analysis than that obtained using the methodology of Ellison and Glaeser. In this section, we extend our research into two distinct directions. First, we measure the overall levels of localization by regions and try to identify which regions are associated with localized industries. Next, we set out to gain some understanding into the working of the mechanisms that influence the location decisions of firms.

4.4.1 Regional Analysis

The standardized residuals from the Dirichlet-Multinomial regressions (Pearson residuals) provide useful information because they may be regarded as an indirect estimate of the level of localization economies (and natural advantages) for each region and for each specific industry. The analysis of these residuals allows us to link industries with regions. This enables us to find the regions that are associated with
the more localized industries. We can go one step further and construct an index of overall localization for each region by aggregating the residuals for all sectors.

Regions Associated with the Most Localized Industries In this subsection we use the standardized residuals from the Dirichlet-Multinomial regressions to map out those regions that are associated with the 27 most localized industries shown on Table 1. To do this, we identify, for each of these industries, all residuals that exceed 1.96.\textsuperscript{15} A summary of the information is presented in Table 3. For the group of most localized traditional industries, we found that using the residuals to analyze "excess of concentration" resulting from localization economies makes sense. Indeed, these residuals show, for each of the localized traditional industries, the "concelhos" that one would expect to find based on knowledge about the historical patterns of specialization of Portuguese regions. As expected, the highest residuals for the motorcycles and bicycles industry (SIC 354) are found for the Âgueda and Anadia "concelhos"; the tannery industry (SIC 191) is localized at Alcanena and the largest spatial concentration of firms in the jewelry industry (SIC 362) is found for Gondomar; the textile industry (SICs 172, 173, 171, 176 and 177) is localized, as expected, in the "concelhos" of the Ave Valley (Guimarães, Santo Tirso, and VN. Famalicão), in the region around Serra da Estrela (Covilhã, Gouveia, Seia, Guarda, and Castelo Branco), and in the Barcelos "concelho"; the highest residuals for the footwear industry (SIC 193) are found for Oliveira de Azeméis, Felgueiras, Santa Maria da Feira and Guimarães, well known regions of clustering of firms for the Portuguese footwear industry; this is also the case of the "concelhos" of Barcelos and Alcobaça, for the ceramic industries (SIC262), and Santa Maria da Feira and Montijo, for the cork industry (SIC 205).

As indicated before, some of the more technologically advanced industries also present levels of localization above the mean. Table 3 indicates for these industries the associated "concelhos". We found significant clusters of firms in the automobile industries (SICs 341 and 343) in three different regions: in the north side of the country, around the cities of Porto and Aveiro (VN. Gaia, Maia, Oliveira de Azeméis, Ovar, and Águeda); in the central part of the country, around the cities of Viseu and Coimbra (Mangualde and Cantanhede); and, more to the south, around Lisbon (Rio Maior, Abrantes, Palmela, and Loures). The pharmaceutical industry (SIC 244) follows a similar pattern with "excessive concentration" of firms around Porto (Penafiel and VN. Famalicão); around Coimbra and Viseu (Penela, Coimbra, Condeixa-a-Nova, Oliveira de Frades, Tondela, and Mortágua); and around the city of Lisbon (Sintra). The largest significant clusters for SIC 323 (Fabrication of material for radio and television apparatus) are found for Mafra and Proença-a-Nova. For industries that produce machinery and equipments (SICs 291 and 295), "excess of concentration" resulting from localization economies is particularly important in Marinha Grande, Braga, Leiria, Oliveira de Azeméis, and Seixal.

Table 3 also makes patent the importance of natural advantages for the shipbuilding industry (SIC 351), as well as for several industries that process sea products.

\textsuperscript{15}The justification for this approach lies in the fact that the distribution of Pearson residuals for a Poisson distribution with moderately sized counts is known to approximate the standardized normal distribution.
Location and Localization of Portuguese Manufacturing Industries

(SIC 152). These are localized, as expected, in “concelhos” pertaining to the coastal side of the country.

**Overall Levels of Localization by Region**  The standardized residuals from the Dirichlet-Multinomial regressions can also be used to measure the overall level of localization by regions. To gain some insight into this question we computed a new variable as the sum (by “concelho”) of all positive standardized residuals across 3-digit SIC industries. In figure 3 we show these data for the 275 Portuguese “concelhos”. The “concelhos” are classified according to the quartiles of the variable, with darker intensity corresponding with increasing values. As can be seen, “excess of concentration” is particularly important in the coastal corridor between Porto and Lisbon. More precisely, in this axis we found three different areas where localization economies (natural advantages of the regions) are important: a large stretch of territory (consisting of 23 adjacent “concelhos”) is found around the cities of Porto and Aveiro; a second large area is centered around the cities of Coimbra and Leiria; and a third but small one (grouping 8 “concelhos”) can be detected around the city of Lisbon.

Figure 2 also shows that levels of spatial concentration explained by natural advantages or external economies associated with firms’ clustering are high for a large area in the central hinterland of the country. This large area (around Serra da Estrela) encompasses the “concelhos” of Viseu, Tondela, Mangularde, Guarda, Covilhã, Fundão, Seia, and Castelo Branco.

Beyond the Porto-Lisbon coastal corridor and the central hinterland of the country around Serra da Estrela, there are a few isolated “concelhos” which also stand out. This is the case of the “concelhos” of Chaves and Mirandela (in the northeast part of the country), Elvas and Évora (in the Alentejo province), and Loulé, Silves and Faro (in the southern part of the country).

**4.4.2 Evaluating the Impact of Changes on Location Factors**

Given that we explicitly model the location decision process of firms, we are able to perform exercises of comparative statics to determine what changes would be obtained under an alternative scenario for the allocation of the regional resources. The results of this analysis will shed some light into the effect of potential policies aimed at influencing the location decisions of firms. Consider first the situation where there is a 1 per cent increase in the level of one of the variables entering the profit function, say variable $k$, in region $j$, everything else remaining constant. The impact of that change on the probability of locating in that region for a given industry amounts to $\theta_k p_{j/\eta} (1 - p_{j/\eta})$. Two obvious implications follow. First, the impact of the change is higher for the variable with the highest profit elasticity ($\theta$) in absolute value. The second implication is that that impact will be higher for the more attractive regions (i.e. regions with higher $p_{j/\eta}$) and even higher in regions that benefit from localization.

---

16Note that in our specification all variables are in logarithmic form.
17Unless $p_{j/\eta} > 0.5$, a very unlikely situation.
economies or natural advantages.\textsuperscript{18} It is straightforward from here to infer that an across the board increase of 1 per cent in the level of a given factor will impact differently across regions and thus will change the spatial distribution of that industry.

But are we allowed to conclude that the level of localization (i.e. excessive concentration) of that industry will increase? To answer that question we need to measure the impact on the localization index $\tilde{\gamma}$ resulting from a change in one of the variables entering in the firms’ profit function. It is obvious that anything that will directly increase industry profits will reduce the weight that localization economies (or natural advantages) have on driving the location decisions of firms. This can be ascertained by looking at expression (18). If we compute the elasticity of $\tilde{\gamma}$ with respect to one of the variables entering the profit equation, say variable $k$, we obtain $-\theta_k(1 - \tilde{\gamma})$ (again we are taking into account the fact that all explanatory variables are already entered in logarithmic form). Thus, those variables that are more capable of affecting profits (with higher profit elasticities) are precisely the ones that offer the highest potential to counterbalance the effects of local spillovers and natural advantages of the regions. This means that if wages have the highest profit elasticity (assumed negative) then a decrease of 1 per cent in the average cost of the workforce across regions will increase profits everywhere and will diminish the relative importance of localization economies and natural advantages leading to a smaller level of “excessive concentration”, more than an equivalent percentual change in any of the other factors affecting profits. But, on the other hand, we can see in the above expression for the elasticity of $\tilde{\gamma}$, that the impact of any change is smaller for those industries that are more localized.

With that in mind, it becomes relevant to identify for each of the 3-digit SIC industries those factors with the highest profit elasticities. In Figure 4 we summarize the results of our calculations.\textsuperscript{19} We find that wages (with a negative sign and thus capturing the cost of the workforce) have the highest profit elasticity for 15 industries (out of 92) while wages (with a positive sign and thus more likely to proxy the quality of the workforce) have the largest significant elasticity for 17 industries. Land costs and urbanization economies are the variables with higher significant impact for 3 and 9 industries, respectively. On the other hand, large-scale accessibility has the largest (negative and significant) impact for 38 industries in contrast with small-scale accessibility which is more relevant for only 3 industries. Thus, it seems fair to conclude that accessibility to the Porto-Lisbon coastal corridor (and transportation costs) is the factor with the highest potential to reshape the Portuguese industrial landscape.

To gain additional insight into this matter we computed the mean impact on $\tilde{\gamma}$ across 3-digit SIC industries resulting from a 10 per cent increase (across regions)

\textsuperscript{18}To get an idea about the extent of this difference across regions we evaluated, for a non-localized industry (SIC 311), the above expression for all the 275 “concelhos”. Next, we computed the ratio of the average value between the upper quartile (the more attractive regions) and the lower quartile of the distribution (the least attractive regions). We find a ratio of 48 indicating that a change of 1% in a given variable for the more attractive regions will have the same redistributive effect as a change of 48% in the least attractive regions. For localized industries that value will be much higher.

\textsuperscript{19}Elasticities were computed for 92 3-digit SIC industries. As indicated before, for a small number of industries (8) the model did not converge.
for each one of the variables entering in the profit equation. The results are shown in Figure 5. As can be seen, large-scale accessibility is the variable with the highest average impact. A 10 per cent increase in this variable across regions results in a 13.73 per cent average increase of $\gamma$ across 3-digit SIC industries, while the same averages for wages, land costs, urbanization economies and small-scale accessibility are 6.64, 2.05, 4.78, and 0.04 per cent, respectively. These results have relevant policy implications. Abstracting from issues pertaining to the cost of alternative policies, it seems that if public authorities want to reshape the industrial landscape and spread industries across the country, they are more likely to succeed if they invest in transportation infrastructures in order to reduce travel time distance (and improve accessibility) from "concelhos" to the Porto-Lisbon coastal corridor.

5 Conclusion

In this paper we have offered an empirical investigation into the level of localization of Portuguese manufacturing industries. We take a critical view of the existing methods for measuring localization and argue that despite its popularity, the EG index has some serious pitfalls that prevent it from being a satisfactory approach. Building on the McFadden's Random Utility (Profit) Maximization framework, we develop an alternative index which is more consistent with the theoretical construct underlying the work of Ellison and Glaeser. With our approach, we are able to simultaneously compute the locational probabilities and the localization index.

When we applied our methodology to Portuguese data, we found evidence that a high percentage of industries are localized, a result that corroborates previous analyses based on the EG index for others countries. But even though Portuguese industries are in general excessively concentrated, the majority of industries exhibit modest localization levels and only a few sectors display a strong tendency to cluster. These sectors are mostly dominated by traditional industries for which localization is determined by the historical specialization of particular regions in Portugal. Past static externalities may justify the clustering of firms in these sectors. We also find that some more technologically advanced industries exhibit high or moderate localization levels. This result indicates that current dynamic knowledge spillovers may also be important in explaining the clustering strategies of Portuguese firms. As expected, some resource oriented industries are found to be highly localized.

We have also precisely identified regions associated with the most localized industries and measured levels of localization by regions. According to our results, spatial concentration induced by external economies and natural advantages of the regions are particularly important in the Porto-Lisbon coastal corridor and in the central hinterland of the country around Serra da Estrela. Our model also allowed us to draw some conclusions regarding economic policy. First, we should note that the stimulation of economic activity through firm creation requires a much larger effort for those regions with a less favorable set of locational factors. Second, we note that localized industries are less likely to respond to policies aimed at redistributing economic activity. And finally, we find that accessibility to the more urbanized coastal
side of the country (and transportation costs) is the variable with the highest potential to reshape the Portuguese industrial landscape, counterbalancing the weight that localization economies and natural advantages of the regions have on driving the location decisions of firms. Thus, if public authorities want to disperse industries across the country and attract private investment to new localities, they are more likely to succeed if they direct their efforts toward the improvement of transportation infrastructures aimed at reducing the travel time distance (and improve accessibility) from “concelhos” to the Porto-Lisbon coastal corridor.
References


APPENDIX A

In the context of the EG model, the number of investments of a given industry in region \( j \), conditional on the total number of investments in the industry, and on the vector of locational probabilities \( (p = p_1, p_2, ..., p_J) \), follows a binomial law with parameters:

\[
E(n_j / p) = np_j \\
V(n_j / p) = np_j(1 - p_j)
\]

We now define a raw index of concentration as:

\[
G_F = \sum_{j=1}^{J} \left( \frac{n_j}{n} - x_j \right)^2
\]

Expanding terms we obtain

\[
G_F = \frac{1}{n^2} \sum_{j=1}^{J} n_j^2 + \sum_{j=1}^{J} x_j^2 - \frac{2}{n} \sum_{j=1}^{J} n_j x_j
\]

and the expected value of the above equation gives:

\[
E(G_F / p) = \frac{1}{n^2} E \left( \sum_{j=1}^{J} n_j^2 \right) + \sum_{j=1}^{J} x_j^2 - \frac{2}{n} E \left( \sum_{j=1}^{J} n_j x_j \right)
\]

\[
= \frac{1}{n^2} \left( \sum_{j=1}^{J} np_j - np_j^2 + n^2 p_j^2 \right) + \sum_{j=1}^{J} x_j^2 - 2 \sum_{j=1}^{J} p_j x_j
\]

Applying the law of iterated expectations we get,

\[
E(G_F) = \frac{1}{n} + \frac{(n-1)}{n} E \left( \sum_{j=1}^{J} p_j^2 \right) + \sum_{j=1}^{J} x_j^2 - 2 \sum_{j=1}^{J} p_j x_j
\]

\[
= \frac{1}{n} + \frac{(n-1)}{n} \gamma \sum_{j=1}^{J} x_j^2 - 2 \sum_{j=1}^{J} p_j x_j
\]

\[
= \frac{1 + \gamma(n-1)}{n} \left( 1 - \sum_{j=1}^{J} x_j^2 \right)
\]

and, as in Ellison & Glaeser (1997), the estimator for \( \gamma \) is obtained by replacing the \( E(G_F) \) by the observed value of \( G_F \) and solving for \( \gamma \). The proposed estimator is:

\[
\hat{\gamma}_A = \frac{nG_F - \left( 1 - \sum_{j=1}^{J} x_j^2 \right)}{(n-1) \left( 1 - \sum_{j=1}^{J} x_j^2 \right)}.
\]
APPENDIX B

Under the null hypothesis that $\gamma = 0$, the $p_j = x_j$ for all $j$ and the observed spatial distribution of the investments for the particular industry follows a multinomial distribution,

$$P(n_1, n_2, ..., n_J) = n! \prod_{j=1}^{J} \frac{x_j^{n_j}}{n_j!}$$

Because we can associate a probability of occurrence to each possible distribution of the $n$ investments we may also construct a distribution for the estimator of $\hat{\gamma}_A$ under the null hypothesis that $\gamma = 0$. To do this, we may simply enumerate all possible values of the multinomial distribution. A simple example will help understand the argument. Suppose that we have 3 regions and 4 investments. Admit for the moment that $(x_1 = x_2 = x_3 = 1/3)$. The next table lists all possible spatial distributions of these investments, the associated probability and the estimated concentration index ($\hat{\gamma}_A$):

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$\hat{\gamma}_A$</th>
<th>$P(n_1, n_2, n_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.2346%</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td>4.9383%</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
<td>4.9383%</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0.00</td>
<td>14.8148%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-0.25</td>
<td>7.4074%</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0.00</td>
<td>14.8148%</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0.25</td>
<td>4.9383%</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-0.25</td>
<td>7.4074%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-0.25</td>
<td>7.4074%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0.25</td>
<td>4.9383%</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1.00</td>
<td>1.2346%</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0.25</td>
<td>4.9383%</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0.00</td>
<td>14.8148%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0.25</td>
<td>4.9383%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1.00</td>
<td>1.2346%</td>
</tr>
</tbody>
</table>

This information can be used to construct the distribution for $\hat{\gamma}_A$ which simply aggregates all common estimates and their probability. Thus, the distribution of $\hat{\gamma}_A$ given $x_1 = x_2 = x_3 = 1/3$, $n = 4$, and $\gamma = 0$ is:

<table>
<thead>
<tr>
<th>$\hat{\gamma}_A$</th>
<th>$f(\hat{\gamma}_A)$</th>
<th>$P(\hat{\gamma}_A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.25</td>
<td>44.44%</td>
<td>44.44%</td>
</tr>
<tr>
<td>0.00</td>
<td>22.22%</td>
<td>66.67%</td>
</tr>
<tr>
<td>0.25</td>
<td>29.63%</td>
<td>96.30%</td>
</tr>
<tr>
<td>1.00</td>
<td>3.70%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

From this simple example, we can see that if we had obtained an estimate of 1 for $\gamma$ we could be fairly confident that $\gamma > 0$, given that the probability of that happening...
was only 3.7 per cent. But any other estimate would be a plausible outcome if the true value of $\gamma$ were 0. Using this approach, we can test the probability that $\gamma = 0$ for any given number of investments and vector of locational probabilities.

However, it is not always feasible to construct the distribution of $\hat{\gamma}_A$ by numerically evaluating all possible distributions of investments by regions (as we did in Table A.1). The number of terms that will need to be computed amounts to $\binom{n + J - 1}{J - 1}$. If, for example, $n = 20$ and $J = 10$, then we get 10,015,005 different cases. If $n$ is increased to 40 we will have 2,054,455,634 different cases. In this case, instead of computing the exact distribution we will randomly sample from this known distribution and generate an empirical cumulative distribution function for $\hat{\gamma}_A$. Thus, in our application we will test our hypothesis for each sector by generating a large number of draws (say 10,000) from a multinomial distribution with parameters $(n; x_1, x_2, \ldots, x_J)$. For each one of these samples, we will compute an estimate of $\gamma$ and the value reported for our test will be the value of the empirical cumulative distribution evaluated at the observed value for $\hat{\gamma}_A$. 

Desenvolvimento Económico Português no Espaço Europeu
Figure 1: Histogram of $\gamma$ at the "concelho" level
Table 1: Geographic Concentration, by Most Localized Industries According to γ

<table>
<thead>
<tr>
<th>3-digit SIC Industry (Portuguese CAE-Rev2)</th>
<th>γ</th>
<th>Number of Plants</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>354- Fab. de motociclos e bicicletas</td>
<td>0.126</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>191- Curtimenta e acabamento de peles sem pelo</td>
<td>0.115</td>
<td>110</td>
<td>2</td>
</tr>
<tr>
<td>362- Fab. de joalharia. ourivesaria e artigos similares</td>
<td>0.060</td>
<td>561</td>
<td>3</td>
</tr>
<tr>
<td>172- Tecelagem de têxteis</td>
<td>0.052</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>173- Acabamento de têxteis</td>
<td>0.048</td>
<td>275</td>
<td>5</td>
</tr>
<tr>
<td>193- Indústria do Calçado</td>
<td>0.048</td>
<td>1932</td>
<td>6</td>
</tr>
<tr>
<td>171- Preparação e fiação de fibras têxteis</td>
<td>0.048</td>
<td>226</td>
<td>7</td>
</tr>
<tr>
<td>351- Construção e reparação naval</td>
<td>0.038</td>
<td>155</td>
<td>8</td>
</tr>
<tr>
<td>323- Fab. de aparelhos receptores e material de rádio e de televisão. etc.</td>
<td>0.037</td>
<td>28</td>
<td>9</td>
</tr>
<tr>
<td>176- Fab. de tecidos de malha</td>
<td>0.036</td>
<td>284</td>
<td>10</td>
</tr>
<tr>
<td>335- Fab. de relógios e de material de relojoaria</td>
<td>0.032</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>247- Fab. de fibras sintéticas ou artificiais</td>
<td>0.029</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>152- Ind. transformadora da pesca e da aquacultura</td>
<td>0.029</td>
<td>106</td>
<td>13</td>
</tr>
<tr>
<td>192- Fab. de artigos de viagem e de uso pessoal</td>
<td>0.029</td>
<td>244</td>
<td>14</td>
</tr>
<tr>
<td>341- Fab. de veículos automóveis</td>
<td>0.027</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>177- Fab. de artigos de malha</td>
<td>0.026</td>
<td>715</td>
<td>16</td>
</tr>
<tr>
<td>275- Fundição de metais ferrosos e não-ferrosos</td>
<td>0.025</td>
<td>154</td>
<td>17</td>
</tr>
<tr>
<td>313- Fab. de fios e cabos isolados</td>
<td>0.021</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>205- Fab. de outras obras de madeira e ind. da cortiça</td>
<td>0.020</td>
<td>1197</td>
<td>19</td>
</tr>
<tr>
<td>264- Fab. de tijolos. telhas e outros</td>
<td>0.019</td>
<td>197</td>
<td>20</td>
</tr>
<tr>
<td>262- Fab. de produtos cerâmicos diversos</td>
<td>0.018</td>
<td>685</td>
<td>21</td>
</tr>
<tr>
<td>332- Fab. de aparelhos de medida. verificação. controlo. navegação e outros</td>
<td>0.018</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>343- Fab. de componentes e acessórios para veículos auto e seus motores</td>
<td>0.015</td>
<td>182</td>
<td>23</td>
</tr>
<tr>
<td>291- Fab. de máquinas e equipamentos diversos</td>
<td>0.015</td>
<td>117</td>
<td>24</td>
</tr>
<tr>
<td>244- Fab. de produtos farmacêuticos</td>
<td>0.015</td>
<td>101</td>
<td>25</td>
</tr>
<tr>
<td>202- Fab. de folheados. contraplacados. painéis. etc.</td>
<td>0.014</td>
<td>37</td>
<td>26</td>
</tr>
<tr>
<td>295- Fab. de outras máquinas e equipamento para uso específico</td>
<td>0.014</td>
<td>831</td>
<td>27</td>
</tr>
</tbody>
</table>
Table 2: Geographic Concentration, by Non Localized Industries According to $\gamma$

<table>
<thead>
<tr>
<th>3-digit SIC Industry</th>
<th>$\gamma$</th>
<th>Number of Plants</th>
<th>Rank</th>
<th>$\gamma$</th>
<th>$\gamma_A$</th>
<th>$\gamma_{EG}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>160- Indústria do tabaco</td>
<td>0.000</td>
<td>2</td>
<td>76</td>
<td>69</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>355- Fab. de outro material de transporte, ne.</td>
<td>0.000</td>
<td>4</td>
<td>76</td>
<td>69</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>296- Fab. de armas e munições</td>
<td>0.000</td>
<td>7</td>
<td>76</td>
<td>69</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>363- Fab. de instrumentos musicais</td>
<td>0.000</td>
<td>8</td>
<td>76</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>283- Fab. de geradores de vapor (excepto caldeiras para aquecimento central)</td>
<td>0.000</td>
<td>8</td>
<td>76</td>
<td>69</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>242- Fab. de pesticidas e de outros produtos agro-químicos</td>
<td>0.000</td>
<td>10</td>
<td>76</td>
<td>69</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>314- Fab. de acumuladores e de pilhas eléctricas</td>
<td>0.000</td>
<td>11</td>
<td>76</td>
<td>69</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>271- Siderurgia e fabricação de ferro</td>
<td>0.000</td>
<td>13</td>
<td>76</td>
<td>69</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>232- Fab. de produtos petrolíferos refinados</td>
<td>0.000</td>
<td>13</td>
<td>76</td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>364- Fab. de artigos de desporto</td>
<td>0.000</td>
<td>13</td>
<td>76</td>
<td>69</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>353- Fab. de aeronaves e de veículos espaciais</td>
<td>0.000</td>
<td>13</td>
<td>76</td>
<td>69</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>272- Fab. de tubos</td>
<td>0.000</td>
<td>15</td>
<td>76</td>
<td>69</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>333- Fab. de equipamentos de controlo de processos industriais</td>
<td>0.000</td>
<td>15</td>
<td>76</td>
<td>69</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>223- Reprodução de suportes gravados</td>
<td>0.000</td>
<td>16</td>
<td>76</td>
<td>27</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>352- Fab. e reparação de material ferroviário</td>
<td>0.000</td>
<td>20</td>
<td>76</td>
<td>69</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>322- Fab. de aparelhos de rádio, televisão, telefonia e telegraafia</td>
<td>0.000</td>
<td>23</td>
<td>76</td>
<td>69</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>183- Preparação e fabricação de artigos de peles com pêlo</td>
<td>0.000</td>
<td>27</td>
<td>76</td>
<td>69</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>334- Fab. de material óptico, fotográfico e cinematográfico</td>
<td>0.000</td>
<td>28</td>
<td>76</td>
<td>21</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>268- Fab. de outros produtos minerais não metálicos</td>
<td>0.000</td>
<td>29</td>
<td>76</td>
<td>69</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>365- Fab. de jogos e brinquedos</td>
<td>0.000</td>
<td>29</td>
<td>76</td>
<td>69</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>273- Outras actividades da primeira transformação de ferro e do aço</td>
<td>0.000</td>
<td>30</td>
<td>76</td>
<td>69</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>371- Reciclagem de sucata e de desperdícios metálicos</td>
<td>0.000</td>
<td>37</td>
<td>76</td>
<td>69</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>265- Fab. de cimento, cal e gesso</td>
<td>0.000</td>
<td>56</td>
<td>76</td>
<td>38</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>263- Fab. de azulejos, ladrilhos, etc.</td>
<td>0.000</td>
<td>57</td>
<td>76</td>
<td>30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>311- Fab. de motores, geradores e transformadores eléctricos</td>
<td>0.000</td>
<td>83</td>
<td>76</td>
<td>69</td>
<td>65</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Box-Whisker plots for the three localization indexes.
Table 3: "Concelhos" Associated with the 27 Most Localized Industries

<table>
<thead>
<tr>
<th>3-digit SIC Industry</th>
<th>Concelhos and Standardized Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>354- Fab. de motociclos e bicicletas</td>
<td>Águeda (17.61); Anadia (4.01)</td>
</tr>
<tr>
<td>191- Curtimenta e acabamento de peles sem pelo</td>
<td>Alcanena (19.46)</td>
</tr>
<tr>
<td>362- Fab. de joalharia, ourivesaria e artigos similares</td>
<td>Gondomar (16.92); Póvoa de Lanhoso (2.72)</td>
</tr>
<tr>
<td>172- Tecelagem de têxteis</td>
<td>Guimarães (10.65); Covilhã (6.20); V. N. Famalicão (4.18); Castelo Branco (2.71)</td>
</tr>
<tr>
<td>173- Acabamento de têxteis</td>
<td>Guimarães (7.05); Santo Tirso (4.90); Barcelos (4.41); V. N. Famalicão (3.69)</td>
</tr>
<tr>
<td>193- Indústria do Calçado</td>
<td>Oliveira de Azeméis (13.01); Felgueiras (9.96); S. M. da Feira (5.50); Guimarães (2.92)</td>
</tr>
<tr>
<td>171- Preparação e fição de fibras têxteis</td>
<td>Guimarães (8.21); V. N. Famalicão (4.88); Santo Tirso (4.84); Covilhã (4.24); Barcelos (3.63); Gouveia (3.01); Seia (2.21); Guarda (2.20)</td>
</tr>
<tr>
<td>351- Construção e reparação naval</td>
<td>Seixal (4.53); Almada (4.46); Aveiro (3.31); Ilhavo (2.57); Palmela (2.54); Setúbal (2.35); Peniche (2.35); Olhão (2.09)</td>
</tr>
<tr>
<td>323- Fab. de aparelhos receptores e material de rádio e de televisão, etc.</td>
<td>Mafra (6.85); Proença-a-Nova (4.85); Viseu (3.22); Sintra (2.96); Ovar (2.69); Sesimbra (2.61); Braga (2.47); V. N. de Ourém (2.44); Alcobaca (2.23)</td>
</tr>
<tr>
<td>176- Fab. de tecidos de malha</td>
<td>Barcelos (11.23); Guimarães (6.43); V. N. Famalicão (3.37); Marco de Canaveses (2.39)</td>
</tr>
<tr>
<td>335- Fab. de relógios e de material de relojoaria</td>
<td>Alcobaca (8.79); Cantanhede (7.52); Guimarães (4.69); Mira (3.43); Braga (2.07); V. N. Famalicão (2.02)</td>
</tr>
<tr>
<td>247- Fab. de fibras sintéticas ou artificiais</td>
<td>Portalegre (19.75); Vale de Cambra (3.79); Espinho (3.56); V. N. de Ourém (2.62); Ovar (2.61); Barreiro (2.05)</td>
</tr>
</tbody>
</table>

continued on next page...
<table>
<thead>
<tr>
<th>Code</th>
<th>Industry Description</th>
<th>Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>152</td>
<td>Ind. transformadora da pesca e da aquacultura</td>
<td>Ilhavo (9.65); Peniche (6.91); Matosinhos (5.03); Vila do Conde (4.59);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Olhão (3.43); Figueira da Foz (2.35)</td>
</tr>
<tr>
<td>192</td>
<td>Fab. de artigos de viagem e de uso pessoal</td>
<td>Alcobaça (10.40); Gondomar (6.70); Braga (4.94); S. M. da Feira (2.71);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V. N. de Gaia (2.49)</td>
</tr>
<tr>
<td>341</td>
<td>Fab. de veículos automóveis</td>
<td>Manguadel (14.70); Ovar (7.19); Rio Maior (2.76); Abrantes (2.48)</td>
</tr>
<tr>
<td>177</td>
<td>Fab. de artigos de malha</td>
<td>Barcelos (13.83); Guimarães (5.22); Fafe (5.11); Alcanena (4.67); V. N.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Famalicão (2.68); Santo Tirso (2.20)</td>
</tr>
<tr>
<td>275</td>
<td>Fundição de metais ferrosos e não-ferrosos</td>
<td>Águeda (8.70); Gondomar (5.58); Braga (4.07); V. N. de Gaia (3.65);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chaves (2.27)</td>
</tr>
<tr>
<td>313</td>
<td>Fab. de fios e cabos isolados</td>
<td>Elvas (4.39); Ovar (4.31); Castelo Branco (4.05); Sintra (4.03); V. N.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>de Gaia (3.77); Guarda (3.53); Esposende (2.14)</td>
</tr>
<tr>
<td>205</td>
<td>Fab. de outras obras de madeira e ind. da cortiça</td>
<td>S. M. da Feira (43.18); Montijo (3.74)</td>
</tr>
<tr>
<td>264</td>
<td>Fab. de tijolos, telhas e outros</td>
<td>Tavira (8.43); Santarém (5.40); Torres Vedras (4.53); Porto de Mós (4.39);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Águeda (4.36); Pombal (3.29); Chaves (2.29); Oliveira do Bairro (2.03)</td>
</tr>
<tr>
<td>262</td>
<td>Fab. de produtos cerâmicos diversos</td>
<td>Barcelos (20.32); Alcobaça (10.24); Reguengos de Monsarás (5.15); Porto</td>
</tr>
<tr>
<td></td>
<td></td>
<td>de Mós (3.73); Caldas das Rainha (3.15); Nazaré (2.37); Mafra (2.20);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Redondo (2.00)</td>
</tr>
<tr>
<td>332</td>
<td>Fab. de aparelhos de medida, verificação, controlo, navegação e outros</td>
<td>Mirandela (7.03); V. N. Famalicão (4.87); Sintra (4.04); Loures (3.74);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Albergaria-a-Velha (3.08); Lagoa (2.92); Coimbra (2.30)</td>
</tr>
</tbody>
</table>

continued on next page...
Table 3: (continued)

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>343</td>
<td>Fab. de componentes e acessórios para veículos auto e seus motores</td>
<td>Águeda (6.58); V. N. de Gaia (5.31); Palmela (4.23); Oliveira de Azeméis (3.99); Loures (3.36); Cantanhede (2.33); V. N. de Cerveira (2.30); Maia (2.17)</td>
</tr>
<tr>
<td>291</td>
<td>Fab. de máquinas e equipamentos diversos</td>
<td>Braga (6.79); Seixal (3.70); Fundão (2.84); V. N. de Gaia (2.78); Figueira da Foz (2.60)</td>
</tr>
<tr>
<td>244</td>
<td>Fab. de produtos farmacêuticos</td>
<td>Tondela (5.03); Penafiel (4.19); V. N. Famalicão (3.01); Mortágua (2.97); Penela (2.88); Sintra (2.71); Condeixa-a-Nova (2.35); Coimbra (2.30); Oliveira de Frades (2.13)</td>
</tr>
<tr>
<td>202</td>
<td>Fab. de folheados, contraplacados, painéis, etc.</td>
<td>Paredes (6.61); Mangularde (6.32); Viana do Castelo (4.16); Proença-a-Nova (3.50); Setúbal (3.38); Chaves (3.35); Nelas (2.99); Monção (2.39); Oliveira do Hospital (2.15); Castelo de Paiva (1.96)</td>
</tr>
<tr>
<td>295</td>
<td>Fab. de outras máquinas e equipamento para uso específico</td>
<td>Marinha Grande (12.79); Leiria (6.14); Oliveira de Azeméis (4.24); Sintra (3.46); Alcobaça (2.97); V. N. de Gaia (2.52)</td>
</tr>
</tbody>
</table>
Figure 3: Overall levels of localization by "concelhos".
Figure 4: Highest profit-elasticities by 3-digit SIC industries.
Figure 5: Mean impact on $\gamma$ across 3-digit SIC industries.