United Nations University
CRIS
Comparative Regional Integration Studies

UNU-CRIS Occasional Papers

O-2006/22

Disproportionality Measures of Concentration, Specialization, and Polarization

Frank Bickenbach, Eckhardt Bode*

Revised version, September 2006

* Kiel Institute for the World Economy, D-24100 Kiel, e-mail frank.bickenbach@ifw-kiel.de, eckhardt.bode@ifw-kiel.de. The authors would like to thank Christiane Krieger-Boden and Serge Rey for helpful comments and suggestions. Financial support by the EU Commission, project EURECO, is gratefully acknowledged.

Disproportionality Measures of Concentration, Specialization, and Polarization

Frank Bickenbach, *Eckhardt Bode*

Abstract:

This paper extends the methodological toolbox of measures of the regional concentration of industries and the industrial specialization of regions. It, first, defines the class of disproportionality measures of concentration and specialization, and proposes a taxonomy that gives rise to a modular construction system for these measures. The taxonomy helps researchers to adjust measures flexibly to their research purpose and data. Second, the paper generalizes the taxonomy to (i) disproportionality measures of economic polarization that evaluate specialization and concentration simultaneously, and to (ii) spatial disproportionality measures of industrial concentration that deal with the checkerboard problem and the MAUP.

Keywords: disproportionality measures, specialization, concentration, polarization, spatial weighting

JEL classification: C43, F15, R12

Introduction

The new economic geography has raised concerns that economic integration at the regional and international level may heighten the regional concentration of industries (henceforth *concentration* for short) and the industrial specialization of regions (*specialization*). Innovative, dynamic industries may concentrate in core regions, leaving peripheral regions with aging, torpid industries. If the core regions specialize in dynamic, and the peripheral regions in torpid industries, both groups of regions will be more vulnerable to adverse macroeconomic shocks, and the peripheral regions will grow slower in terms of income and employment.

Various studies have explored the evolutions of concentration and specialization in Europe and other regions using statistical inequality measures borrowed directly or indirectly from the income inequality literature.[1] The results emerging from these studies are remarkably inconclusive for at least three reasons (Combes and Overman 2004): First, many of the studies lack a clear-cut research purpose and test hypothesis. Second, the studies differ in the

[1] Examples are the Theil index, the Gini index, the coefficient of variation, and the so-called Krugman index (relative mean deviation). In addition, the Herfindahl index, dartboard measures, and statistics based on Ripley's K have been used to measure concentration or specialization. See Bode et al. (2003), Combes and Overman (2004), and Nijkamp et al. (2003) for recent reviews.

sectoral and spatial scales of the data used to calculate the measures. The choice of the sectoral and spatial scales may affect the values of the measures due to the modifiable areal unit problem (MAUP) and the checkerboard problem.[2] And third, the choice of the statistical measure applied has been largely ad hoc, neglecting the sensitivity of the measure's adequacy to the research purpose and the sectoral and spatial scales of the data.

To facilitate a more substantive analysis of concentration and specialization patterns, Combes and Overman (2004) set up a catalog of 'baseline criteria' for a 'perfect' measure. In a nutshell, these criteria require that the measure

(i) be suitable for specifying an unambiguous and meaningful null hypothesis of no concentration or specialization that captures both systematic variations, suggested by economic theory, and random variations in the data;

(ii) be comparable across industrial and regional units and scales, and unbiased by the MAUP and the checkerboard problem in both the sectoral and the regional dimension; and

(iii) be suitable for statistical testing.

Rather than trying to develop a 'perfect' measure, the purpose of the present paper is to improve upon the *inequality* measures used in the concentration and specialization literature so far (and upon their choice by the researcher). First, the paper extends the inequality measures to – what will be called – *disproportionality* measures, and it proposes a taxonomy for disproportionality measures of concentration and specialization. The disproportionality measures can be adjusted more flexibly to the research purpose and data at hand than the inequality measures. The taxonomy gives rise to a modular system of three characteristic features of any disproportionality measure: the projection function, the reference distribution, and the weighting scheme. The taxonomy can be used to define a great variety of disproportionality measures by determining each of the three features – largely independently of the other two features.[3] It thereby facilitates choosing the measure that meets the requirements of a specific research hypothesis (baseline criterion i) and data (criterion ii) most closely. The taxonomy can also be used for performing sensitivity tests (criterion iii) by selectively modifying the

---

2 The MAUP (Openshaw and Taylor 1979; Arbia 1989) arises from discretizing heterogeneous continuous variables. It comes under two guises: (i) Discretizing space averages away heterogeneity, such that results are sensitive to the scale of aggregation (*scale* problem), and (ii) the boundaries between the discrete spatial units may be misplaced (*arbitrary boundary* problem). The checkerboard problem (Arbia 2001) arises from neglecting relevant information on the locations of or distances between regions or industries. Problems similar to the MAUP and the checkerboard problem arise in the sectoral dimension as well.

3 The bulk of the literature so far has taken measures as a fixed combination of two of these features, the projection function and the reference distribution. An important step towards using a more flexible combination of features is taken by Brülhart and Träger (2005), who consider using different references for the same projection function. However, they do not consider varying the weighting scheme independently of the reference distribution.

realization of a single characteristic feature.[4] In addition, it may also give the researcher guidance for specifying his null hypothesis properly, and for alleviating the effects of a suboptimal sectoral and spatial disaggregation of the data on his inferences.

The taxonomy covers most of the measures used frequently in the literature on concentration and specialization, including the Gini coefficient, the Krugman index, the Theil index, and the coefficient of variation.[5] It does not, however, cover measures that differ conceptually from inequality measures, such as the so-called 'dartboard' measures (Ellison and Glaeser 1997; Maurel and Sédillot 1999), or distance-based statistics based on Ripley's K functions proposed recently by Duranton and Overman (2005; 2006) and Marcon and Puech (2003; 2005).

Second, the paper extends the set of available measures – and generalizes their taxonomy – to disproportionality measures of polarization of an economy (*polarization measures* for short) and spatial disproportionality measures of concentration (*spatial concentration measures*). Polarization measures assess the concentration of industries and specialization of regions within an economy simultaneously. As generalizations of the disproportionality measures of concentration and specialization discussed in the first part, polarization measures can be used for a nested analysis of concentration and specialization patterns at different sectoral and regional scales. Spatial concentration measures take into account the neighborhood structures of the regions as well as additional empirical or theoretical information about the unobserved intra-regional distributions of the variable of interest. Spatial disproportionality measures help deal with the MAUP and the checkerboard problem. They are thus alternatives to the distance-based concentration statistics based on Ripley's K functions.

The organization of the paper is as follows: Section 2 defines disproportionality measures of concentration and specialization, introduces the taxonomy for these measures, and illustrates how specific disproportionality measures of concentration and specialization can be defined using the modular system of characteristic features. Section 3 extends the taxonomy to disproportionality measures of polarization, and Section 4 extends it to spatial disproportionality measures of concentration. Section 5 concludes, and discusses directions for future research. An empirical illustration for selected measures is given in Bickenbach et al. (2006). A more detailed tabulation of the various measures discussed in the paper is available at http://www.uni-kiel.de/ifw/staff/bode/measures.htm.

---

[4] A detailed discussion of the conceptual and technical issues of assessing the statistical significance of changes of a measure over time, or of the differences between regions or industries, is beyond the scope of the present paper. Among the few studies performing rigorous statistical tests are Brülhart and Träger (2005) and Mori et al. (2005). While the tests proposed by Mori et al. (2005) are specific to the Kullback-Leibler D statistic, which is akin to a Theil index, the bootstrap tests proposed by Brülhart and Träger (2005) for the Theil index and the CV can, in principle, be applied to all the disproportionality measures discussed in this paper.

[5] The Herfindahl index, another measure used in this literature, is closely related to the coefficient of variation.

1. The Taxonomy

This section introduces a taxonomy of disproportionality measures of concentration and specialization. All the measures covered by the taxonomy can be characterized as measures of the disproportionality of the distribution of a population across a set of mutually exclusive characteristics and a predetermined reference distribution. Since the available data is discrete in most applications, the discussion will focus on discrete versions of the measures. The population may be workers, establishments, or units of value added; the characteristics may be industries or regions. For expositional convenience, the following discussion will exemplify measures of the regional concentration of employment in an industry.[6] Thus, the population is *workers within an industry*, and their characteristics are the *regions* of their workplaces.[7]

Formally, for a finite set of industries, $i \in \mathbf{I} = \{1, \ldots, I\}$, and a set of regions, $r \in \mathbf{R} = \{1, \ldots, R\}$, let $\mathbf{L}_{(ir)} = (L_{ir}: ir \in \mathbf{I} \times \mathbf{R})$ denote the industry-region employment pattern and $\mathbf{L}_{i(r)} = (L_{ir}: r \in \mathbf{R})$ the distribution of industry i employment across regions. For a given reference distribution, $\mathbf{\Pi}_{(r)} = (\Pi_r: r \in \mathbf{R})$, and absolute region-specific weights, $\mathbf{W}_{(r)} = (W_r: r \in \mathbf{R})$, the disproportionality measure $M_i^{W\Pi}$ is defined as

$$M_i^{W\Pi} = f_M\left(\mathbf{W}_{(r)}, \frac{\mathbf{L}_{i(r)}}{\mathbf{\Pi}_{(r)}}\right). \tag{1}$$

M reflects the projection function, f_M, and the superscripts, $W\Pi$, denote the choice of the weights and the references. The projection function is such that the *region-specific proportionality factors*, L_{ir}/Π_r [=: X_{ir}], ($r \in \mathbf{R}$), are always scaled by their weighted average across all regions $r \in \mathbf{R}$, i.e., by $\overline{X}_i = \sum_{r=1}^{R} w_r X_{ir} = \sum_{r=1}^{R} \frac{W_r}{\sum_r W_r} \frac{L_{ir}}{\Pi_r}$. Technically, M_i is a function of w_r and $X_{ir}/\overline{X}_i$ only, similar to inequality measures. A disproportionality measure, however, describes the inequality across regions of the *proportions* of the variable of main interest and its reference, L_{ir}/Π_r, rather than just the inequality of the variable of main interest. The reference, Π_r, can be chosen by from a wide array of possible references, depending on the research purpose at hand. The population mean serves as a scaling factor. It ensures that the measure assumes a minimum value of zero, if the region-specific proportionality factors are the same in all regions ($L_{ir}/\Pi_r = L_{is}/\Pi_s \ \forall\ r, s \in R$). In addition, it makes the measure invariant to the scales of both the variable of main interest and the reference. The scale of the reference may consequently deviate arbitrarily from that of the variable of main interest.

---

6 Hence, employment in an industry is the ‘variable of main interest’.

7 For measures of specialization, the population is *workers within a region* and their characteristics are the *industries*. Formally, specialization measures can thus be obtained from concentration measures by merely switching the indices for regions and industries.

The taxonomy builds on the three characteristic features of the measures in equation (1): (i) the region-specific weights, $\mathbf{W}_{(r)}$, (ii) the references, $\mathbf{\Pi}_{(r)}$, and (iii) the projection function, f_M. Together with the variable of main interest, $\mathbf{L}_{i(r)}$, the three features unambiguously define a measure. For any empirical investigation, the specification of each characteristic feature should follow directly from the research purpose or the test hypothesis at hand, and take into account the specificities of the available data.

(i) The *region-specific weights*, $\mathbf{W}_{(r)}$, reflect the researcher's choice of the basic units of the analysis (Brülhart and Träger 2005): for measures of concentration, the basic units are spatial units. The variable of main interest is defined as, say, the number of industry i workers *per basic spatial unit*.[8] The region-specific weights ensure that each basic unit is assigned the same weight in calculating the measure. Disproportionality measures allow a variety of different spatial basic units to be specified, provided the variable of main interest as well as the references can be measured consistently in terms of these basic units. Only three types of basic units have, nonetheless, been used in the literature so far: First, the regions themselves have been chosen as basic units, which implies assigning all regions the same weight, independent of their actual sizes or any other characteristics. These basic units are represented by the region-specific weights $\mathbf{W}_{(r)} = \mathbf{1}_{(r)} = (1, \ldots, 1)$ in equation (1).[9] Second, square kilometers (km²) have been chosen as basic units, which implies weighting each region by its geographical size (A_r), i.e., $\mathbf{W}_{(r)} = \mathbf{A}_{(r)} = (A_1, \ldots, A_R)$.[10] And third, the average size of the area attributed to a worker in the region has been chosen as basic units, which implies weighting each region by its total employment, i.e., $\mathbf{W}_{(r)} = \mathbf{L}_{\bullet(r)} = (L_{\bullet 1}, \ldots, L_{\bullet R})$. $L_{\bullet r}$ [$= \Sigma_i L_{ir}$] denotes the sum of workers over all industries in region r. Each worker in region r is taken to represent a share of $1/L_{\bullet r}$ of the region's area.

Measures using regions as basic units will be labeled *unweighted* measures; those using non-uniform region-specific weights *weighted* measures. Weighted measures are invariant to dividing a region into subregions,[11] provided the weights represent the sizes of the regions, and the sub-regions exhibit, or are assumed to exhibit, identical concentration patterns.

(ii) The *reference distribution*, $\mathbf{\Pi}_{(r)}$, reflects the researcher's choice of the benchmark, or the null hypothesis of "no" or "no unusual concentration". Economically meaningful inferences require the reference distribution to pick up any systematic components in the observed

---

8 For measures of specialization, the basic units are units of (sectoral) activities, such that the variable of main interest is defined as, say, the number of region r workers per sectoral unit.

9 By standardizing the sum of weights to one, each region is assigned the relative weight $w_r = W_r/\Sigma_r W_r = 1/R$.

10 As the spatial distribution of workers within the regions cannot be observed in most cases, workers are assumed to be distributed uniformly across space within the region.

11 Haaland et al. (1998) attribute this invariance to relative measures. The taxonomy here makes clear that it is solely due to the choice of the weights.

regional employment patterns that the researcher is *not* willing to label concentration for the research purpose at hand (Combes and Overman 2004). Similarly, anything the researcher wants to label concentration should show up as a deviation of the variable of main interest from its reference. Disproportionality measures allow a great variety of references to be specified, provided the references are defined over the basic units. Only three types of references have, nonetheless, been used in the literature so far. First, the uniform distribution has been chosen as the reference, which implies assuming all regions to be of the same size under the H_0. This reference is represented by $\mathbf{\Pi}_{(r)} = \mathbf{1}_{(r)}$ in equation (1). Uniform references reflect the researcher's emphasis on the qualitative characteristics of regions, or on administrative issues. Second, the distribution of employment observed at a higher-level sectoral aggregate (total regional employment, for example) has been chosen as the reference, which implies assuming the spatial distribution of the industry under investigation to equal that of total employment across all industries under the H_0, i.e., $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$. Aggregate references reflect the researcher's emphasis on controlling for systematic differences between regions in the sizes of the labor force, in the attractiveness to firms or workers, the regulatory frameworks, or other institutional or political factors. And third, the distribution of the geographical sizes of regions has been chosen as the reference, which implies assuming employment in the industry under investigation to be distributed evenly across space under the H_0, i.e., $\mathbf{\Pi}_{(r)} = \mathbf{A}_{(r)}$. Measures based on the uniform reference will henceforth be labeled *absolute* measures, those based on a nonuniform reference *relative* measures.[12]

(iii) The *projection function*, f_M, reflects the researcher's relative emphasis on region-specific proportionality factors of different magnitude. Some measures, such as the Theil index, emphasize variations in the range of low values of the region-specific proportionality factors (e.g., industry is strongly underrepresented in a region), others, such as GE measures with a sensitivity parameter $\alpha > 2$, emphasize variations in the range of high values of the region-specific proportionality factors (e.g., industry is strongly overrepresented).[13] Again others, such as the relative mean deviation (RMD), emphasize changes in the balance between regions with over- and underrepresented industries and incidences of regions "jumping across" the reference.[14]

---

[12] Brülhart and Träger (2005) introduce the term *topographic* measures for measures using the area as a reference. This reference is just one of many possible nonuniform references.

[13] See Table 1 below for a formal definition of the different projection functions.

[14] See Cowell (2000) and Cowell and Flachaire (2002) for more details. In addition to being subject to theoretical considerations, the choice of the projection function may be subject to practical considerations. A projection function that does not put too much emphasis on extreme (positive and/or negative) region-specific proportionality factors may be preferred to reduce the effects of indivisibilities in firm sizes or 'outliers' on the measure. Alternatively, or in addition, the sensitivity of the results can be assessed by comparing the results for different projection functions.

In the literature, concentration and specialization measures have so far been classified by their projection function and their reference distribution (Haaland et al. 1998). The reference and the weights have always been assumed to be the same. Varying the references independently of the region-specific weights has not been considered an option. The present paper argues that this is unnecessarily restrictive. By distinguishing carefully between references and weights, the taxonomy adds one additional degree of freedom to the opportunities to choose an appropriate measure.

Disentangling references and weights is useful for two reasons: First, the research purpose or test hypothesis may require using a reference that differs from the weighting scheme. For example, a study of local policies may require choosing the sphere of influence of local governments as the basic units, i.e., $\mathbf{W}_{(r)} = \mathbf{1}_{(r)}$, while the aggregate regional employment is the proper benchmark, i.e., $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$. Or the research purpose may require to compare the spatial distribution of an industry to that of total employment, i.e., $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$, while controlling for the geographical size of regions, i.e., $\mathbf{W}_{(r)} = \mathbf{A}_{(r)}$. Second, by clarifying the distinct functionality of the weights and the references, the taxonomy facilitates sensitivity testing. Selectively changing the region-specific weights or the references will help in assessing the sensitivity of the preferred measure to a variation of the basic units or the null hypothesis.

Using the taxonomy, four different groups of measures can be defined for each projection function: An *unweighted absolute* measure and various *unweighted relative*, *weighted absolute*, and *weighted relative* measures. Table 1, which will be discussed below in more detail, gives an overview of the general principle of defining disproportionality measures of concentration for selected projection functions: the generalized entropy (*GE*) class of measures, the Theil index (*T*), the coefficient of variation (*CV*),[15] the *RMD*, and the Gini coefficient (*G*). The first column of Table 1 gives, for each projection function, a general form that can be used to derive all related measures. For a given region-industry employment pattern, $\mathbf{L}_{i(r)}$, a measure may be unambiguously defined by choosing a reference distribution, region-specific weights and a projection function. The remaining three columns of Table 1 give three examples of measures obtained for different combinations of weights and references: the *unweighted absolute*, an *unweighted relative*, and a *weighted relative* measure.[16] In order to

---

[15] The Theil index and (a transformed version of) the CV are actually members of the GE class of measures. Owing to their popularity in the literature, they are nonetheless listed separately in Table 1. Table 1 can easily be extended to projection functions based on other measures discussed in the inequality literature (see, e.g., Cowell 1995; Silber 1999).

[16] To save space, weighted absolute measures are omitted, and the relative and the weighted measures are exemplified only for total regional employment as a reference or as weights. All three variants of the measures listed in Table 1 have actually been employed in studies of concentration or specialization, though not for all the projection functions: Among the *weighted relative* measures used in the literature are (i) the so-called Krugman index (weighted relative *RMD*) used, e.g., by Krugman (1991), Hallet (2002), Dohse et al. (2002), and Traistaru et al. (2003); (ii) the so-called relative Theil index (e.g., Brülhart and Träger 2005), (iii)

compare the values of different measures directly, it may be useful to normalize the measures to the (0, 1) interval by dividing them by their upper bounds.[17]

To illustrate the taxonomy, consider first the so-called Krugman index, which, for the concentration of industry i, is defined as

$$K_i = \sum_{r=1}^{R} \left| \lambda_{ir} - \lambda_r \right| := \sum_{r=1}^{R} \left| \frac{L_{ir}}{L_{i\bullet}} - \frac{L_{\bullet r}}{L_{\bullet\bullet}} \right|. \tag{2}$$

As K_i is calculated as the unweighted sum of the absolute region-specific differences in employment shares for industry i, $\lambda_{ir} := L_{ir}/\Sigma_r L_{ir} = L_{ir}/L_{i\bullet}$, and the 'reference', $\lambda_r := L_{\bullet r}/\Sigma_i \Sigma_r L_{ir} = L_{\bullet r}/L_{\bullet\bullet}$, it can be interpreted easily and intuitively: A value of, say, $K_i = 0.5$ indicates that a share of at least one fourth ($\frac{1}{2} K_i$) of the industry's total workforce has to move to another region for the employment distribution to exactly correspond to the reference distribution. The Krugman index has traditionally been classified as a 'relative' measure.

the relative CV (e.g., Brülhart and Träger 2005), and (iv) the 'locational' Gini coefficient, as used by Krugman (1991), Amiti (1998) and Brülhart (2001). An *unweighted relative* measure is the 'locational' Gini coefficient, as used by Südekum (2006). And among the *unweighted absolute* measures are (i) the traditional Gini coefficient (e.g., Aiginger and Leitner 2002, Midelfart-Knarvik et al. 2002), (ii) the Theil index as used by Aiginger and Davies (2004), and (iii) the CV as used by Aiginger and Leitner (2002).

[17] For details on the calculation of upper bounds, see http://www.uni-kiel.de/ifw/staff/bode/ measures.htm.

Table 1 — Disproportionality measures of industrial concentration for selected projection functions, reference distributions and basic units[a]

projection function	general form	unweighted absolute references: $\Pi_r := 1$ weights: $W_r := 1 \Rightarrow w_r = 1/R$	unweighted relative references: $\Pi_r := L_{\bullet r}$ weights: $W_r := 1 \Rightarrow w_r = 1/R$	weighted relative references: $\Pi_r := L_{\bullet r}$ weights: $W_r := L_{\bullet r} \Rightarrow w_r = \lambda_r$
$GE_{(\alpha)}$	$GE_{(\alpha)i} = \omega \sum_{r=1}^{R} w_r \left[\left(\frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \right)^{\alpha} - 1 \right]$	$GE_{(\alpha)i}^{UA} = \omega \left[\frac{1}{R} \sum_{r=1}^{R} (R\lambda_{ir})^{\alpha} - 1 \right]$	$GE_{(\alpha)i}^{UR} = \omega \left[\frac{1}{R} \sum_{r=1}^{R} \left(R \frac{l_{ir}}{\sum_r l_{ir}} \right)^{\alpha} - 1 \right]$	$GE_{(\alpha)i}^{WR} = \omega \left[\sum_{r=1}^{R} \lambda_r (LC_{ir})^{\alpha} - 1 \right]$
Theil $= GE_{(1)}$	$T_i = \sum_{r=1}^{R} w_r \frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \ln \left(\frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}} \right)$	$T_i^{UA} = \sum_{r=1}^{R} \lambda_{ir} \ln (R \cdot \lambda_{ir})$	$T_i^{UR} = \sum_{r=1}^{R} \frac{l_{ir}}{\sum_r l_{ir}} \ln \left(R \frac{l_{ir}}{\sum_r l_{ir}} \right)$	$T_i^{WR} = \sum_{r=1}^{R} \lambda_{ir} \ln (LC_{ir})$
CV $= 2(GE_{(2)})^{0.5}$	$CV_i = \left[\sum_{r=1}^{R} w_r \left[\left(\frac{L_{ir}/\Pi_r}{\sum_r w_r L_{ir}/\Pi_r} \right)^2 - 1 \right] \right]^{0.5}$	$CV_i^{UA} = \left(R \sum_{r=1}^{R} \lambda_{ir}^2 - 1 \right)^{0.5}$	$CV_i^{UR} = \left[R \sum_{r=1}^{R} \left(\frac{l_{ir}}{\sum_r l_{ir}} \right)^2 - 1 \right]^{0.5}$	$CV_i^{WR} = \left(\sum_{r=1}^{R} \lambda_r LC_{ir}^2 - 1 \right)^{0.5}$
RMD / Krugman	$RMD_i = \frac{\sum_{r=1}^{R} w_r \left\lvert \frac{L_{ir}}{\Pi_r} - \sum_r w_r \frac{L_{ir}}{\Pi_r} \right\rvert}{\sum_r w_r L_{ir}/\Pi_r}$	$RMD_i^{UA} = \sum_{r=1}^{R} \left\lvert \lambda_{ir} - \frac{1}{R} \right\rvert$	$RMD_i^{UR} = \sum_{r=1}^{R} \left\lvert \frac{l_{ir}}{\sum_r l_{ir}} - \frac{1}{R} \right\rvert$	$RMD_i^{WR} = \sum_{r=1}^{R} \lvert \lambda_{ir} - \lambda_r \rvert$
Gini	$G_i = \frac{\sum_{r=1}^{R} \sum_{s=1}^{R} w_r w_s \left\lvert \frac{L_{ir}}{\Pi_r} - \frac{L_{is}}{\Pi_s} \right\rvert}{2 \sum_r w_r L_{ir}/\Pi_r}$	$G_i^{UA} = \frac{1}{2R} \sum_{r=1}^{R} \sum_{s=1}^{R} \lvert \lambda_{ir} - \lambda_{is} \rvert$	$G_i^{UR} = \frac{1}{2R \sum_r l_{ir}} \sum_{r=1}^{R} \sum_{s=1}^{R} \lvert l_{ir} - l_{is} \rvert$	$G_i^{WR} = \frac{1}{2} \sum_{r=1}^{R} \sum_{s=1}^{R} \lambda_r \lambda_s \lvert LC_{ir} - LC_{is} \rvert$

[a] The lower bounds of all measures are 0. For the upper bounds see http://www.uni-kiel.de/ifw/staff/bode/measures.htm. Notation: $L_{i\bullet} := \Sigma_r L_{ir}$; $L_{\bullet r} := \Sigma_i L_{ir}$; $l_{ir} := L_{ir}/L_{\bullet r}$; $l_i := L_{i\bullet}/\Sigma_i L_{i\bullet}$; $\lambda_{ir} := L_{ir}/L_{i\bullet}$; $\lambda_r := L_{\bullet r}/\Sigma_r L_{\bullet r}$; $LC_{ir} := l_{ir}/l_i = \lambda_{ir}/\lambda_r$; $\omega := (\alpha^2 - \alpha)^{-1}$.

The taxonomy proposed in the present paper suggests looking at K_i in a slightly different way. By rearranging (2), the Krugman index can be shown to be a *weighted relative* RMD (RMD_i^{WR}; see Table 1, last column):

$$RMD_i^{WR} = f_{RMD}\left(\mathbf{L}_{i(r)}, \mathbf{L}_{\bullet(r)}, \mathbf{L}_{\bullet(r)}\right) = \sum_{r=1}^{R} \frac{L_{\bullet r}}{L_{\bullet\bullet}} \left| \frac{\frac{L_{ir}}{L_{\bullet r}}}{\sum_r \frac{L_{\bullet r}}{L_{\bullet\bullet}} \frac{L_{ir}}{L_{\bullet r}}} - 1 \right| \\ = \sum_{r=1}^{R} \lambda_r \left| LC_{ir} - 1 \right| . \tag{3}$$

Here, $LC_{ir} = \lambda_{ir}/\lambda_r$ denotes the location coefficient for industry i and region r. By setting $\mathbf{\Pi}_{(r)} = \mathbf{W}_{(r)} = \mathbf{L}_{\bullet(r)}$, (3) can alternatively be derived directly from the general definition of the RMD given in the first column of Table 1.

The first line of (3) clarifies the constructive principle of all the disproportionality measures discussed in the present paper: Any disproportionality measure first determines the proportionality factor for each region by comparing the value for the region-industry, L_{ir}, to the corresponding reference value, $L_{\bullet r}$. Second, it converts the region-specific proportionality factors into its specific metric by applying the projection function. The projection function of the *RMD* requires to (i) scale the region-specific proportionality factors by their weighted mean across all regions, $\Sigma_r \lambda_r L_{ir}/L_{\bullet r} = L_{i\bullet}/L_{\bullet\bullet}$ [$:= l_i$], employing the weights determined by the choice of the basic units; (ii) subtract 1; (iii) take the absolute value; and (iv) take the weighted average over all regions, again employing the weights determined by the basic units.

Following the same general principle, any of the three characteristic features of the disproportionality measure may be varied separately. Setting $\mathbf{W}_{(r)} = \mathbf{1}_{(r)}$, and $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$, an *unweighted relative RMD* is obtained as

$$RMD_i^{UR} = f_{RMD}\left(\mathbf{L}_{i(r)}, \mathbf{1}_{(r)}, \mathbf{L}_{\bullet(r)}\right) = \sum_{r=1}^{R} \left| \frac{l_{ir}}{\sum_r l_{ir}} - \frac{1}{R} \right|$$

where $l_{ir} = L_{ir}/L_{\bullet r}$ (see Table 1, third column),. Setting $\mathbf{\Pi}_{(r)} = \mathbf{W}_{(r)} = \mathbf{1}_{(r)}$ gives the *unweighted absolute RMD*,

$$RMD_i^{UA} = f_{RMD}\left(\mathbf{L}_{i(r)}, \mathbf{1}_{(r)}, \mathbf{1}_{(r)}\right) = \sum_{r=1}^{R} \left| \lambda_{ir} - \frac{1}{R} \right| \tag{4}$$

(see Table 1, second column). Comparing the Krugman index, or weighted relative RMD in (2) to the unweighted absolute RMD in (4) clarifies the usefulness of the taxonomy vis-à-vis the traditional distinction of absolute and relative measures: according to the traditional distinction, the two measures differ in just one characteristic, namely the reference. The taxonomy makes clear that the two measures actually differ in two characteristics, the reference *and* the region-specific weights.[18]

Second, consider the generalized entropy class of measures, $GE_{(\alpha)}$. In contrast to the *RMD*, which is frequently characterized as an ad-hoc measure, the GE measures have several useful properties defined by a set of axioms (see, e.g., Cowell 1995; Litchfield 1999). One useful property is decomposability: the total inequality within a population can be decomposed into the inequality within and that between any set of subgroups of the population. The GE class of measures is generally defined as

$$GE_{(\alpha)} = f_{GE(\alpha)}\left(\mathbf{Y}_{(n)}\right) = \frac{1}{\alpha(\alpha-1)} \frac{1}{N} \sum_{n=1}^{N} \left[\left(\frac{Y_n}{\overline{Y}} \right)^{\alpha} - 1 \right] \qquad (-\infty \leq \alpha \leq \infty) \tag{5}$$

for the vector of some characteristics $\mathbf{Y}_{(n)} = (Y_1, \ldots, Y_N)$ of a population. The members of the population are the "basic units", and $\overline{Y} = \frac{1}{N}\sum_n Y_n$ is the mean across the basic units. The parameter α governs the sensitivity of the projection function to changes in the ranges of high and low values of the $Y_n / \overline{Y}$ ratios.[19] The most prominent GE measures are those given by $\alpha = 2$, which is a simple monotonic transformation of the coefficient of variation, $GE_{(2)} = ½CV^2$, and by $\alpha \rightarrow 1$, which is the Theil index, i.e., $GE_{(1)} = T$.

$GE_{(\alpha)}$ in (5) can be decomposed into a within-groups ($GE_{(\alpha)w}$) and a between-groups component ($GE_{(\alpha)b}$), such that $GE_{(\alpha)} = GE_{(\alpha)w} + GE_{(\alpha)b}$. For H subgroups with N_h basic units in subgroup h ($h = 1, \ldots, H$), the between group component is given by

$$GE_{(\alpha)b} = f_{GE(\alpha)}\left(\overline{Y}_{(h)}, \mathbf{N}_{(h)}\right) = \frac{1}{\alpha(\alpha-1)} \sum_{h=1}^{H} \frac{N_h}{N} \left[\left(\frac{\overline{Y}_h}{\overline{Y}} \right)^{\alpha} - 1 \right], \tag{6}$$

where $\overline{Y}_h = \frac{1}{N_h}\sum_{n=1}^{N_h} Y_{nh}$ is the unweighted mean of subgroup h, and $\overline{Y} = \sum_h \frac{N_h}{N} \overline{Y}_h \left[= \frac{1}{N}\sum_n Y_n \right]$ the weighted average of all subgroup means. Y_{nh} denotes the characteristic of the n^{th} member of the h^{th} subgroup.

---

18 For testing the sensitivity of, for example, the weighted relative RMD in (3) to changes in the weighting scheme, it may also be informative to selectively change the weights of the region-specific proportionality factors while keeping the scaling factor of the region-specific proportionality factors unchanged, i.e., to compare (3) to $\Sigma_r|LC_{ir}-1|$. Although the latter expression is not an RMD in terms of the present taxonomy, the comparison may still yield valuable information on the sensitivity of the preferred RMD to the weighting scheme.

19 With $\alpha < 2$, the measure is more sensitive to (mean-preserving) changes in observations with low values of $Y_n / \overline{Y}$; with $\alpha > 2$, it is more sensitive to (mean-preserving) changes in observations with high values of $Y_n / \overline{Y}$ (e.g., Cowell 2000; Cowell and Flachaire 2002).

Traditionally, the [unweighted] absolute $GE_{(\alpha)}$ measures of concentration have been derived from (5) and the [weighted] relative measures from the between-group component (6), assuming the unobservable within-group component to be zero (Brülhart and Träger 2005). The taxonomy proposed in this paper instead suggests using a generalized form of the between-group component (6) as the unique basis for all GE measures of concentration (see Table 1, first row, for examples):

$$GE_{(\alpha)i} = f_{GE(\alpha)}\left(\mathbf{L}_{i(r)}, \mathbf{W}_{(r)}, \mathbf{\Pi}_{(r)}\right) = \frac{1}{\alpha(\alpha-1)}\sum_{r=1}^{R} w_r \left[\left(\frac{\frac{L_{ir}}{\Pi_r}}{\sum_r w_r \frac{L_{ir}}{\Pi_r}}\right)^{\alpha} - 1\right]. \qquad \text{(6a)}$$

All disproportionality measures derived from (6a) share the usual properties of GE measures, provided the variables of main interest, the weights, and the references are related consistently to the basic units.

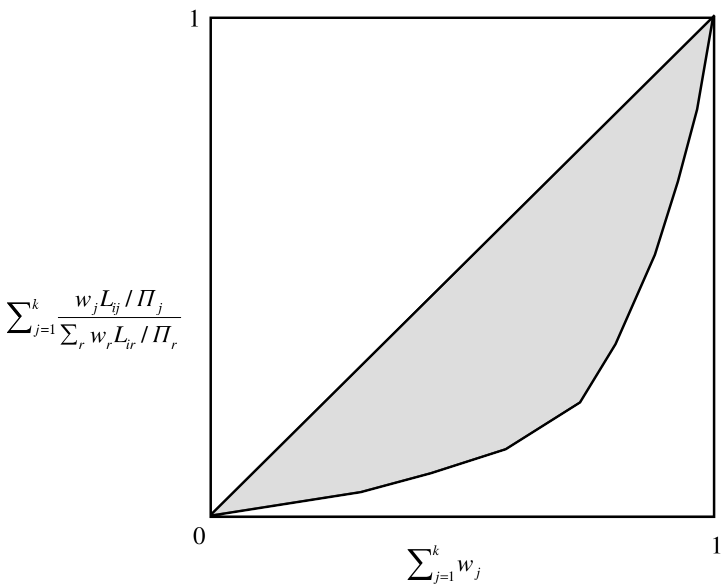
Consider the two examples discussed earlier in this section, and assume the Theil index, $GE_{(1)}$, to be the appropriate projection function in both examples. In the first example, the study of local policy measures, which suggests choosing $\mathbf{W}_{(r)} = \mathbf{1}_{(r)}$ and $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$, the preferred measure should be the unweighted relative Theil index $T_i^{UR} = \sum_{r=1}^{R} \lambda_{ir} \ln(R\lambda_{ir})$. In the second example, the study comparing the spatial distribution of an industry to that of total employment for each km², which suggests choosing $\mathbf{W}_{(r)} = \mathbf{A}_{(r)}$ and $\mathbf{\Pi}_{(r)} = \mathbf{L}_{\bullet(r)}$, the preferred measure should be the weighted relative Theil index $T_i^{WR} = \sum_{r=1}^{R} \frac{A_r}{A} \frac{l_{ir}}{\sum_r \frac{A_r}{A} l_{ir}} \ln\left(\frac{l_{ir}}{\sum_r \frac{A_r}{A} l_{ir}}\right)$.

Notice that the decomposition of the corresponding GE measure or Theil index for the basic units requires assuming that the fraction of industry-i workers in the total workforce does not vary across space within a region, i.e., $L_{in}/L_{\bullet n} = L_{ir}/L_{\bullet r}$ $\forall\, n = 1, \ldots, A_r$. L_{in} and $L_{\bullet n}$ denote industry-i and total employment on the n^{th} km² in region r.

The third and final group of measures illustrated here are measures based on the Gini projection function.[20] The Gini coefficient is generally defined as two times the area between the Lorenz curve and the 45° line (shaded area in Figure 1) in a box plot of cumulated shares of individuals in the population on the horizontal axis and the cumulated shares of their characteristics on the vertical axis. In terms of the taxonomy of the present paper, the population, depicted on the horizontal axis, consists of the basic units, the shares of which are represented by the (relative) region-specific weights, w_r. The characteristics, depicted on the vertical axis,

[20] As the RMD, the Gini coefficient is an intuitively appealing ad hoc measure. It meets the requirements of the axiomatic approach, including decomposability, only under specific conditions.

Figure 1 – Lorenz curve of industrial concentration



are the weighted region-specific proportionality factors, the shares of which are $w_r \frac{L_{ir}}{\Pi_r} \Big/ (\sum_r w_r \frac{L_{ir}}{\Pi_r})$. All observations are sorted in ascending order by the region-specific proportionality factors, L_{ir}/Π_r,. This convention gives rise to

$$
\begin{aligned}
G_i = f_G\left(\mathbf{L}_{i(r)}, \mathbf{W}_{(r)}, \mathbf{\Pi}_{(r)}\right) &= 1 + \frac{1}{\sum_r w_r \dfrac{L_{ir}}{\Pi_r}} \sum_{k=1}^{R} w_k \left(w_k \frac{L_{ik}}{\Pi_k} - 2 \sum_{j=1}^{k} w_j \frac{L_{ik}}{\Pi_j} \right) \\
&= \frac{1}{2 \sum_r w_r \dfrac{L_{ir}}{\Pi_r}} \sum_{r=1}^{R} \sum_{s=1}^{R} w_r w_s \left| \frac{L_{ir}}{\Pi_r} - \frac{L_{is}}{\Pi_s} \right|
\end{aligned}
\tag{7}
$$

as the general form of all Gini disproportionality measures (see Table 1).[21] k in (7) and Figure 1 indexes the observation with the k^{th} lowest region-specific proportionality factor.[22] The Gini coefficients for the various basic units and references can be defined along the same lines as the *RMD* and *GE* measures above (see Table 1).

[21] Note that the expression in the second line of (7) does not require sorting observations by the ratio of the values depicted on the vertical and horizontal axes.

[22] Equation (7) includes all Gini coefficients used in the concentration and specialization literature as special cases. See footnote 17.

2. Generalization 1: Disproportionality Measures of Polarization

Disproportionality measures of polarization of an economy evaluate the concentration of industries and the specialization of regions within the economy simultaneously. Formally, they are straightforward generalizations of the disproportionality measures of concentration or of specialization discussed in Section 2. Rather than evaluating the employment pattern in just one dimension, i.e., in either one industry, $\mathbf{L}_{i(r)}$, or one region, $\mathbf{L}_{(i)r}$ [= (L_{ir}: $i \in \mathbf{I}$)], measures of polarization cover both dimensions simultaneously, thus evaluating all elements of the industry-region employment pattern, $\mathbf{L}_{(ir)}$. Similar measures for two-dimensional data have recently been discussed in the sociological segregation literature (Reardon and Firebaugh 2002).[23] The present paper generalizes these "segregation measures" to integrate them into the taxonomy introduced in the preceding section.

In terms of the taxonomy, polarization measures require, first, specifying an (Ix1) vector of *industry-specific weights*, $\mathbf{W}_{(i)}$ [= (W_i: $i \in \mathbf{I}$)], in addition to the (Rx1) vector of region-specific weights, $\mathbf{W}_{(r)}$. Similar to the region-specific weights, the industry-specific weights reflect the choice of basic units in the sectoral dimension. The sectoral basic units may, for example, be whole industries, represented by the industry-specific weights $\mathbf{W}_{(i)} = \mathbf{1}_{(i)}$, or be related to the activities of individual workers, represented by $\mathbf{W}_{(i)} = \mathbf{L}_{(i)\bullet}$. Second, polarization measures require the reference distribution to be a bivariate distribution represented by the (IxR) matrix $\mathbf{\Pi}_{(ir)}$. For absolute measures the matrix is $\mathbf{1}_{(ir)}$. For relative measures it may take various values. If the references are total employment by industry and region, the matrix is $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}{}^T$, the matrix with element (i, r) equal to $L_{i\bullet}L_{\bullet r}$. If the references for industries are total employment by industry and those for regions are the area by region, the matrix is $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{A}_{(r)}{}^T$. Notice that the definitions and scales of the references may differ across industries or regions.[24]

Table 2 depicts the general forms of the measures for several projection functions, similar to the first column of Table 1. The various weighted and unweighted absolute and relative measures can be derived from these general forms in a way similar to that outlined in the preceding section. To give an example, a weighted relative GE measure of polarization for $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}{}^T$, $\mathbf{W}_{(i)} = \mathbf{L}_{(i)\bullet}$, and $\mathbf{W}_{(r)} = \mathbf{L}_{\bullet(r)}$ can be derived from the general form of the GE polarization measure in Table 2 as

[23] Aiginger and Davies (2004) discuss the relationship between entropy measures of concentration and specialization that are formally equivalent to the (negative) difference between the unweighted absolute Theil index and its upper bound. Cutrini (2006) discusses measures of 'localization' which are similar to the relative weighted polarization measures of this paper.

[24] The references for some industries may, for example, be related to the regions' areas, while those for other industries are related to the regions' total employment.

Table 2 — Disproportionality measures of economic polarization for selected projection functions: general forms[a]

measure	general form
$GE_{(\alpha)}$	$GE_{(\alpha)} = \frac{1}{\alpha(\alpha-1)} \sum_{i=1}^{I}\sum_{r=1}^{R} w_i w_r \left[\left(\frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_i w_r \frac{L_{ir}}{\Pi_{ir}}} \right)^{\alpha} - 1 \right]$
Theil = $GE_{(1)}$	$T = \sum_{i=1}^{I}\sum_{r=1}^{R} w_i w_r \frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_i w_r \frac{L_{ir}}{\Pi_{ir}}} \ln \left(\frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_i w_r \frac{L_{ir}}{\Pi_{ir}}} \right)$
CV = $2(GE_{(2)})^{0.5}$	$CV = \left[\sum_{i=1}^{I}\sum_{r=1}^{R} w_i w_r \left[\left(\frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_i w_r \frac{L_{ir}}{\Pi_{ir}}} \right)^{2} - 1 \right] \right]^{0.5}$
RMD / Krugman	$RMD = \sum_{i=1}^{I}\sum_{r=1}^{R} w_i w_r \left| \frac{\frac{L_{ir}}{\Pi_{ir}}}{\sum_{i,r} w_i w_r \frac{L_{ir}}{\Pi_{ir}}} - 1 \right|$
Gini	$G = \frac{1}{2\sum_{i,r} w_i w_r \frac{L_{ir}}{\Pi_{ir}}} \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{r=1}^{R}\sum_{s=1}^{R} w_i w_j w_r w_s \left| \frac{L_{ir}}{\Pi_{ir}} - \frac{L_{js}}{\Pi_{js}} \right|$

[a] The corresponding unweighted absolute, unweighted relative and weighted relative measures are obtained from the general forms in the same way as described for concentration measures in Section 2 and Table 1. The lower bounds of all measures are 0. For the upper bounds see http://www.uni-kiel.de/ifw/staff/bode/measures.htm.

$$GE_{(\alpha)}^{WR} = f_{GE(\alpha)}\left(\mathbf{L}_{i(r)}, \mathbf{L}_{(i)\bullet}, \mathbf{L}_{\bullet(r)}, \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}\right) = \frac{1}{\alpha(\alpha-1)} \sum_{i=1}^{I}\sum_{r=1}^{R} l_i \lambda_r \left(LC_{ir}^{\alpha} - 1\right); \tag{8}$$

an unweighted relative polarization GE for $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}^{T}$, $\mathbf{W}_{(i)} = \mathbf{1}_{(i)}$, and $\mathbf{W}_{(r)} = \mathbf{1}_{(r)}$ as

$$GE_{(\alpha)}^{UR} = f_{GE(\alpha)}\left(\mathbf{L}_{i(r)}, \mathbf{1}_{(i)}, \mathbf{1}_{(r)}, \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}\right) = \frac{1}{\alpha(\alpha-1)} \frac{1}{IR} \sum_{i=1}^{I}\sum_{r=1}^{R} \left[\left(IR \frac{LC_{ir}}{\sum_i \sum_r LC_{ir}} \right)^{\alpha} - 1 \right]. \tag{9}$$

For addressing baseline criterion (ii) in Combes and Overman (2004), which requires that measures are comparable across industries and regions, the decomposition of GE measures of polarization is a particularly useful tool. The Theil index of polarization (see Table 2), for example, can be decomposed in both the industrial and the regional dimension. Decomposing it in the industrial dimension, for example, yields a within-industries component that is a weighted average of the Theil indices of concentration of the individual industries (see Table 1), and a between-industries component that is the Theil index of specialization in the aggregate economy.[25]

With weighted relative polarization measures that use industry and region totals as references and weights, i.e., $\mathbf{\Pi}_{(ir)} = \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}{}^{T}$, $\mathbf{W}_{(i)} = \mathbf{L}_{(i)\bullet}$, and $\mathbf{W}_{(r)} = \mathbf{L}_{\bullet(r)}$, the polarization measures are simply the weighted averages of the corresponding concentration or specialization measures (Reardon and Firebaugh 2002; Cutrini 2006). This is true not only for the GE measures but also for those measures that do not meet the general decomposability requirement. For the RMD, for example, one gets

$$\begin{aligned} RMD^{WR} = f_{MRD}\left(\mathbf{L}_{(ir)}, \mathbf{L}_{(i)\bullet}, \mathbf{L}_{\bullet(r)}, \mathbf{L}_{(i)\bullet}\mathbf{L}_{\bullet(r)}^{T}\right) &= \sum_{i=1}^{I}\sum_{r=1}^{R} l_i \lambda_r \left| LC_{ir} - 1 \right| \\ &= \sum_{i=1}^{I} l_i RMD_i^{WR} = \sum_{r=1}^{R} \lambda_r RMD_r^{WR}. \end{aligned} \tag{9}$$

3. Generalization 2: Spatial Disproportionality Measures of Concentration

All the disproportionality measures discussed so far are invariant to the spatial ordering of the regions under investigation. Ignoring the spatial ordering of the data gives rise to the checkerboard problem: the measures systematically understate the true degree of concentration if the industry is clustered at a spatial scale larger than the regions under investigation. To avoid the checkerboard problem, measures need to take into account the spatial ordering of the observations. Arbia (2001) suggests combining inequality measures such as the Gini coefficient with statistics of spatial association such as Moran's I or the Getis-Ord statistic.[26] While the inequality measure is informative as to aspatial concentration, the spatial statistic gives an indication of the spatial clustering.

[25] With the Theil index of polarization given by $T = \sum_{i,r} w_i w_r (X_{ir} / \overline{X}) \ln\left(X_{ir} / \overline{X}\right)$, where $X_{ir} = L_{ir}/\Pi_{ir}$ and $\overline{X} = \sum_{i,r} w_i w_r X_{ir}$, the decomposition yields $T = \sum_i w_i (\overline{X}_i / \overline{X}) T_i + T_b$, where $\overline{X}_i = \sum_r w_r X_{ir}$, $T_i = \sum_r w_r (X_{ir} / \overline{X}_i) \ln\left(X_{ir} / \overline{X}_i\right)$, and $T_b = \sum_i w_i \overline{X}_i / \overline{X} \ln\left(\overline{X}_i / \overline{X}\right)$.

[26] Lafourcade and Mion (2005) do essentially the same but test in addition for the effects of firm size by combining the dartboard measure and Moran's I statistic.

Rather than combining aspatial and spatial measures in an ad hoc way, this paper suggests introducing the spatial dimension directly into the disproportionality measures. The resulting measures, which are labeled *spatial* disproportionality measures of concentration, are actually generalizations of the corresponding aspatial measures discussed in Section 2 (see Table 1). The basic idea is to complement the information on the industry in question from each region by the corresponding information on the industry from nearby regions. Reardon and O'Sullivan (2004) suggest doing so in a way similar to a kernel density estimation, or a geographically weighted analysis.[27] More specifically, they suggest defining a measure in terms of the spatially weighted averages of the variables of main interest and the reference. This approach helps lessen, though not completely avoid, the MAUP and the checkerboard problem inherent to any analysis of concentration based on regional aggregates.

To extend the taxonomy introduced in Section 2 to spatial disproportionality measures, L_{ir} and Π_r are redefined as spatially weighted averages,

$$L_{ir}^S = \sum_{q=1}^{R} \phi_{rq} L_{iq} \text{ and } \Pi_r^S = \sum_{q=1}^{R} \phi_{rq} \Pi_q \,,$$

where $\phi_{rq} := \Phi_{rq} / \sum_{q=1}^{R} \Phi_{rq}$ is a nonnegative spatial weight, or spatial discount factor that reflects the 'closeness' of region q to r, and the superscript S denotes spatially weighted averages. The closeness between regions may generally depend on geographic distances, neighborhood patterns, or accessibility. To meet the usual regularity conditions, the weights are row-normalized, such that the weights sum up to one for each region, i.e., $\sum_{q=1}^{R} \phi_{rq} = 1$.

Extending the set of characteristic features of the concentration measures discussed in Section 2 by the row-normalized (RxR) matrix $\mathbf{\Phi}_{(r)} = (\phi_{rq}\!: r, q \in \mathbf{R}; \Sigma_{q\in\mathbf{R}}\, \phi_{rq} = 1)$,[28] and substituting L_{ir}^S and Π_r^S for L_{ir} and Π_r, all measures in Table 1 can be extended to spatial measures of concentration. The general form of the spatial GE measures, e.g., reads

$$GE_{(\alpha)i}^S = f_{GE(\alpha)}\left(\mathbf{L}_{i(r)}, \mathbf{W}_{(r)}, \mathbf{\Pi}_{(r)}, \mathbf{\Phi}_{(r)}\right) = \frac{1}{\alpha(\alpha-1)} \sum_{r=1}^{R} w_r \left[\left(\frac{\dfrac{\sum_{q=1}^{R} \phi_{rq} L_{iq}}{\sum_{q=1}^{R} \phi_{rq} \Pi_q}}{\sum_r w_r \dfrac{\sum_{q=1}^{R} \phi_{rq} L_{iq}}{\sum_{q=1}^{R} \phi_{rq} \Pi_q}} \right)^{\alpha} - 1 \right]. \tag{10}$$

---

[27] Reardon and O'Sullivan (2004) discuss this approach in the context of spatial segregation measures and continuous space. As in the previous sections, the following discussion will focus on disproportionality measures for regional aggregates.

[28] $\phi_{rr} = 1$ and $\phi_{rq} = 0$ for $q \neq r$ gives the corresponding a-spatial measures (Table 1). For notational convenience, $\mathbf{\Phi}_{(r)}$ denotes the matrix of row-normalized weights (ϕ_{rq}) rather than absolute weights (Φ_{rq}). In matrix notation, $\mathbf{L}_{(r)}^S = \mathbf{\Phi}_{(r)} \mathbf{L}_{(r)}$ and $\mathbf{\Pi}_{(r)}^S = \mathbf{\Phi}_{(r)} \mathbf{\Pi}_{(r)}$.

Due to the geographical weighting, the effect of region r on the measure is magnified if industry i is overrepresented (or underrepresented) in both the region itself and its neighbors. And it is reduced if industry i is over- (under-) represented in region r but under- (over-) represented in nearby regions. It should be noted that decomposing the spatial GE measures in the usual way is not possible due to the interdependencies introduced by the geographical weights.[29] A change in concentration between subregions of one region (e.g., a country) may influence the concentration within another region by affecting its subregion-specific proportionality factors nonuniformly.

The spatial measures are capable of reducing biases resulting from the checkerboard problem and the MAUP. The checkerboard problem is reduced by taking into account the geographical ordering of the regions. The MAUP is reduced by geographical smoothing, which addresses the arbitrary boundary problem, and by carefully specifying the intra- and interregional weights, which addresses the scale problem.[30] For georeferenced microdata that provide information on the distances between any pairs of establishments, Duranton and Overman (2005; 2006) and Marcon and Puech (2003; 2005) have recently proposed describing concentration using functions based on the Ripley's K function. The K-based functions, which assign each possible distance a frequency of observations,[31] arguably provide the currently most sophisticated measures of concentration because they avoid the checkerboard problem and the MAUP. The spatial disproportionality measures proposed in the present section are an alternative to the K-based functions. Both approaches may in principle be used for analyzing aggregate or disaggregate data. For any given level of regional aggregation, they are capable of dealing with the checkerboard problem and the MAUP to a similar extent.

4. Conclusion

This paper improves and extends the methodological toolbox for analyzing the regional concentration of industries and industrial specialization of regions. First, it proposes a taxonomy for disproportionality measures of concentration and specialization. The disproportionality

[29] See Reardon and O'Sullivan (2004) for more specific ways of decomposing spatial disproportionality measures.

[30] If transport costs or other spatial transaction costs are considered the determinants of regional interdependencies, the geographical weights could be operationalized by some functions of the geographical or economic distance between any two regions, i.e., $\phi_{rq} = \phi(D_{rq})$, where D_{rq} denotes the distance between the regions q and r, and $\partial\phi/\partial D_{rq} < 0$. See Anselin (1988) for a discussion of alternative forms of geographical weights. One possible way to specify the unobservable *intra*regional distances is to assume that all workers are concentrated at a single regional center. In this case, $D_{rr} = 0$ (but $\phi_{rr} > 0$), and the interregional distances are just the distances between the regional centers. Other possible ways are to assume that all workers are distributed uniformly over space within each region, or to estimate the intraregional distributions from a finer partition of regions provided, for example, by population or electoral statistics.

[31] See Marcon and Puech (2005) for a comparative survey of different K-based functions.

measures can be adjusted more flexibly to the research purpose and data at hand than the inequality measures used in the concentration and specialization literature so far. The taxonomy gives rise to a modular construction system that enables a researcher to unambiguously *define* the disproportionality measure by three characteristic features, the projection function, the reference distribution, and the weighting scheme. Each feature can be determined largely independently of the other two features. The modular system is also useful for systematically evaluating the robustness of the inferences against a variation of the individual features of the measure.

Second, the paper extends and generalizes the taxonomy to disproportionality measures of economic polarization, and spatial disproportionality measures of industrial concentration. Measures of polarization evaluate industrial concentration and regional specialization patterns simultaneously, and render possible a nested analysis of the polarization, specialization and concentration patterns at different spatial and industrial scales. Spatial measures of concentration help address the checkerboard problem and the MAUP, thus posing a promising alternative to K-based statistics. Using spatially weighted averages of the relevant data as an input, the spatial measures allow the specific characteristics of neighboring regions as well as the intra-regional distributions of the variable of interest and the reference to be taken into account.

We are confident that the taxonomy for disproportionality measures proposed in this paper will prove useful for a wide range of empirical studies on concentration, specialization and polarization. Future research should contribute to extending and refining the disproportionality measures and their taxonomy in several respects. First, the taxonomy should be generalized to spatial polarization measures. Second, ways of coping with the counterparts of scale, arbitrary boundary, and checkerboard problems in the sectoral dimension should be explored. Unlike the spatial dimension, where geographical distance or traveling time is widely accepted as a metric for relating the locations of individual units to each other, the sectoral dimension is still lacking a widely accepted metric. A metric for the distances between industries may be based on the coefficients of input-output tables, or on proxies of the similarity of the firms or industries in terms of their input markets, output markets, or technologies (see Conley and Dupor 2003; Bloom et al. 2005). Based on distances between basic units in both the regional and the sectoral dimension, the spatial polarization measures may be extended to spatially and sectorally weighted polarization measures that account for the MAUP and the checkerboard problem in both dimensions. And third, the comparative pros and cons of the spatial disproportionality measures and the K-based functions should be investigated in more detail for both micro and macro data.

References

Aiginger, K., and S.W. Davies (2004), Industrial Specialization and Geographic Concentration: Two Sides of the Same Coin? Not for the European Union. *Journal of Applied Economics* 7 (24): 231–248.

Aiginger K., and W. Leitner (2002), Regional Concentration in the USA and Europe: Who Follows Whom? *Weltwirtschaftliches Archiv* 138 (4): 652–679.

Amiti, M. (1998), New Trade Theories and Industrial Location in the EU. *Oxford Review of Economic Policy* 14 (2): 45–53.

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

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

— (2001), The Role of Spatial Effects in the Empirical Analysis of Regional Concentration. Journal of Geographical Systems 3 (3): 271–281.

Bickenbach, F., E. Bode, and C. Krieger-Boden (2006), Economic Polarization in Europe. Mimeo. Kiel Institute for the World Economy.

Bloom, N., M. Schankerman and J. Van Reenen (2005), Identifying Technological Spillovers and Product Market Rivalry. CEP Discussion Paper 675. London School of Economics.

Bode, E., J. Bradley, G. Fotopoulos et al. (2003), European Integration, Regional Structural Change and Cohesion: A Survey of Theoretical and Empirical Literature. EURECO Working Paper 1. Bonn: Center of European Integration.

Brülhart, M. (2001), Evolving Geographical Concentration of European Manufacturing Industries. *Weltwirtschaftliches Archiv* 137 (2): 215–243.

Brülhart, M., and R. Träger (2005), An Account of Geographic Concentration Patterns in Europe. *Regional Science and Urban Economics* 35 (6): 597–624.

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

Conley, T.G., and B. Dupor (2003), A Spatial Analysis of Sectoral Complementarity. *Journal of Political Economy* 111 (2): 311–351.

Cowell, F.A. (1995), *Measuring Inequality*. 2nd Edition. London: Prentice Hall.

— (2000), Measurement of Inequality. In: A.B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution*, Volume 1. Amsterdam: Elsevier.

Cowell, F.A., and E. Flachaire (2002), Sensitivity of Inequality Measures to Extreme Values. Discussion paper 60, Distributional Analysis Research Programme, STICERD, London School of Economics, London.

Cutrini, E. (2006). The Balassa Index Meets the Theil Index – A Decomposition Methodology for Location Studies. Mimeo, Universita Politecnica delle Marche, Ancona.

Dohse, D., C. Krieger-Boden, and R. Soltwedel (2002). EMU and Regional Labor Market Disparities in Euroland. In: J.R. Cuadrado-Roura, M. Parellada (eds.), *Regional Convergence in the European Union.* Berlin, Heidelberg.

Duranton, G., and H. Overman (2005). Testing for Localisation Using Micro-Geographic Data. *Review of Economic Studies* 72 (4): 1077–1106.

Duranton, G., and H. Overman (2006). Exploring the Detailed Location Patterns of UK Manufacturing Industries using Microgeographic Data. Mimeo.

Ellison, G., and E.L. Glaeser (1997), Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* 105 (5): 889-927.

Haaland, J.I., H.J. Kind, K.H. Midelfart-Knarvik and J. Torstensson (1998), What Determines the Economic Geography of Europe? Discussion paper 98, 19, Norwegian School of Economics and Business Administration, Bergen.

Hallet, M. (2002), Regional Specalization and Concentration in the EU. In J.R. Cuadrado-Roura (ed.), *Regional Convergence in the European Union*. Berlin: Springer.

Krugman, P.R. (1991), *Geography and Trade*. Leuven: Leuven Univ. Press.

Lafourcade, M., and G. Mion (2005), Concentration, Agglomeration and the Size of Plants. Forthcoming in Regional Science and Urban Economics.

Litchfield, J.A. (1999), Inequality Methods and Tools. London: London School of Economics (Download: http://www1.worldbank.org/prem/poverty/inequal/methods/litchfie.pdf).

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

—, — (2005), Measures of the Geographic Concentration of Industries: Improving Distance-based Methods. Mimeo (download: http://e.marcon.free.fr/publications/index.htm).

Maurel, F., and B. Sédillot (1999), A Measure of the Geographic Concentration in French Manufacturing Industries. *Regional Science and Urban Economics* 29: 575-604.

Midelfart-Knarvik, K.H., H.G. Overman, S.J. Redding and A.J. Venables (2002), The Location of European Industry. European Economy, Special Report 2/2002. Brussels: European Commission.

Mori, T., K. Nishikimi and T.E. Smith (2005), A Divergence Statistic for Industrial Localization. *Review of Economics and Statistics* 87 (4): 635–651.

Nijkamp, P., L. Resmini and I. Traistaru (2003), European Integration, Regional Specialization and Location of Industrial Activity: a Survey of Theoretical and Empirical Literature. In I. Traistaru, P. Nijkamp and L. Resmini (eds.), The Emerging Economic Geography in EU Accession Countries. Aldershot: Ashgate.

Openshaw, S., and P.J. Taylor (1979), A Million or So Correlated Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In N. Wrigley and R.J. Bennet (eds), *Statistical Applications in the Spatial Sciences*. London: Pion.

Reardon, S.F., and G. Firebaugh (2002), Measures of Multigroup Segregation. *Sociological Methodology* 32 (1): 33–67.

Reardon, S.F., and D. O'Sullivan (2004), Measures of Spatial Segregation. *Sociological Methodology* 34 (1): 121–162.

Silber, J. (1999) (ed.), *Handbook of Income Inequality Measurement*. Boston: Kluwer.

Südekum, J. (2006), Concentration and Specialisation Trends in Germany since Re-Unification. *Regional Studies* 41 (forthcoming).

Traistaru, I., P. Nijkamp and S. Longhi (2003), Specialization of Regions and Concentration of Industries in EU Accession Countries. In I. Traistaru, P. Nijkamp and L. Resmini (eds.), *The Emerging Economic Geography in EU Accession Countries*. Aldershot: Ashgate.