FEP WORKING PAPERS
FEP WORKING PAPERS
Research Work in Progress

n. 353, Dec. 2009

Accounting for Neighboring Effects in Measures of Spatial Concentration

Paulo Guimarães [1 2]
Octávio Figueiredo [2 3]
Douglas Woodward[1]

[1] University of South Carolina
[2] CEMPRE
[3] Faculdade de Economia da Universidade do Porto

U.PORTO
FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

Accounting for Neighboring Effects in Measures of Spatial Concentration[1]

Paulo Guimarães

University of South Carolina and CEMPRE

Octávio Figueiredo

Universidade do Porto and CEMPRE

Douglas Woodward

University of South Carolina

December 16, 2009

[1]The authors acknowledge the support of FCT, the Portuguese Foundation for Science and Technology. An earlier version of this paper was presented at the 2009 North American Meetings of the Regional Science Association International, Spatial Modeling and Location Sessions. We thank Arthur Getis for commenting the paper.

Abstract

A common problem with spatial economic concentration measures (e.g. Gini, Herfindhal, entropy and Ellison-Glaeser indices) is accounting for the position of regions in space. While they purport to measure spatial clustering, these statistics are confined to calculations within individual areal units. They are insensitive to the proximity of regions – to neighboring effects. Clearly, economic clusters may cross the boundaries of the regions. Yet with current measures, any industrial agglomeration that traverses boundaries will be chopped into two or more pieces. Activity in adjacent spatial units is treated in exactly the same way as activity in far-flung, non-adjacent areas. This paper shows how some popular measures of spatial concentration relying on areal data can be modified to account for neighboring effects and spatial autocorrelation. With a U.S. application, we also show that the new instruments we propose are useful and easy to implement.

JEL classification: R12, R39, C49.

1 Introduction

In recent years there has been renewed interest in the study of spatial concentration patterns of economic activity. Starting with Krugman (1991*b*), the "new economic geography" models called for stylized facts about the distribution of industries across space. Researchers responded with a substantial amount of applied work. Beyond the ongoing research in North America, empirical work on spatial concentration of economic activities can be found for regions across the world.

Often, these studies rely on statistical concentration indices–Herfindhal, Gini, entropy–adapted to spatial problems [e.g. Krugman (1991*a*), Kim (1995), Brülhart & Traeger (2005) and Mori, Nishikimi & Smith (2005)]. Other researchers have advanced new tools specifically for spatial analysis. Most notable among these is Ellison & Glaeser's (1997) index (henceforth EG index) and related statistics formally derived from Carlton's (1983) location choice model [Maurel & Sedillot (1999), Devereux, Griffith & Simpson (2004) and Guimarães, Figueiredo & Woodward (2007)]. Applications of these indices is popular in part because they can use readily available data, typically spatially aggregated employment and/or plant counts by industry and region.[1]

A major problem with the commonly used concentration measures is that they ignore the position of regions in space, notably neighboring regions, even though they are based on spatial data. These statistics only account for concentration within pre-defined areal units and ignore "neighboring effects."

[1] Computation of the EG index and some of the related measures also requires data on Herfindhal indices measuring each industry concentration level in terms of plant sizes.

Current spatial concentration measures are insensitive to permutations of the spatial position of the regions. Indeed economic activity in adjacent spatial units is treated no differently than activity at opposite ends of the country. In reality, however, the spatial concentration of economic activity does not recognize areal units. Certainly, clusters of industries cross areal boundaries. In turn, measures of concentration must address the position, or ordering of regions in space.

Recently, two solutions borrowed from the spatial statistics literature have been put forth to deal with the "neighboring effect" problem when measuring spatial concentration. Some researchers [Arbia (2001) and Lafourcade & Mion (2007)] have proposed ad-hoc solutions that complement information from spatial concentration measures, such as the locational Gini coefficient or the EG index, with that of measures specifically developed to account for spatial autocorrelation, e.g. the Moran's I statistic. Yet in the context of these ad-hoc approaches it is unclear how to combine the information from concentration measures with that from spatial autocorrelation statistics.

A more sophisticated approach was proposed by both Marcon & Puech (2003) and Duranton & Overman (2005). Based on the theory of spatial point processes, the authors used detailed micro-level data on the employment and location for each individual plant to calculate distance-based indicators of spatial concentration. Clearly, this approach is able to measure spatial concentration without being affected by the "neighboring effect" problem. However, it requires highly detailed geographic data, with precise information on the exact location of each business unit. This level of precision is unavailable for most countries and regions. In practice, most researchers have access to spatially aggregated areal data only. Therefore, in most cases we lack a single statistic that summarizes spatial concentration, while at the same time accounting for "neighboring effects."

In this paper, we tackle this issue and propose an alternative approach to solve the problem within the context of areal data. The paper is organized as follows. In the next section, we offer several examples to motivate the "neighboring effect" problem when researchers work with available data to construct spatial concentration measures. Then, in section 3, we show how some of these measures can be modified to account for "neighboring effects." Section 4 presents an illustration using U.S. data, while section 5 summarizes and concludes the paper.

2 Spatial Concentration and the "Neighboring Effect" Problem

In this section we use some numerical examples borrowed from Arbia (2001) to illustrate the "neighboring effect" problem when dealing with areal data and the standard spatial concentration measures. Consider the three hypothetical distributions of 12 firms across 16 regions shown in Table 1. Given the position of regions in space, it is clear that the level of spatial concentration is the highest in situation 1.a – and higher in 1.b than in 1.c. However, because commonly used measures of spatial concentration only account for information within each areal unity they will give exactly the same result when applied to any of the three cases shown in Table 1.

Table 1: Three Hypothetical Distributions of 12 Firms

3	3	0	0
3	3	0	0
0	0	0	0
0	0	0	0

1.a

3	0	0	0
3	0	0	0
3	0	0	0
3	0	0	0

1.b

0	0	0	0
0	3	0	3
0	0	0	0
0	3	0	3

1.c

Consider the Herfindhal index, an absolute measure of concentration. In matrix form, this index is given by,

$$H = \mathbf{s}'\mathbf{s} = \sum_{j=1}^{J} s_j^2 \ , \tag{1}$$

where $\mathbf{s}' = [s_1, s_2, ..., s_J]$ is a vector containing the regional shares of a measure of interest (number of firms in our example) and J is the total number of regions in the economy (16 regions in the example shown in Table 1). It is clear from the formula in (1) why H takes the same value for any of the three situations described above [see the results in Table 2]. This statistic is insensitive to the position of the regions in space. It does not account for "neighboring effects." Thus, permutations of the spatial ordering of the regions leave the results unchanged.

Another measure used in the literature is the Ellison & Glaeser's (1997) "raw concentration" index – a slightly modified version of Hoover's coefficient of location. This index is a relative concentration measure because it compares the distribution of interest with a reference distribution. It is defined as,

$$G = (\mathbf{s} - \mathbf{x})'(\mathbf{s} - \mathbf{x}) = \sum_{j=1}^{J} (s_j - x_j)^2 \ , \tag{2}$$

where $\mathbf{x}' = [x_1, x_2, ..., x_J]$ is a vector containing the elements of the reference distribution. Again, as shown in Table 2, this measure fails to recognize the three different situations.[2] The same is also true for the EG index,

$$\gamma = \frac{G - H_I\,(1 - \mathbf{x}'\mathbf{x})}{(1 - H_I)\,(1 - \mathbf{x}'\mathbf{x})}\ , \tag{3}$$

where H_I is an Herfindhal index measuring industry concentration.[3] Note also that other commonly used measures of spatial concentration–such as the Gini coefficient or entropy indices–would also fail in ranking situations 1.a to 1.c in descending order. Actually, they will give exactly the same result when applied to any of the three situations.

Thus, to be able to correctly rank the three cases shown in Table 1, we need a measure that captures the position of regions in space accounting for "neighboring effects". An obvious candidate is, as noted by Arbia (2001) and Lafourcade & Mion (2007), the Moran's I index of spatial autocorrelation,

$$M = \frac{(\mathbf{s} - J^{-1}\mathbf{i}\mathbf{i}'\mathbf{s})'\mathbf{W}(\mathbf{s} - J^{-1}\mathbf{i}\mathbf{i}'\mathbf{s})}{(\mathbf{s} - J^{-1}\mathbf{i}\mathbf{i}'\mathbf{s})'(\mathbf{s} - J^{-1}\mathbf{i}\mathbf{i}'\mathbf{s})}\ ,$$

where $\mathbf{i}$ is a vector of ones and $\mathbf{W}$ is a row-standardized spatial contiguity matrix. This statistic is a spatial version of Pearson's correlation coefficient and is usually contained in the [-1, 1] interval.

[2] For simplicity, in our calculations, the reference distribution assumes equal probability for all regions.

[3] Hence, $H_I = \mathbf{z}'\mathbf{z}$, where $\mathbf{z}$ is a vector containing each firm's employment share in the industry's total employment. In our calculations, we assume that all firms are equally sized and thus $H_I = 1/12$.

Table 2: Concentration Measures for the Examples

	1.a	1.b	1.c	2.b
H	0.2500	0.2500	0.2500	0.0833
G	0.1875	0.1875	0.1875	0.0208
γ	0.1273	0.1273	0.1273	-0.0667
M	0.6111	0.4861	-0.3333	0.4861

As seen in Table 2, Moran's I correctly distinguishes the three cases displayed in Table 1.[4] However, a spatial autocorrelation statistic *per se* is not a good measure of spatial concentration. This point is well illustrated in Arbia (2001) who considered the two alternative scenarios given in Table 3. In this table, spatial dispersion is clearly higher in situation 2.b than in 1.b. To be sure, if we ignore the position of regions in space, then situation 2.b is the most dispersed that is possible to obtain with the 12 firms in our example. All concentration measures in Table 2 reflect this and show higher values for case 1.b. However, if we use Moran's I statistic as a measure of spatial concentration we are led to conclude that the degree of concentration is exactly the same in cases 1.b and 2.b. This is because Moran's I is designed to account for the degree of similitude between values in adjacent areas–taking into account "neighboring effects"–but fails to account for the information within each areal unity *per se*.[5] Hence, Moran's I statistic does not fully capture the concept of spatial concentration. In the next section, we propose

[4]In our calculations, each element of the **W** matrix takes a value of one for contiguous units and zero otherwise. The elements of the main diagonal are all zero and a rook's case definition of neighbors is assumed.

[5]Note that the diagonal elements of the spatial contiguity matrix in the Moran's I are set to zero.

a modification of some of the measures discussed above that allows us to simultaneously capture spatial concentration and account for "neighboring effects."

Table 3: Two Hypothetical Distributions of 12 Firms

3	0	0	0
3	0	0	0
3	0	0	0
3	0	0	0

1.b

0	1	1	1
0	1	1	1
0	1	1	1
0	1	1	1

2.b

3 Accounting for Neighboring Effects when Measuring Spatial Concentration

Consider again the Herfindhal index. A straightforward modification of the measure allows us to introduce "neighboring effects." This statistic can be seen as a quadratic form associated with the identity matrix $\mathbf{I}$. We can account for spatial interactions between all terms by writing a more general version of the Herfindhal such as,

$$H_S = \mathbf{s}'\mathbf{\Psi}\mathbf{s} \ , \tag{4}$$

where $\mathbf{\Psi}$ is a matrix of spatial weights with generic element ψ_{ij}, and non-null elements in the main diagonal. The matrix $\mathbf{\Psi}$ can be constructed in many different ways but here we will consider the case where $\mathbf{\Psi} = \mathbf{I} + \mathbf{W}$, where $\mathbf{W}$ is the conventional row-standardized contiguity matrix with zeros in the diagonal. Now, if there are no "neighboring effects", $\mathbf{\Psi} = \mathbf{I}$, and we obtain the conventional Herfindhal index. The spatially weighted version of the Herfindhal in (4) is bounded–if all regions have neighbors–in the $[2/J, 1]$ interval. Interestingly enough, as shown in Table 4, the H_S is able to distinguish the four different cases in Tables 1 and 3 discussed so far and it behaves as one would expect. This is because the spatially weighted Herfindhal in (4) mixes information from the conventional Herfindhal with that of Moran's I statistic. To see the relation note that,

$$\begin{aligned} H_S &= \mathbf{s}'(\mathbf{I}+\mathbf{W})\mathbf{s} \\ &= H + \mathbf{s}'\mathbf{W}\mathbf{s} \ , \end{aligned}$$

and given that,

$$\mathbf{s}'\mathbf{W}\mathbf{s} = M(H - J^{-1}) + J^{-1}\mathbf{W}\mathbf{s} \ ,$$

we obtain a relation between Moran's I and the spatially weighted version of the Herfindhal,

$$H_S = M(H - J^{-1}) + H + k \ , \tag{5}$$

where $k = J^{-1}\mathbf{W}\mathbf{s}$ is a spatially weighted mean of the shares. Inspection of (5) reveals that H_S is an increasing function of H (a measure of spatial concentration within regions) and M (a measure of spatial concentration across regions that picks the "neighboring effects").

This idea can be extended to the "raw concentration index" discussed above. In this case, a spatially weighted version can be constructed as,

$$G_S = (\mathbf{s}-\mathbf{x})'\mathbf{\Psi}(\mathbf{s}-\mathbf{x}) \ . \tag{6}$$

Table 4 shows that the G_S also has the expected behavior across all the four examples discussed up till now. The relation between the G_S and

Moran's I is also very simple,

$$G_S = G + (\mathbf{s} - \mathbf{x})'\mathbf{W}(\mathbf{s} - \mathbf{x}) \ ,$$

and thus,

$$G_S = G + M_D G \ , \tag{7}$$

where we add the subscript "D" to M to highlight the fact that Moran's I is applied to the differences in the shares.[6]

Table 4: Weighted Concentration Measures for the Examples

	1.a	1.b	1.c	2.b
H_s	0.4271	0.3958	0.2500	0.1586
G_s	0.3021	0.2786	0.1250	0.0310
γ_s	0.2857	0.2565	0.0649	−0.0523
M	0.6111	0.4861	−0.3333	0.4861

Using the G_S as a starting point we can also apply Ellison & Glaeser's (1997) procedure to derive a spatially weighted version of the EG index. It can be shown (see the Appendix) that this weighted version of the EG is given by,

$$\gamma_S = \frac{G_S - H_I\left(1 - \mathbf{x}'\mathbf{\Psi}\mathbf{x}\right)}{\left(1 - H_I\right)\left(1 - \mathbf{x}'\mathbf{\Psi}\mathbf{x}\right)} \ . \tag{8}$$

Again, for $\mathbf{\Psi} = \mathbf{I}$ the index collapses to the standard Ellison & Glaeser's (1997) measure. Note also that the spatially weighted EG behaves much like the other spatially weighted indices derived above (see Table 4). Indeed, γ_S

[6] However, in our example, because we are assuming that the elements of $\mathbf{x}$ are constant across regions, it is indifferent whether Moran's I is computed on $\mathbf{s}$ or on the differences in the shares, $(\mathbf{s} - \mathbf{x})$.

is a reparametrization of the G_S and as such will behave monotonically with G_S for a given spatial structure ($\mathbf{\Psi}$) and reference distribution ($\mathbf{x}$). This means that γ_S is also monotonically related to M_D (Moran's I applied to the differences in the shares).

4 An Empirical Application

We now illustrate the implementation of the spatially weighted EG index derived above with an application to the U.S. economy. To implement our approach, we needed aggregated data on employment by industry and region, as well as data on each establishment employment size to compute the H_I indices. We obtained these data from the web page of Thomas J. Holmes at the University of Minnesota. The data are for the year of 2000 and were originally obtained from the U.S. Census Bureau, *County Business Patterns* (CBP). There are approximately seven million establishments in the CBP database for this year. The CBP contains cell counts for establishments by industry (six-digit NAICS), employment size category, and county. Yet, the "raw" county-level CBP database has a large number of disclosure problems. The advantage of the Holmes' data is that an imputation procedure was used to fill all the empty cells. Moreover, Holmes provided estimates of average employment for all employment size categories. These can be used as point estimates of each establishment individual employment size.[7] In our calculations, we restrict analysis to the manufacturing industries and make

[7] This was the procedure adopted by Thomas Holmes. The database is described in more detail in Holmes & Stevens (2004) and can be found at www.econ.umn.edu/~holmes/data/cbp.

use of the four-digit level classification (86 industries) of the North American Industrial Classification System (NAICS).

To compute the spatially weighted EG index we also needed spatial weight matrices. We obtained these matrices from the REPEC data repository. The package in this repository contains two Stata data files with spatial matrices at the state and county levels for the 48 U.S. contiguous states.[8] These are first-order contiguity matrices, giving an weight of one to all contiguous locations and zero otherwise, including the own location.[9]

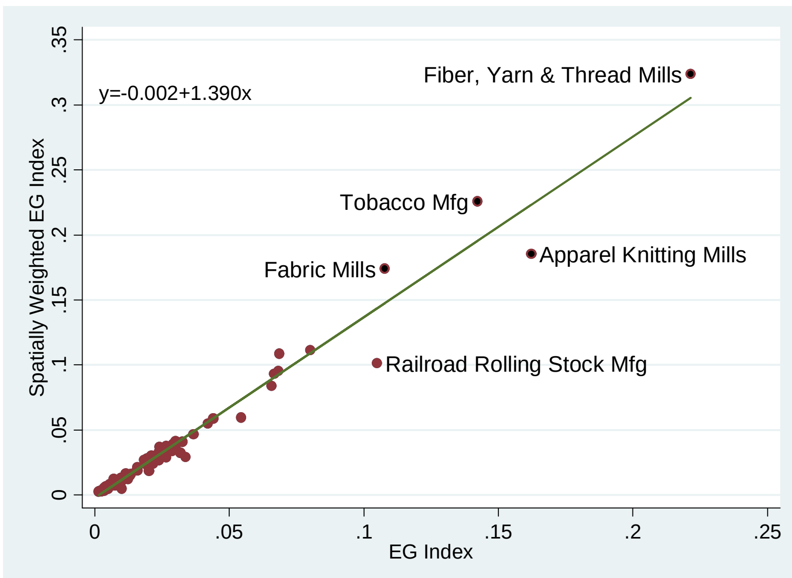
Figure 1 shows employment EG indices computed at the state level for the 86 four-digit NAICS industries. In this figure we plot the EG index in the x-axis against its spatially weighted version in the y-axis, identifying the five most concentrated sectors according to the unweighted EG measure. We also plot in the figure the fitted line between the two indices. As the regression shows, on average, the spatially weighted index is 39 percent higher than its counterpart. This result is in line with Duranton & Overman's (2005) critique of the unweighted EG index. The authors note that the EG, as well as related unweighted measures, may be downwardly biased due to the "neighboring effect" problem. Any agglomeration of an industry that crosses the states boundaries will be chopped into two or more pieces. Yet, Figure 1 also shows that for some sectors we can also obtain lower values for the spatially weighted

[8]The *usswm* package, developed by Scott Merryman, can be found at http://fmwww.bc.edu/RePEc/bocode/u. The original spatial weight matrices were created by Luc Anselin.

[9]The matrices were row-standardized using the Stata command *spatwmat* described in Pisati (2001).

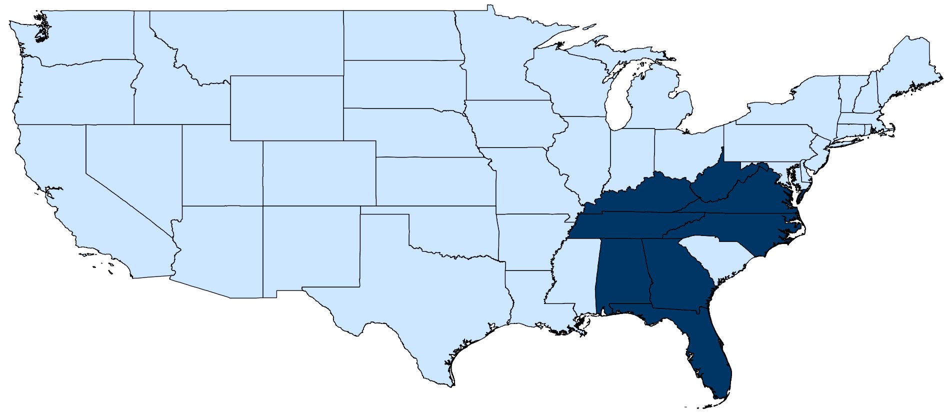
EG measure. This is the case for one of the most concentrated industries, the Railroad Rolling Stock industry (NAICS 3365), highlighted in the figure. In contrast, for the Tobacco industry (NAICS 3122) the weighted EG index increased well above the average.

Figure 1: **EG Indices by 4–digit NAICS at the State Level**



To gain some insight into this issue we mapped the excess of concentration, $(s_j - x_j)$, for these two sectors. In Figures 2 and 3, the darker color represents states with industry employment concentration above expected levels while the lighter color stands for states with below expected industry employment levels. These figures are reminiscent of the two situations (1.a and 1.c) shown above in Table 1. For the Tobacco industry, there is only a single pocket of concentration in the southeastern part of the United States, while for the Railroad Rolling industry there are several pockets of

Figure 2: **Excess Concentration in NAICS 3132–Tobacco Mfg**

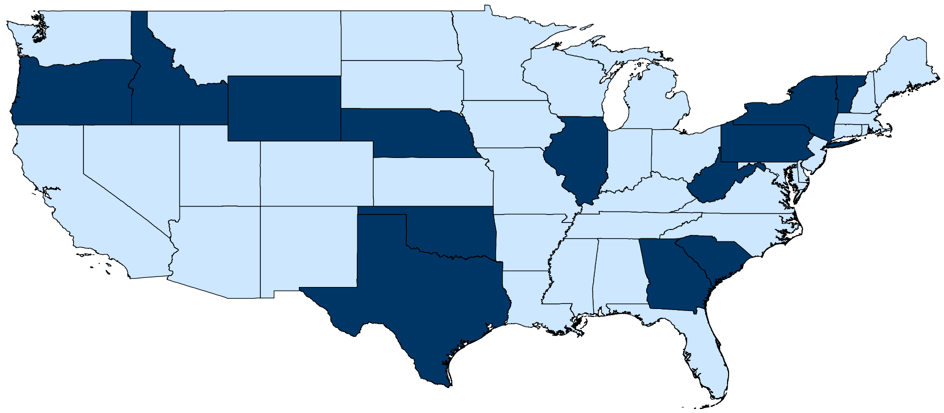


concentration scattered around the country. Thus, as expected, spatial autocorrelation, as measured by Moran's I, reinforces spatial concentration of Tobacco, while it downplays concentration for the Railroad Rolling industry.

Typically, EG indices are used to rank industries by level of spatial concentration. In our application–with data aggregated at the state level and using a first-order contiguity matrix–we find a significant amount of concordance among the two indices. The hierarchy of individual industries for the 86 sectors remains basically unchanged, as a Spearman rank correlation coefficient of 98 per cent between γ and γ_S indicates. Yet there are some noteworthy differences. Among these, we discover that the Engine, Turbine and Power Equipment (NAICS 3336) ranks 29th according to the unweighted index, but it moves up to the 19th place according to the weighted measure. On the other hand, the Motor Vehicle Body and Trailer industry (NAICS 3362) is the 15th most concentrated industry according to the unweighted EG index but drops to 31st in terms of the γ_S.

It should also be noted that differences in rankings may be much larger

Figure 3: **Excess Concentration in NAICS 3365–Railroad Rolling Stock Mfg**



when considering other spatial structures. To gain some insight into this question, we also measured the level of industries' spatial concentration within each one of the 48 contiguous states using county level data. At this level, in order to have sufficient information in each cell, we used data aggregated by three-digit codes of the North American Industrial Classification System (21 industries).[10] In Table 5, we report the four highest and lowest Spearman rank correlation coefficients between γ and γ_S. As can be seen in this table, there are considerable changes in the hierarchy of individual industries for some states. Table 6 shows detailed information for selected cases that evidence striking changes in the ordering of industries' spatial concentration within states. For example, the Textile Mills industry (NAICS 313) ranks 11th according to the unweighted EG index but is the most spatially concentrated industry in Virginia if we rely on the γ_S (see the

[10]When working at the county level with four-digit NAICS codes there are too many cells with zero values.

numbers in parenthesis in Table 6). In contrast, the Electrical Equipment & Components sector (NAICS 335) is the 7th highest ranked in South Dakota according to γ but drops to the 13th place if we account for "neighboring effects." To let the readers gain some insight on these situations, we mapped in Figures 4 and 5 concentration in excess, $(s_j - x_j)$, for these two cases.

Table 5: Spearman Rank Correlations

Smallest		Highest	
Delaware	41.0%	Texas	99.7%
Massachusetts	69.6%	California	99.5%
Virginia	87.3%	Nebraska	99.5%
Wisconsin	88.8%	Arizona	99.5%

Table 6: Comparison of γ and γ_s: Some selected cases

State	NAICS-3d	γ	γ_s
$\gamma < \gamma_s$			
Virginia	Textile Mills	0.007 (11)	0.034 (1)
Wisconsin	Elect Equip & Comp	0.003 (17)	0.016 (7)
North Carolina	Food	0.005 (17)	0.011 (8)
$\gamma > \gamma_s$			
South Dakota	Elect Equip & Comp	0.116 (7)	0.040 (13)
Louisiana	Textile Mills	0.042 (4)	0.018 (9)
North Carolina	Transport Equip	0.010 (9)	0.008 (14)

Figure 4: **Excess Concentration in Virginia for NAICS 313–Textile Mills Mfg**

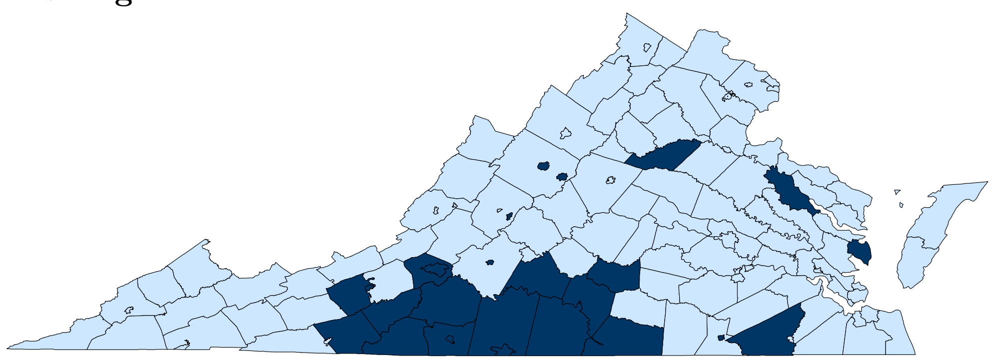
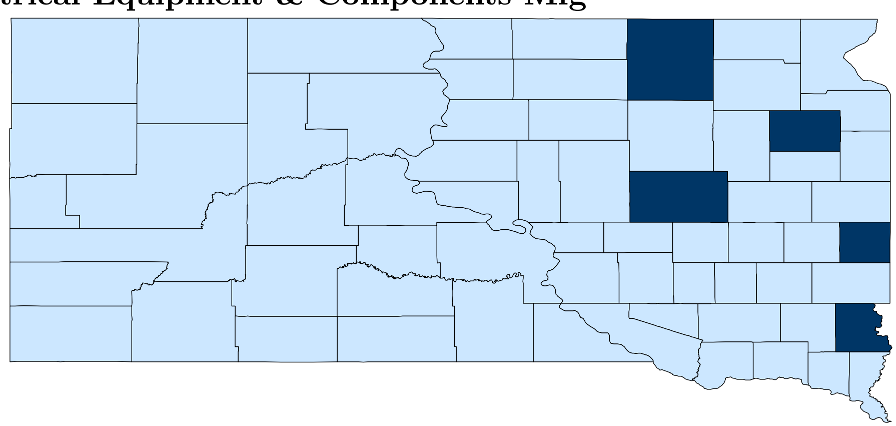


Figure 5: **Excess Concentration in South Dakota for NAICS 335–Electrical Equipment & Components Mfg**



5 Conclusion

A well known problem with most measures of spatial economic concentration is that they eschew the position of regions in space. These statistics only account for information within pre-defined areal units, leaving out "neighboring effects." Yet, as this paper stresses, clusters of industries may cross the boundaries of regions and any industrial agglomeration that traverses areal units will be chopped into two or more pieces. Thus, activity in adjacent spatial units can not be considered in exactly the same way as activity elsewhere in the country.

A complex approach to deal with the problem was been proposed by Marcon & Puech (2003) and Duranton & Overman (2005). Relying on the theory of spatial point processes, these authors calculate distance-based indicators of spatial concentration and avoid the "neighboring effect" problem. Yet their approach requires information on the exact location of each business unit, which is unavailable for most countries or regions. In practice, most regional researchers have access to spatially aggregated, areal data. They lack, however, a single statistic that summarizes spatial concentration along with spatial autocorrelation.

In this paper, we propose an alternative approach to solve the problem with existing areal data and spatial concentration indices. We show how some popular measures relying on areal data can be modified to account for "neighboring effects" and spatial autocorrelation when quantifying industrial concentration in space. With a U.S. application, we compare the Ellison-Glaeser index to its spatial weighted counterpart. The results make sense. The modified index seems to do a better job of quantifying spatial economic concentration. This empirical test demonstrates that the new instruments we propose are useful and tractable, given the databases and tools to build spatial weight matrices commonly available in Europe, North America, and elsewhere.

References

Arbia, G. (2001), 'The role of spatial effects on the empirical analysis of regional concentration', *Journal of Geographical Systems* **3**(3), 271–281.

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

Carlton, D. (1983), 'The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables', *The Review of Economic and Statistics* **65**(3), 440–449.

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

Duranton, G. & Overman, H. (2005), 'Testing for localisation using micro-geographic data', *Review of Economic Studies* **72**(4), 1077–1106.

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

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

Holmes, T. J. & Stevens, J. J. (2004), Spatial distribution of economic activities in North America, *in* J. V. Henderson & J. F. Thisse, eds, ‘Handbook of Regional and Urban Economics’, Elsevier, New York, pp. 2797–2843.

Kim, S. (1995), ‘Expansion of markets and the geographic distribution of economic activities: the trends in U.S. regional manufacturing structure, 1860-1987’, *Quarterly Journal of Economics* **110**(4), 881–908.

Krugman, P. (1991*a*), *Geography and Trade*, M.I.T. Press, Cambrige, MA.

Krugman, P. (1991*b*), ‘Increasing returns and economic geography’, *Journal of Political Economy* **99**, 483–499.

Lafourcade, M. & Mion, G. (2007), ‘Concentration, agglomeration and the size of plants’, *Regional Science and Urban Economics* **37**(1), 46–68.

Marcon, E. & Puech, F. (2003), ‘Evaluating the geographic concentration of industries using distance-based methods’, *Journal of Economic Geography* **3**(4), 409–428.

Maurel, F. & Sedillot, B. (1999), ‘A measure of the geographic concentration in french manufacturing industries’, *Regional Science and Urban Economics* **29**(5), 575–604.

Mori, T., Nishikimi, K. & Smith, T. (2005), ‘A divergence statistic for industrial localization’, *The Review of Economic and Statistics* **87**(4), 635–651.

Pisati, M. (2001), ‘Tools for spatial data analysis’, *Stata Technical Bulletin* **60**, 21–37.

Appendix: Derivation of the Spatially Weighted EG Index

Following Ellison & Glaeser (1997), we let the location decision of firm m be represented by a multinomial vector, $\mathbf{u}_m$, with expected probabilities given by the vector $\mathbf{p}$. All N firms in the industry are assumed to have independent location decisions and to face the same multinomial distribution. Firms have an exogenous given employment size and firm's m share of total employment in the industry is given by z_m. Thus, we can express the regional shares for that industry employment as,

$$\mathbf{s} = \sum_{m=1}^{N} z_m \mathbf{u}_m \ . \tag{A.1}$$

From here it follows that,

$$E(\mathbf{s}) = \sum_{m=1}^{N} z_m E(\mathbf{u}_m) = \sum_{m=1}^{N} z_m \mathbf{p} = \mathbf{p} \ ,$$

and, given the independence assumption of the $\mathbf{u}$ vectors,

$$\begin{aligned} V(\mathbf{s}) &= V\left(\sum_{m=1}^{N} z_m \mathbf{u}_m\right) \\ &= V(\mathbf{u}_m) \sum_{m=1}^{N} z_m^2 \\ &= [diag(\mathbf{p}) - \mathbf{p}\mathbf{p}']H_I \ , \end{aligned}$$

where H_I is an Herfindhal index measuring the industry concentration level in terms of firms' employment sizes.

The spatially weighted version of G is,

$$G_S = (\mathbf{s} - \mathbf{x})'\mathbf{\Psi}(\mathbf{s} - \mathbf{x}) \ ,$$

which we can expand to,

$$G_S = \mathbf{s}'\mathbf{\Psi}\mathbf{s} - 2\mathbf{x}'\mathbf{\Psi}\mathbf{s} + \mathbf{x}'\mathbf{\Psi}\mathbf{x} \ .$$

Taking the expected value of G_S, conditional on $\mathbf{p}$, we obtain,

$$E(G_S|\mathbf{p}) = E(\mathbf{s}'\mathbf{\Psi}\mathbf{s}|\mathbf{p}) - 2E(\mathbf{x}'\mathbf{\Psi}\mathbf{s}|\mathbf{p}) + \mathbf{x}'\mathbf{\Psi}\mathbf{x} \ .$$

For simplicity, we can solve the above expression by parts,

$$\begin{aligned}
E(\mathbf{s}'\mathbf{\Psi}\mathbf{s}|\mathbf{p}) &= tr[\mathbf{\Psi}V(\mathbf{s})] + \mathbf{p}'\mathbf{\Psi}\mathbf{p} \\
&= tr(H_I\mathbf{\Psi}[diag(\mathbf{p}) - \mathbf{p}\mathbf{p}']) + \mathbf{p}'\mathbf{\Psi}\mathbf{p} \\
&= H_I \sum_{j=1}^{J} \psi_{jj} p_j - H_I tr(\mathbf{\Psi}\mathbf{p}\mathbf{p}') + \mathbf{p}'\mathbf{\Psi}\mathbf{p} \\
&= H_I \sum_{j=1}^{J} \psi_{jj} p_j + (1 - H_I)\mathbf{p}'\mathbf{\Psi}\mathbf{p} \ ,
\end{aligned}$$

and, since $E(\mathbf{x}'\mathbf{\Psi}\mathbf{s}|\mathbf{p}) = \mathbf{x}'\mathbf{\Psi}\mathbf{p}$, then,

$$E(G_S|\mathbf{p}) = H_I \sum_{j=1}^{J} \psi_{jj} p_j + (1 - H_I)\mathbf{p}'\mathbf{\Psi}\mathbf{p} - 2\mathbf{x}'\mathbf{\Psi}\mathbf{p} + \mathbf{x}'\mathbf{\Psi}\mathbf{x} \ .$$

Following Ellison & Glaeser (1997), we assume that,

$$\begin{aligned}
E(\mathbf{p}) &= \mathbf{x} \\
V(\mathbf{p}) &= \gamma\,[\mathrm{diag}(\mathbf{x}) - \mathbf{x}\mathbf{x}'] \ ,
\end{aligned}$$

which we need to calculate the unconditional expected value of G_S. Using the law of iterated expectations we obtain,

$$E(G_S) = H_I \sum_{j=1}^{J} \psi_{jj} x_j + (1 - H_I)E(\mathbf{p}'\mathbf{\Psi}\mathbf{p}) - 2\mathbf{x}'\mathbf{\Psi}\mathbf{x} + \mathbf{x}'\mathbf{\Psi}\mathbf{x} \ . \qquad \text{(A2)}$$

Because,

$$E(\mathbf{p}'\mathbf{\Psi}\mathbf{p}) = \mathrm{tr}(\mathbf{\Psi}\mathbf{V}(\mathbf{p})) + \mathbf{x}'\mathbf{\Psi}\mathbf{x}$$
$$= \gamma\sum_{j=1}^{J}\psi_{jj}x_j - \gamma\mathbf{x}'\mathbf{\Psi}\mathbf{x} + \mathbf{x}'\mathbf{\Psi}\mathbf{x}\ ,$$

we can replace $E(\mathbf{p}'\mathbf{\Psi}\mathbf{p})$ in (A2) to obtain,

$$E(G_S) = H_I\sum_{j=1}^{J}\psi_{jj}x_j + \gamma[(1-H_I)\sum_{j=1}^{J}\psi_{jj}x_j - (1-H_I)\mathbf{x}'\mathbf{\Psi}\mathbf{x}] - H_I\mathbf{x}'\mathbf{\Psi}\mathbf{x}\ .$$

Solving for γ and replacing $E(G_S)$ by its actual value, G_S, we obtain,

$$\gamma_S = \frac{G_S - H_I\left(\sum_{j=1}^{J}\psi_{jj}x_j - \mathbf{x}'\mathbf{\Psi}\mathbf{x}\right)}{(1-H_I)\left(\sum_{j=1}^{J}\psi_{jj}x_j - \mathbf{x}'\mathbf{\Psi}\mathbf{x}\right)}\ ,$$

and, given that in our parametrization we assume that $\psi_{jj} = 1$, then the above formula simplifies to,

$$\gamma_S = \frac{G_S - H_I\,(1-\mathbf{x}'\mathbf{\Psi}\mathbf{x})}{(1-H_I)\,(1-\mathbf{x}'\mathbf{\Psi}\mathbf{x})}\ .$$

It is also worthy to note that if H_I=$1/N$ we obtain a spatially weighted version of the particular case of the "EG index for plant counts" derived in Maurel & Sedillot (1999) and Guimarães et al. (2007),

$$\gamma_S = \frac{NG_S - (1-\mathbf{x}'\mathbf{\Psi}\mathbf{x})}{(N-1)\,(1-\mathbf{x}'\mathbf{\Psi}\mathbf{x})}\ .$$

Recent FEP Working Papers

Nº 352	Vasco Leite, Sofia B.S.D. Castro and João Correia-da-Silva, "*A third sector in the core-periphery model: non-tradable goods*", December 2009
Nº 351	João Correia-da-Silva and Joana Pinho, "*Costly horizontal differentiation*", December 2009
Nº 350	João Correia-da-Silva and Joana Resende, "*Free daily newspapers: too many incentives to print?*", December 2009
Nº 349	Ricardo Correia and Carlos Brito, "*Análise Conjunta da Dinâmica Territorial e Industrial: O Caso da IKEA – Swedwood*", December 2009
Nº 348	Gonçalo Faria, João Correia-da-Silva and Cláudia Ribeiro, "*Dynamic Consumption and Portfolio Choice with Ambiguity about Stochastic Volatility*", December 2009
Nº 347	André Caiado, Ana Paula Africano and Aurora A.C. Teixeira, "*Firms' perceptions on the usefulness of State trade missions: an exploratory micro level empirical analysis*", December 2009
Nº 346	Luís Pinheiro and Aurora A.C. Teixeira, "*Bridging University-Firm relationships and Open Innovation literature: a critical synthesis*", November 2009
Nº 345	Cláudia Carvalho, Carlos Brito and José Sarsfield Cabral, "*Assessing the Quality of Public Services: A Conceptual Model*", November 2009
Nº 344	Margarida Catarino and Aurora A.C. Teixeira, "*International R&D cooperation: the perceptions of SMEs and Intermediaries*", November 2009
Nº 343	Nuno Torres, Óscar Afonso and Isabel Soares, "*Geographic oil concentration and economic growth – a panel data analysis*", November 2009
Nº 342	Catarina Roseira and Carlos Brito, "*Value Co-Creation with Suppliers*", November 2009
Nº 341	José Fernando Gonçalves and Paulo S. A. Sousa, "*A Genetic Algorithm for Lot Size and Scheduling under Capacity Constraints and Allowing Backorders*", November 2009
Nº 340	Nuno Gonçalves and Ana Paula Africano, "*The Immigration and Trade Link in the European Union Integration Process*", November 2009
Nº 339	Filomena Garcia and Joana Resende, "*Conformity based behavior and the dynamics of price competition: a new rational for fashion shifts*", October 2009
Nº 338	Nuno Torres, Óscar Afonso and Isabel Soares, "*Natural resources, economic growth and institutions – a panel approach*", October 2009
Nº 337	Ana Pinto Borges, João Correia-da-Silva and Didier Laussel, "*Regulating a monopolist with unknown bureaucratic tendencies*", October 2009
Nº 336	Pedro Rui Mazeda Gil, "*Animal Spirits and the Composition of Innovation in a Lab-Equipment R&D Model*", September 2009
Nº 335	Cristina Santos and Aurora A.C. Teixeira, "*The evolution of the literature on entrepreneurship. Uncovering some under researched themes*", September 2009
Nº 334	Maria das Dores B. Moura Oliveira and Aurora A.C. Teixeira, "*Policy approaches regarding technology transfer: Portugal and Switzerland compared*", September 2009
Nº 333	Ana Sofia Ferreira, Leonídio Fonseca and Lilian Santos, "*Serão os 'estudantes empreendedores' os empreendedores do futuro? O contributo das empresas juniores para o empreendedorismo*", August 2009
Nº 332	Raquel Almeida, Marina Silva and Tiago Soares, "*Coesão Territorial - As relações de fronteira entre Portugal e Espanha*", August 2009
Nº 331	Custódia Bastos, Suzi Ladeira and Sofia Silva, "*Empreendedorismo nas Artes ou Artes do Empreendedorismo? Um estudo empírico do 'Cluster' da Rua Miguel Bombarda*", August 2009
Nº 330	Filipe A. Ribeiro, Ana N. Veloso and Artur V. Vieira, "*Empreendedorismo Social: Uma análise via associativismo juvenil*", August 2009
Nº 329	Argentino Pessoa, "*Outsourcing And Public Sector Efficiency: How Effective Is Outsourcing In Dealing With Impure Public Goods?*", July 2009
Nº 328	Joana Almodovar, Aurora A.C. Teixeira, "*Conceptualizing clusters through the lens of networks: a critical synthesis*", July 2009
Nº 327	Pedro Mazeda Gil, Fernanda Figueiredo and Óscar Afonso, "*Equilibrium Price Distribution with Directed Technical Change*", July 2009

Nº 326	Armando Silva, Ana Paula Africano and Óscar Afonso, "*Which Portuguese firms are more innovative? The importance of multinationals and exporters*", June 2009
Nº 325	Sofia B. S. D. Castro, João Correia-da-Silva and Pascal Mossay, "*The core-periphery model with three regions*", June 2009
Nº 324	Marta Sofia R. Monteiro, Dalila B. M. M. Fontes and Fernando A. C. C. Fontes, "*Restructuring Facility Networks under Economy of Scales*", June 2009
Nº 323	Óscar Afonso and Maria Thompson, "*Costly Investment, Complementarities and the Skill Premium*", April 2009
Nº 322	Aurora A.C. Teixeira and Rosa Portela Forte, "*Unbounding entrepreneurial intents of university students: a multidisciplinary perspective*", April 2009
Nº 321	Paula Sarmento and António Brandão, "*Next Generation Access Networks: The Effects of Vertical Spillovers on Access and Innovation*", April 2009
Nº 320	Marco Meireles and Paula Sarmento, "*Incomplete Regulation, Asymmetric Information and Collusion-Proofness*", April 2009
Nº 319	Aurora A.C. Teixeira and José Sequeira, "*Determinants of the international influence of a R&D organisation: a bibliometric approach*", March 2009
Nº 318	José Sequeira and Aurora A.C. Teixeira, "*Assessing the influence of R&D institutions by mapping international scientific networks: the case of INESC Porto*", March 2009
Nº 317	João Loureiro, Manuel M. F. Martins and Ana Paula Ribeiro, "*Cape Verde: The Case for Euroization*", March 2009
Nº 316	Ester Gomes da Silva and Aurora A.C. Teixeira, "*Does structure influence growth? A panel data econometric assessment of 'relatively less developed' countries, 1979-2003*", March 2009
Nº 315	Mário A. P. M. Silva, "*A Model of Growth with Intertemporal Knowledge Externalities, Augmented with Contemporaneous Knowledge Externalities*", March 2009
Nº 314	Mariana Lopes and Aurora A.C. Teixeira, "*Open Innovation in firms located in an intermediate technology developed country*", March 2009
Nº 313	Ester Gomes da Silva, "*Capital services estimates in Portuguese industries, 1977-2003*", February 2009
Nº 312	Jorge M. S. Valente, Maria R. A. Moreira, Alok Singh and Rui A. F. S. Alves, "*Genetic algorithms for single machine scheduling with quadratic earliness and tardiness costs*", February 2009
Nº 311	Abel Costa Fernandes, "*Explaining Government Spending: a Cointegration Approach*", February 2009
Nº 310	João Correia-da-Silva, "*Uncertain delivery in markets for lemons*", January 2009
Nº 309	Ana Paula Ribeiro, "*Interactions between Labor Market Reforms and Monetary Policy under Slowly Changing Habits*", January 2009
Nº 308	Argentino Pessoa and Mário Rui Silva, "*Environment Based Innovation: Policy Questions*", January 2009
Nº 307	Inês Drumond and José Jorge, "*Basel II Capital Requirements, Firms' Heterogeneity, and the Business Cycle*", January 2009
Nº 306	Adelaide Maria Figueiredo, Fernanda Otília Figueiredo and Natália Pimenta Monteiro, "*Labor adjustments in privatized firms: a Statis approach*", December 2008
Nº 305	Manuela A. D. Aguiar and Sofia B. S. D. Castro, "*Chaotic and deterministic switching in a two-person game*", December 2008

Editor: Sandra Silva (sandras@fep.up.pt)
Download available at:
http://www.fep.up.pt/investigacao/workingpapers/
also in http://ideas.repec.org/PaperSeries.html

www.fep.up.pt

Faculdade de Economia da Universidade do Porto
Rua Dr. Roberto Frias, 4200-464 Porto | Tel. 225 571 100
Tel. 225571100 | www.fep.up.pt